Interaction in Depth

Maurice H.P.H. van Beurden

The work described in this thesis has been carried out at the Human-Technology Interaction group at the Eindhoven University of Technology, within the EC FP6 MUTED project (Multi-User 3D Television Display) and the EC FP7 HELIUM3D project (High Efficiency Laser-Based Multi-User Multi-Modal 3D Display).

Interaction in Depth / by Maurice Hubertus Petrus Helena van Beurden.
Eindhoven: Eindhoven University of Technology, 2013. –Proefschrift.-

A catalogue record is available from the Eindhoven University of Technology Library.

ISBN: 978-90-8891-687-8
NUR: 778
Keywords: 3D interaction / 3D displays / embodiment

Cover design: Proefschriftmaken.nl || Uitgeverij BOXPress
Printed & Lay Out by: Proefschriftmaken.nl || Uitgeverij BOXPress
Published by: Uitgeverij BOXPress, ’s-Hertogenbosch

Interaction in Depth

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven,
op gezag van de rector magnificus prof.dr.ir. C.J. van Duijn,
voor een commissie aangewezen door het College voor Promoties,
in het openbaar te verdedigen op woensdag 25 september 2013 om 16:00 uur

door

Maurice Hubertus Petrus Helena van Beurden

geboren te Middelburg

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof.dr. A.G.L. Romme
1e promotor:	prof.dr. W.A. IJsselsteijn
2e promotor:	prof.dr. J.F. Juola
copromotor(en):	dr.ir. Y.A.W. de Kort
leden:	prof.dr.ir. J.-B.O.S. Martens
	prof.dr. H. de Ridder (TUD)
	prof.dr. D.K.J. Heylen (UT)

Contents

CHAPTER 5

CHAPTER 6

CHAPTER 1

General introduction

Rapid developments in sensor, processing, storage and communication technologies provide a significant increase in the availability of data, and create a need to visualize these data in an intuitive fashion. The wealth of data has outstripped the capabilities of conventional 2D displays, which are becoming a bottleneck in the human-computer interaction process. Three-dimensional displays and visualization techniques offer a potential improvement, especially when combined with intuitive means of interacting with 3D datasets. Interaction with 3D content will be intuitive when actions performed on a display correspond with how we act in the real world. From the day we are born we use our bodies when interacting with the environment, and this knowledge can be used when interacting with 3D displays.

For a long time computer visualizations have been presented two-dimensionally, allowing interaction with computers and virtual environments on a 2D plane. Although this remains sufficient for many purposes, the growing amount of 3D (spatial) data (e.g., virtual environments, medical imaging, and geophysical data) that have become available, requires these data to be processed and visualized intuitively. Techniques such as object rotation and shadowing increase our depth perception and understanding of 2D visualizations of 3D content. The recent popularity of stereoscopic displays allows more realistic visualization of spatial data, presenting full three-dimensional views. 3D visualization methods are interesting for both entertainment as well as professional contexts. In the second chapter of this thesis the focus will be on performance-oriented contexts discussing how stereoscopic visualization impacts performance of tasks involving 3D content.

The development of 3D content and presentation of this content on stereoscopic displays require interaction methods different from those in more common interactions used for two dimensions (e.g., mouse-based interaction). Much research has focused on the development of interaction techniques that can be used in three dimensions, such as the Cubic Mouse (Fröhlich & Plate, 2000) or the 3D mouse developed by 3DConnexion. For an extensive overview of 3D interaction devices and 3D User Interfaces we refer to Bowman, Kruijff, LaViola and Poupyrev (2005). Although much research has been aimed at developing innovative methods of interaction, the mouse and keyboard still dominate the way we interact with computers today. However, recently there is a trend towards interactions that allow the use of arms and hands to interact with 3D content (e.g., Nintendo Wii, PlayStation Move, Microsoft Kinect[1]), similar to the way we interact in daily life. This change from traditional desktop computing towards interaction that

[1] *Examples of game consoles in which hand and arm movements are used as input when playing games.*

makes use of full body movements may impact the effectiveness of the tasks we perform, our interaction experience, and even might have their repercussions on our perception of the visualized content. Overall, this thesis investigates how 3D interaction maps onto 3D spaces, and to what extent interaction can optimize performance and user-experience, or influence the very nature of perception and understanding of the digital world.

In the current thesis, both stereoscopic visualizations and 3D interaction are studied from a user-centered perspective, involving performance, user experience and perception. Before we discuss our empirical work, we will first review the most relevant literature. We will start with introducing stereoscopic display technology and the basics of binocular vision, followed by an introduction of interaction technologies, and perspectives taken by various researchers when evaluating these technologies. Lastly, we will discuss 3D interaction focusing mainly on two concepts: natural interaction and embodied interaction.

1.1 Stereoscopic displays

1.1.1 Short history

The interest in 3D displays and stereoscopic visualization started around 1833, when Charles Wheatstone created a mirror device, allowing the fusion of two slightly different perspectives of an image. This idea was further developed by David Brewster and Oliver Wendell Holmes into a handheld stereoscope. In Victorian times, stereo images were popular, and many images were produced and sold during that period. Although the principle of stereoscopic cinema was already demonstrated in the early 20th century, the growing popularity of television in people's homes in the early 1950s required cinemas to consider offering something that would enhance the viewers' entertainment experiences (IJsselsteijn, 2003). Between 1952 and 1954 there was a short-lived breakthrough for stereoscopic cinema, but its popularity decreased after 1954 due to issues of visual discomfort and the introduction of competitive formats, such as wide-screen cinema. While the popularity of stereo among the public decreased, research into stereoscopic visualization and display development continued, especially for professional niche markets. Since James Cameron's 3D movie *Avatar* broke box office records in 2009, we are again in a period in which stereoscopic displays and stereoscopic cinema are flourishing. In Hollywood, many major production companies, such as Disney, have committed to producing 3D films, and affordable 3D televisions are currently being introduced in the consumer market. At the same time, stereoscopic 3D computer games are introduced in the home market, with platforms including the handheld Nintendo 3DS, as well as Sony's PlayStation 3 that connects to any 3D-enabled HDTVs. The difference between the current trend and the one in 1952 is that more attention is paid to content generation and avoiding visual discomfort. Meanwhile, the use of stereoscopic displays in professional markets has shown a much more gradual, yet durable acceptance, driven by a number of specialized applications, such as molecular visualization, computer-aided design, remote operation, and volumetric data visualization. Stereoscopic displays have a number of characteristics that make their application to both settings (i.e., professional and entertainment) advantageous. Whereas in entertainment settings the enhanced viewing experience and perceived naturalness are critical to consumer acceptance (IJsselsteijn, 2004; Lambooij, IJsselsteijn, Bouwhuis & Heynderickx, 2010), professional applications benefit most from the enhanced ability to separate an object of interest from its visual surrounding, and to improve relative depth judgment and surface/shape interpretation - e.g., slant, convexities, and concavities (Merritt, 1991; Pastoor, 1993).

Principle of stereoscopic displays

Binocular vision - seeing with two eyes - has various advantages over vision with only one eye. The most important advantages are probably an enlarged field of view and stereopsis. Stereopsis is the ability of our visual system to make depth judgments based on the two unique perspectives of the world provided by the horizontal separation of the eyes. This horizontal separation causes a difference in the relative projections of monocular images onto the left and right retinas. When points from one eye's view are matched to corresponding points in the other eye's view, the retinal disparity variation across the image provides the observer with information about the relative depth structure of objects, as well as the relative distances between objects. Stereopsis thus acts as a strong depth cue, particularly at shorter distances (see Figure 1). A large body of literature focuses on the inner workings of binocular vision, including theories of depth cue combination. For an overview of theories of binocular vision, see Howard and Rogers (2002).

Although binocular depth is an important depth cue, other depth cues such as pictorial depth cues (i.e., cues that can be captured in a photograph or painting), and motion-based depth (i.e., depth created by relative movement of objects separated in depth, induced by the observer or object movement) also enhance our depth perception. Examples of pictorial depth cues include shading, occlusion, relative size, aerial perspective, linear perspective, and texture gradients. Cutting and Vishton (1995) discussed the various sources of visual information that signal depth structure and distance, and estimated their relative depth potency at different distances, based on the available evidence from a broad range of empirical studies on depth perception. They estimated that binocular disparity is a somewhat stronger cue than motion perspective for distances less than 1 meter, and motion perspective is stronger for distances over 1 meter (see Figure 1).

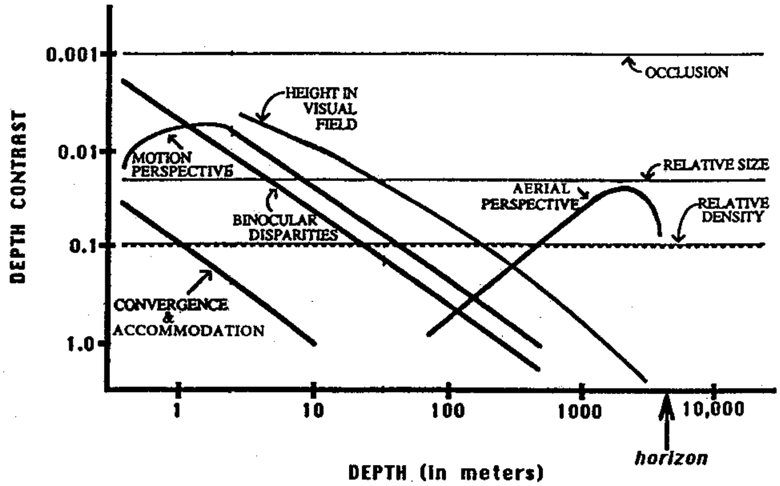


Figure 1. *Just discriminable depth thresholds (depth contrast) as a function of the distance from the observer, showing the relative strength of various depth cues at various distances. The smaller the depth contrast the larger the strength of the depth cue at specific distances. This figure is adopted from Cutting and Vishton (1995).*

For motion-based depth cues, it is useful to distinguish between movement parallax and object motion. Movement parallax is defined as the change in image perspective corresponding to the movements of the user's head position, whereas in object motion the perspective can be changed by observing a moving image or actively manipulating it with an interaction device, such as a computer mouse. Both object motion and movement parallax can enhance depth perception of images presented on a 2D monitor or television. In Chapter 2, we will further elaborate on motion based depth cues and pictorial depth cues in relation to task performance, and compare these to stereoscopic image presentations.

Stereoscopic display techniques are based on the principle of taking two images with a different horizontal perspective, and displaying them in such a way that the left view is seen only by the left eye, and the right view is seen only by the right eye. There are a number of ways of achieving this effect, the most commonly known are shutter glasses, polarised glasses, or anaglyph glasses (blue or green for one eye, red for the other), but in fact there are many more possibilities (Okoshi, 1980; Pastoor, 1997; Sexton & Surman, 1999). Stereoscopic displays can be categorized based on the technique used to channel the right and left images to the appropriate eyes. A distinguishing feature in this regard is whether the display method requires a viewing aid (e.g., glasses) to separate the right and left eye images. Stereoscopic displays that do not require such a viewing aid are known as autostereoscopic displays, having eye-addressing techniques completely integrated into the display itself. Other distinguishing features are whether the display is suitable for more than one viewer (i.e., allows for more than one geometrically correct viewpoint), and whether look-around capabilities are supported. The latter is inherent to some autostereoscopic displays (e.g., holographic or volumetric displays), but requires additional head-tracking when implemented in most other stereoscopic and autostereoscopic displays. In literature on this topic, the term '3D display' is also frequently used in situations in which the content is visualized in 3D perspective, but without the benefit of stereovision. In this thesis, we use the terms *3D display* and *stereoscopic display* interchangeably, to refer to displays in which stereoscopic vision is supported. We reserve the term *perspective imaging* to refer to images that are visualized in perspective on a monoscopic display.

Affordable 3D displays and the increasing amount of 3D content available have given rise to the current popularity of stereoscopy in entertainment settings like cinemas, home entertainment, and digital gaming applications. Although some types of content can be enjoyed passively (e.g., watching movies or photographs), many other applications in both entertainment and professional contexts require active interaction (e.g., selection, manipulation) with the image content. This is likely to influence both experience and task performance, as compared to passive perception of the content. However, interacting with stereoscopic displays requires interaction methods that are in line with the dimensionality of the displayed content, allowing users to manipulate content intuitively in three dimensions. We will provide a brief overview of such 3D interaction technologies in the following section.

1.2 Interaction technologies

Human computer interaction (HCI) is a relatively young research domain. This section will start with a short historical perspective of interaction devices, followed by frequently used paradigms and taxonomies when studying interaction technology.

1.2.1 Historical perspective of interaction devices

There is a general consensus that the field of Human Computer interaction became a professional domain of expertise some thirty years ago, in 1982. In that year, the first conference on human factors in computing was organized (the ACM CHI conference), reflecting the increasing awareness of the importance of the human factor in the development of computer systems. Furthermore, the same year, Time magazine announced that 1982 was the year of the computer (Friedrich, 1983). Since then, the internal components of computers (e.g., processor, hard disk, memory) have become both faster and smaller, doubling computing speed almost every 18 months (i.e., Moore's law). Yet, the interaction methods used to interact with a PC are still dominated by the mouse and keyboard, much like they were developed thirty years ago (see, e.g., Hutchins, Hollan & Norman, 1985). However, the recent introduction of new computing systems (e.g., smartphones and tablet computers) and special game consoles (e.g., Nintendo Wii, Playstation Move, Microsoft Kinect) demonstrate a trend towards interaction methods that support more direct and active methods of interaction (e.g., touch-based and movement-based interaction).

The origin of touch-based interaction can be found in the work done by IBM in the mid-1960 (Buxton, 2010). But research on alternative interaction devices started earlier, around 1950, with the development of the light gun used to identify aircrafts on a screen. The light gun was further developed into the light pen that became a popular method of interacting with displays around 1957, and was used to point and select objects on a screen (Buxton 2012). The first study that compared and evaluated various interaction technologies from a user perspective was performed by English, Engelbart and Berman (1967). In this study they compared the mouse, joystick, light pen, knee control, and some other devices while selecting a character on a screen. The results showed that the mouse and light pen were the most accurate and fastest (interestingly, also knee control resulted in a low completion time, since it did not require additional time to pick up the object). Compared to the mouse, the light pen resulted in a faster but less accurate performance. Novice users perceived the light pen as more natural than mouse-based interaction, however for experienced users the mouse was both faster and more accurate than the light pen. Although the most natural interaction method was the light pen, the mouse (as history has shown) was the most efficient interaction method. Low levels of fatigue, quick transfer to and from the keyboard, and accurate performance made the mouse the most popular device.

1.2.2 Structuring interaction technologies

Within HCI, the user interface generally refers to the intersection at which users interact with technology, including interaction technology and graphical user interface (GUI). The GUI represents the interface that allows us to communicate with technologies such as buttons, icons, or symbols presented on a computer screen. In the current thesis we will reserve the term *interaction technology* for the technology (i.e., hardware/software) through which users interact with a computer - e.g., the infrared sensor of the computer mouse, or the gesture tracker. In addition, we will use the term *interaction method*, to refer to the method in which we communicate with the GUI of a technical product, such as using the mouse or gestures, without a need to specify how gestures are being tracked, or how movements of the mouse are sensed. Previous studies often focused on the GUI and interaction technology, whereas less attention was paid to the method of interaction. Importantly though, when developing a graphical user interface, one should take into account characteristics of the interaction method and technology used during the interaction. For example, when working with touch screens, the user interface should have buttons that can be easily touched by our fingers, and should therefore be larger than icons in the classical WIMP (Windows, Icons, Menus, Pointer) interfaces. For a successful development of new interaction methods, both the user interface and interaction technology are important. In the current thesis, we will focus on the interaction method, since we are interested if the method itself affects experiences of users, independent of the interface through which a user interacts with a system.

Since the breakthrough of the mouse, many other interaction methods have been developed, aiming to increase efficiency and make the interaction more intuitive. The various interaction methods are based on different input classes (e.g., isotonic, elastic, isometric), use different transfer functions (e.g., rate control, position control), or are targeted for specific tasks (e.g., navigating, (3D) manipulating), for more details see Bowman et al. (2005). Several taxonomies were proposed to structure the various developments and interaction devices. These taxonomies classify the devices either based on task characteristics, e.g., selecting, positioning, orienting, navigating, manipulating (Bowman et al., 2005; Foley, Wallace & Chan, 1984), or according to the properties of the interaction technologies, i.e., isotonic or isometric devices, or position-rate control (Buxton, 1983; Card, Mackinlay & Robertson, 1990; Zhai & Milgram, 1993). These taxonomies have been used to identify strong and weak points of the various input devices, and to reveal unexplored future research and development areas concerning interaction technologies. Since HCI as a discipline has its roots in engineering, it is not surprising that many of the proposed taxonomies focused on technical aspects, whereas only a few adopted a user perspective focusing on human performance (Card et al., 1990; Zhai & Milgram, 1993) and the human motor and sensory system (Buxton, 1983). For example, Zhai (1995) conducted a series of studies focusing on user performance, and found that isotonic devices with position control, and isometric devices with rate control, were most appropriate for positioning a 3D object. To study effectiveness and efficiency of interaction devices various models are developed. For example, Fitt's law is used for tasks involving pointing and selecting, predicting the movement time to a target, which depends on both the distance towards a target and the width of that target. The larger the distance and the smaller the target, the longer the movement time will be for a user pointing or selecting a target (see for more detail MacKenzie (1992)). For navigating through a hierarchical menu or 3D worlds, Accot and Zhai (1997) developed the Steering law, describing the relation between movement time and the width of a 'tunnel' through which a user steers a cursor. Investigators generally evaluate experts' task performance for routine tasks (e.g., text editing) with the Keystroke-level model (KLM). In this model, each task (e.g., mouse presses, moving the mouse, decision making) is specified and used to predict completion times (see Hinckley and Wigdor (2002) for a more detailed description). Jacob and Sibert (1992) showed that also the attributes of the task itself affect performance. For example, tasks that have attributes that are related, such as changing size and position, are performed best when the control of these attributes is integrated (e.g., mouse movements in x and y to change the position, and movement in z to change the size of an object). However, for tasks that have attributes that are unrelated, such as changing color and position, performance is better when control of these attributes is separated (i.e., mouse movement in x and y to change the position, and mouse click plus a movement in y to change the color).

What these studies have in common is the focus on usability aspects, measuring the efficiency of the interaction. However, recent studies concerning game applications (e.g., McGloin, Farrar & Krcmar, 2011) showed that although the content is the same, i.e., users played the same game, the experience of users interacting through full body movements was different than that of users interacting via a classic game console. In the next section, we will address the difference between full body interaction and more passive interaction (e.g., classic game console, keyboard) based on two concepts, natural and embodied interaction, which are frequently used in the context of (3D) interaction.

1.3 Natural and embodied interaction

For a long time efforts within HCI were geared towards the development of interaction methods in which users could interact with computers using skills learned in the real world. For example in tangible interfaces, real objects can be used to manipulate and represent digital information, allowing more intuitive interaction through the use of well learned skills when interacting with displayed content (Fitzmaurice, Ishii & Buxton, 1995; Ishii & Ullmer, 1997). In addition, two-handed interaction is useful when users are asked to perform multiple tasks simultaneously, such as changing position and rotating an object at the same time (Buxton & Myers, 1986; Hinckley, Pausch, Proffitt & Kassell, 1998). In literature concerning interaction methods, various terms are used to refer to interaction methods that better correspond with real world interactions (e.g., direct interaction, motion-based interaction). In the current thesis, we will focus on two concepts: natural and embodied interaction. In the literature no consistent definition exists for these concepts, and sometimes they are used to refer to the same phenomena. In this thesis, we will draw a distinction between natural and embodied interaction, since each concept has its own unique perspective on interaction technology and interaction methods.

1.3.1 Natural interaction

Currently we see increasing interest towards interaction methods that adapt to - or are built for - human movements. The result of that trend, at present, is the successful development and introduction of natural interaction techniques, through which users can engage in virtual activities with the same type of movements they would use in the real world. Bowman, McMahan and Ragan (2012) defined both natural interaction and interaction fidelity as "the objective degree with which the actions (characterized by movements, forces, or body parts in use) used for a task in the user interface correspond to the actions used for that task in the real world" (Bowman et al., 2012, p. 79). In our view, naturalness and fidelity are two different concepts, whereas in the definition of Bowman et al (2012) both concepts are related. In Figure 2, we illustrate this by showing the similarities between the three stages of gaining and communicating knowledge in the real world identified by Bruner (1966), and the development of interaction technologies in personal computing.

In Bruner's theory a child first learns to interact with the world by touching, grasping, and manipulating objects around them, which is called enactive knowledge. When growing up, we learn to gain and communicate knowledge through (internal) visual representations, which is called iconic knowledge. The final stage is symbolic knowledge, in which knowledge is represented in symbols like words and numbers. As shown on the right side of Figure 2, the same stages can be identified when interacting with computers, however the order is reversed.

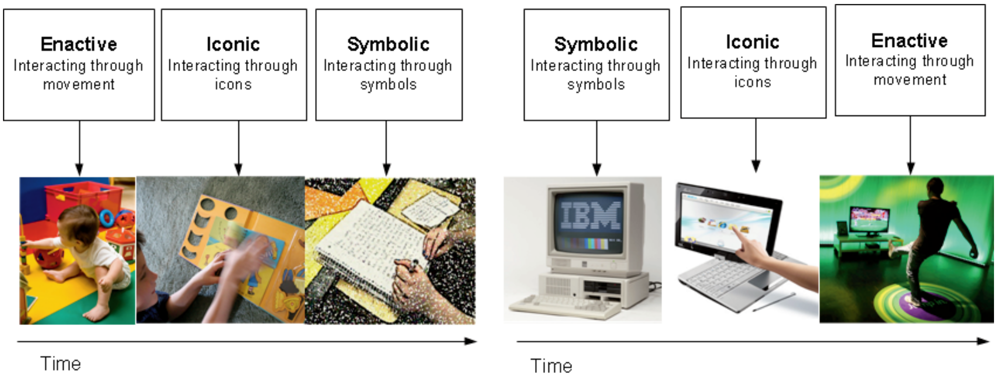


Figure 2. *On the left side, the three stages of cognitive development of a child. On the right side the stages of development of interaction technologies in personal computing.*

The first personal computers only had symbolic interaction, using command languages such as MS-DOS, and using a keyboard for entering text and numbers. The next step was the development of iconic styles of input (WIMP), in which interaction with the computer takes place by interacting with icons on the screen – known, at the time, as direct manipulation interfaces. The final step is enactive interaction, in which movements of the user are used as input to computing systems (e.g., Microsoft Kinect, Nintendo Wii and PlayStation Move). Figure 2 demonstrates two things. First, what is perceived as natural depends on the acquired skills and therefore changes over time. For example, for a child, interacting via symbols is unnatural, whereas for an adult it is a natural method of interacting. In this view also mouse-based interaction is a natural method of interacting with a computer. Second, the enactive stage can be seen as easiest way to interact with content, since it is the first thing we have learned. Therefore, enactive interaction is an intuitive way of interaction for a larger group of users, including naïve users. In addition, interaction based on movements is not the most natural way of interaction for all tasks. For example, the Wii is a natural method of interaction when playing a tennis game, however when creating a word document, Wii-based interaction is unnatural; the more natural interaction method would involve the use of a keyboard. Therefore, in this thesis we define natural interaction as: 'interaction with digital environments, supporting the use of well-learned interaction fitting the characteristics of the tasks'. In this definition, natural interaction is related to one's expertise and objectives, and may change over time. Since we can learn new skills, actions that might be unnatural in the beginning can become natural when users adapt to them. Importantly also, natural interaction is not limited to movements of body parts, but interaction through speech, mouse-based interaction and joysticks can also be regarded as natural.

1.3.2 Embodied Interaction

In contrast to natural interaction, embodied interaction emphasizes bodily engagement during interaction. Before discussing the concept of embodied interaction, we start with a short discussion of the term embodiment. The philosophers Heidegger and Merleau-Ponty were among the first to emphasize the importance of seeing both mind and body as one entity. This view was against popular dualistic approaches, arguing that mind and body should be studied separately, e.g., Descartes. The view that body and mind are closely related influenced many researchers (e.g., Clark, 2008; Dourish, 2001; Gibson, 1979; Noë, 2004). The idea that the human body plays an important role in how we think, feel, and perceive the world is defined as embodied cognition.

Gallagher (2011) recently discussed the various approaches to embodiment, and argued that there is not yet a unified view of what embodiment entails. The different views of embodiment range from minimal involvement of bodily activities on cognitive processes, to an essential role of the body in cognition and perception. In the - what Gallagher called - 'radical embodiment' view, our body and sensory-motor couplings inevitably shape and contribute to consciousness, cognition, and perception. Sensory-motor coupling describes the relation between our movements and corresponding changes in the perceptual (e.g., visual, auditory, tactile, proprioceptive) field. For example, when we move our head from left to right, the world appears to move in the opposite direction, and these visual changes in the apparent environment are used to extract depth and shape of objects. Haans and IJsselsteijn (2012) defined three levels of embodiment: (1) morphology of the body, (2) body schema, and (3) body image. Body morphology drives behavior due to body characteristics (e.g., having wings or legs). Having wings allows birds to fly, and having legs allows humans to walk. The second level is the level of body schema, which allows us to use tools that support our daily activities and experience. Due to the flexible character of our body schema, objects can effortlessly extend our interaction area. The third level of embodiment is the level of body image, which includes our perceptions of our own body, conceptual knowledge we have about our body, and how we experience our body. Although sometimes body image and body schema are used interchangeably, there is a clear distinction between the two. For example, when using a pen when writing a letter, the pen becomes part of our body schema (i.e., it becomes transparent), however we do not experience the pen as part of our body image (i.e., the pen is not part of our conceptual knowledge of our body).

The term embodied interaction was introduced by Dourish (2001) defining it as: "the creation, manipulation and sharing of meaning through engaged interaction with artifacts" (Dourish, 2001, p. 126). In his book, Dourish often refers to tangible computing as an example of embodied interaction, in which real objects are coupled to digital data, and we can use previously gained knowledge and skills when interacting with tangible objects. A second point Dourish makes with the phrase 'sharing meaning' is an essential part of embodied interaction. During interaction, we can use body movements and gestures that are meaningful, and use them in analogous ways to interact with a computer or display screen. In our definition embodiment is the interplay between our body, and the perceptual, cognitive, and emotional responses in the world, using sensory-motor couplings, body representations and meaningful movements. Therefore, in our view, embodied interaction does not represent new, undiscovered interaction methods, but rather a different perspective in which interaction is not only a purposeful means of accomplishing a task, but also impacts cognition, perception, and overall user experience.

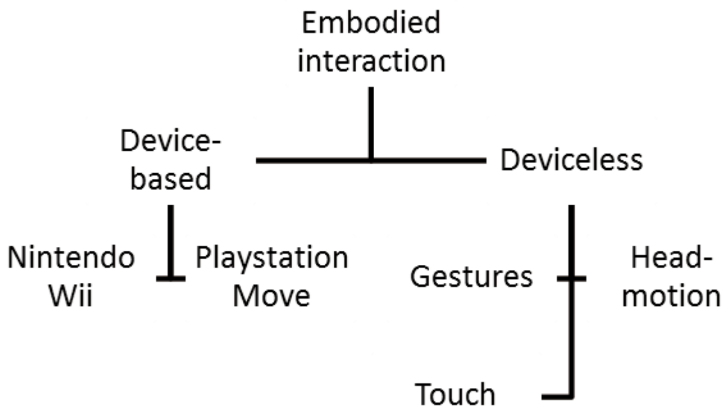


Figure 3. *The different subclasses of embodied interaction.*

In Figure 3, the subclasses of embodied interaction are shown. Embodied interaction can be device-based or deviceless. Examples of device-based embodied interaction are the Wii and the Move, with which we can, for example, use our body movements as input for computer games. Examples of deviceless interaction currently implemented in mobile phones and tablets, is touch-based interaction. Also head motion and gesture-based interaction (e.g., Microsoft Kinect) are potential promising embodied interaction methods, although not yet widely implemented.

1.3.3 What can we learn from the distinction between natural and embodied interaction?

Embodied interaction and natural interaction are two different interaction perspectives, although the two concepts also have certain overlap. As shown in Figure 4, interaction can be embodied, natural, both natural and embodied, or neither embodied nor natural. The last category is reserved for interaction methods that are not appropriate for the task at hand, or when users are not (yet) used to work with the interaction method. An example of an interaction device that is both natural and embodied is the Wii. The Wii makes use of movements as we have learned in daily life, such as playing tennis (i.e., natural), and it corresponds to our previously learned sensory-motor actions when we play tennis (i.e., embodied). On the other hand, the mouse is an example of a natural interaction technology that is not embodied. It is natural since we have learned how to interact with it, however it is not embodied since a forward movement of the mouse is coupled to an upward movements of the cursor, which is not in line with those expected based on only our hand movements.

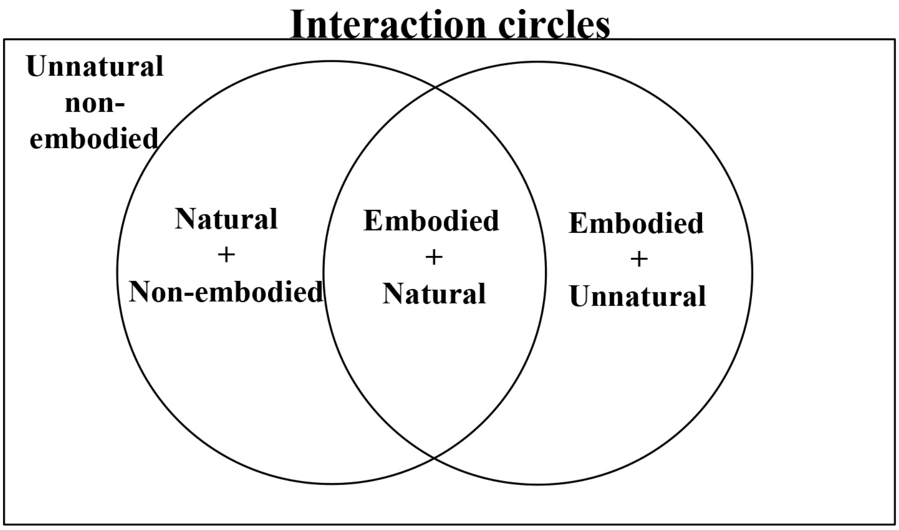


Figure 4. *Interaction Venn diagram representing a classification of natural and embodied interaction. Some interaction methods are either natural or embodied, whereas other interaction technologies are both natural and embodied. The other interactions methods are both unnatural and non-embodied.*

The last category is interaction that is embodied but not natural. An example is a light switch, in which the luminance level is based on how hard one presses the button[2]. This example is embodied, since we use meaning of our bodily action (pressing harder means more light) for the input, however it is unnatural (although it might become more natural after practice), since we are not used to interact with a light switch in such a way. Please note that how natural an interaction is, depends on the skills of the users (e.g., expert, novice) as well as their objectives (e.g., fun, efficient work). Furthermore, devices placed in the diagram might move from one category to another over time. The perspective of embodied interaction aims to make interaction personally relevant in terms of experience, cognition, and perception, and therefore it goes beyond traditional aims when interacting with computers - i.e., making interaction more efficient. Thus, interaction from an embodied perspective should be studied utilizing a measurement approach, which goes beyond traditional usability measures such as efficiency and satisfaction. In the current thesis, we will extend currently applied usability measures towards a broader perspective of user experience when studying 3D interaction (see Chapters 2, 3 and 4). In addition, the role of embodied interaction in perception is studied in Chapter 5.

1.4 Rationale and overview of the thesis

In the general introduction, we suggested that 3D displays can be used for more intuitive visualizations of complex data, and increasing presence and naturalness of the images displayed on a television or monitor. Interaction technologies on the other hand, are often developed for interaction with 2D content, supporting interaction in two dimensions. The increasing computing power and more accurate sensor technologies have inspired both the development of new display devices, as well as new interaction technologies. More natural interaction methods with 3D displays - correctly mapped in spatial dimensions, and corresponding to previously learned skills - will likely enhance the effectiveness and experience of this interaction. Recent developments in interaction technologies also show a more prominent role of the body in the interaction. This potentially makes the interaction not only more natural, but also more embodied. Embodied interaction may positively impact users' emotions and decrease cognitive load, but may also impact users' perception of the environment. Therefore, to understand the effects of natural and embodied interaction, it should be studied from a perspective encompassing more than currently applied usability indicators such as efficiency, satisfaction, and learnability.

In *Chapter 2,* we extend the current 3D display evaluation methods by applying the concept of perceived workload in addition to completion times and accuracy, to

[2] *Not yet published, however a demonstration was given during the Dutch Design Week exhibition 2012 at Eindhoven University of Technology.*

better understand the benefits of stereoscopic visualization in performance-oriented contexts. In three subsequent experiments, we explore the role of disparity level on task performance using stereoscopic displays, and study the contribution of motion-based depth cues vs. stereo visualization in task performance.

Chapter 3 provides a user-centered assessment of embodied interaction, and more specifically on gesture-based interaction. In Experiment 4, we determine the range and variability in gestures that are made naturally when interacting with 2D surfaces and 3D volumes. The outcome of this study is then used in the design of a gesture tracker studied in Experiment 5. In this experiment, gesture-based interaction is evaluated against more traditional mouse-based interaction. For this evaluation we extended the currently applied usability perspective by including experienced hedonic quality and fun. The same measures are then used in Experiment 6, which compares two embodied interaction methods, i.e., Wii (device based) and gesture-based (deviceless) interaction.

Chapter 4 extends the work of chapters 2 and 3, replicating the main experimental comparison, and including additional outcome measures. In Experiment 7 we (1) investigate gesture and mouse-based interaction in a performance oriented context; (2) investigate the effects of stereoscopic presentation on user experience; (3) extend user experience measures with those concerning affect and image quality and, (4) using an optimized disparity level between those used in Experiments 2 and 3 of Chapter 2, to assess efficiency and perceived workload during the task.

Chapter 5 examines the role of embodied interaction in perception, as people perceive an environment not only in terms of its behaviorally independent visual properties, but also in terms of their ability to act in it. When interacting in daily life, the movements we make, and the corresponding actions in the real world have a constant, predictable relationship. In virtual environments this is not necessarily the case, as many parameters can be set depending on the application. In Experiments 8 and 9, we study users' distance estimations between objects on the screen by manipulating the gain of the interaction device (i.e., the relation between our hand movement and corresponding changes on the screen). We study this in 2D and 3D environments.

In *Chapter 6,* we summarize and discuss the main findings of this thesis. In this discussion we will provide an overview of our findings in the context of human-computer interaction. In addition, implications as well as future research directions will be discussed in the light of future computing applications using displays as well as virtual environments.

CHAPTER 2

Stereoscopic Display Evaluation in Performance Oriented Contexts[3]

"If we had three-dimensional environments that allowed us to reach in and move things, then we would appreciate stereo technology more" Colin Ware (2008, p. 94).

[3] *Experiment 1 has been reported in: Beurden van M.H.P.H., IJsselsteijn W.A., Kort de Y.A.W. (2011). Evaluating stereoscopic displays: Both efficiency measures and perceived workload sensitive to manipulations in binocular disparity. Proceedings of SPIE-IS&T Electronic imaging 7863:786316 1 – 786316 7.*

Experiment 2 has been reported in: Beurden van M.H.P.H., Kuijsters A., IJsselsteijn, W.A. (2010). Performance of a path tracing task using stereo and motion based depth cues, Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop; 176-181.

2.1 Introduction

Compared to monocular vision (i.e., using one eye), binocular vision (i.e., using two eyes), is particularly useful for perceiving distances to, and shapes of, objects in daily life (Allison, Gillam & Vecellio, 2009; McKee & Taylor, 2010; Servos, Goodale & Jakobson, 1992). Binocular vision increases depth discrimination between objects (McKee & Taylor, 2010). For aimed movements it produces shorter movement times, higher peak velocities, shorter deceleration phases, and smaller grip apertures than in monocular vision (Servos et al., 1992). Binocular depth cues, i.e., disparity information, provide us with a strong sense of depth, although various other depth cues (e.g., motion, shadow, perspective, occlusion) also help us to understand and interpret the 3D environment in which we live, and to estimate distances and sizes of objects. Many depth cues are generally available when viewing content on standard 2D displays. However, 3D displays capable of presenting stereoscopic images have demonstrated advantages over 2D displays. Studies have shown that the presentation of stereoscopic content enhances the viewing experience and naturalness compared to monocular presentation of content (Lambooij, IJsselsteijn, Bouwhuis & Heynderickx, 2010). Studies also suggested that stereoscopic presentation may improve task performance in terms of lower completion times and/or fewer errors (Getty & Green, 2007; Merritt, 1991; Smith, Cole, Merritt & Pepper, 1979). A recent review of the application of stereoscopic displays in the medical domain revealed mixed results in terms of merits, however a disadvantage for the use of stereoscopic presentations in terms of performance was never found (Beurden, IJsselsteijn & Juola, 2012). The potential advantages of binocular depth cues include: easier relative depth judgments, ability to pick out camouflaged objects, ability to concentrate on objects located at different depth levels, better judgment of surface curvature, and the fact that potential degradations of the 2D image (e.g., lower resolution, limited grey scale, noise) become less disagreeable when they are presented in stereo (Merritt, 1991; Pastoor, 1993). On the other hand, stereoscopic displays are often associated with visual discomfort (Lambooij, IJsselsteijn, Fortuin & Heynderickx, 2009), which potentially leads to decreased task performance (Roufs & Boschman, 1991).

In the current chapter we study stereoscopic displays in a performance-oriented context, i.e., for professional applications such as medical diagnosis, surgery, and aviation. First, we will review literature to understand the relevant factors (e.g., type of depth cues, task difficulty) impacting task performance on 2D and/or 3D displays. In addition, we will look into frequently used measures when determining task performance on 3D displays, and will discuss how these measures can be extended to better understand human performance. These findings lead to three experiments discussed in this chapter, aiming to better understand task performance on stereoscopic displays in different conditions.

2.2 Task performance using 2D vs. 3D displays

Veridical perception of spatial structures and spatial layouts is important for tasks performed both in daily life as well as on computer screens. Depth presented on displays can be enhanced using pictorial or non-pictorial depth cues (Hershenson, 1999; Ware et al., 2008). Pictorial depth cues are those that can be captured in a photograph or painting, whereas non-pictorial depth cues are those gained by motion (object motion or movement parallax) or stereopsis (binocular parallax). Many studies cited in this thesis use the term monoscopic depth cues to refer to both pictorial as well as motion-based depth cues. In this thesis we differentiate between pictorial depth cues, motion-based depth cues and binocular depth cues, which can all be presented on a display. For pictorial and motion-based depth cues a 2D display is sufficient, whereas for binocular depth a 3D display is required.

2.2.1 Task performance using pictorial and stereoscopic depth cues

Various studies have concentrated on how pictorial depth cues can enhance task performance. A comprehensive review of pictorial depth cues is beyond the focus of this thesis, and therefore we limited ourselves discussing the most relevant ones in the present context. For a more comprehensive review we refer to Cutting and Vishton (1995). A number of studies have compared task performance using 3D displays and 2D displays for different type of tasks, such as spatial arrangement of objects (e.g., Hendrix & Barfield, 1995; Hu, Gooch, Creem-Regehr & Thompson, 2002; Hubona, Shirah & Jennings 2004; Hubona & Shirah, 2005; Yeh & Silverstein 1992) or visual motor tasks (e.g., McWhorter, Hodges & Rodriguez, 1991; Smith et al, 1979). These studies have shown that both pictorial and stereoscopic depth cues decrease completion times and error rates compared to scenes that do not provide any of these cues. In the majority of studies, stereoscopic displays increased performance in various tasks compared to monoscopic displays (Hu et al., 2002; Hubona & Shirah, 2005; McWhorter et al., 1991; Yeh & Silverstein, 1992). However, for some tasks (e.g., altitude or azimuth judgment between objects) the use of shadows or drop-lines increased the accuracy of users' judgments to levels beyond which stereo could not further increase performance (Barfield & Rosenberg, 1995; Hendrix & Barfield, 1995). Nevertheless, the previous studies have shown that stereopsis is a powerful depth cue that can enhance performance in a wide variety of tasks (e.g., object placement, resizing and positioning of objects, perceiving distances), either with or without pictorial depth cues (Barfield & Rosenberg, 1995; Hu et al., 2002; Hubona & Shirah, 2005; McWhorter et al., 1991; Yeh & Silverstein, 1992). Furthermore, in scenes in which pictorial depth cues are degraded, scene and task complexity are increased, or ambiguity of objects is higher, the advantage of stereoscopic depth becomes more pronounced (Smith et al., 1997; Yeh & Silverstein, 1992). In our view, stereoscopic displays can be used to optimize the performance in tasks that require accurate and fast interpretation of spatial layout or object placement and manipulation. In addition to pictorial depth cues, motion is another strong depth cue that - similar to stereo - can increase performance in various tasks (Hubona & Shirah, 2005).

2.2.2 *Task performance using motion and stereoscopic depth cues*

A special type of monocular depth cue is motion, providing temporally integrated, successive views of an environment. Mathematically, motion provides the same information to the visual system as spatially integrated (stereoscopic) views (Rogers & Graham, 1982). Motion is therefore frequently used to enhance depth perception. One can identify two types of motion-based depth cues when interacting with computers: movement parallax (MP) and object motion (OM). Movement parallax is defined as the change in image perspective corresponding to movements of the user's head. When participants move their heads, objects on the foreground move faster than objects in the background. This information is used by the brain to extract the relative positions of objects in the environment. The same principle is used during OM. However, OM refers to the perspective changes as a result of, for instance, on-screen object rotation (sometimes referred to as the Kinetic Depth Effect (Wallach & O'Connell, 1953)). It is worth noting that two types of OM can be identified; user-controlled (using an interaction device to rotate an object) and uncontrolled (the object rotates at a constant speed). In contrast to OM, MP is always user-controlled; i.e., the content changes according to the position of the user's head. To date, all computers have the capability of using OM, however MP is not yet widely adopted in current computer systems. One notable exception is virtual reality (VR) using a head-tracked, head-mounted display (HMD). In VR, movement parallax is the essential feature that creates a sense of presence in the VR environment (Dinh, Walker, Song, Kobayashi & Hodges, 1999).

An overview of the studies focusing on the effectiveness of motion-based depth cues, stereo and a combination of motion and stereo is presented in Table 1. In this table, we compared studies on different aspects: nature of the task, number of participants, conditions used in the experiment, and performance measures. The overview shows that in all studies performance is better for stereo than for static monocular presentations: adding stereoscopic depth decreases both the number of errors and completion times (Sollenberger & Milgram, 1993; Faubert, 2001; Ware, Hui & Franck, 1993; Ware & Mitchell, 2008). Adding motion cues also generally improved performance compared to static monocular presentations (Sollenberger & Milgram, 1993; Faubert, 2001; Ware & Mitchell, 2008), the only exception being the study by Ware and colleagues (1993), which did in fact reported better accuracy, but longer completion times for monocular presentation with vs. without motion cues. Direct comparisons of stereo vs. motion cues show that stereo is typically more effective in terms of decreasing completion time (Naepflin & Menozzi, 2001; Ware et al., 1993; Ware & Mitchell, 2008), whereas motion cues generally resulted in better accuracy (Faubert, 2001; Naepflin & Menozzi, 2001; Sollenberger & Milgram, 1993; Ware et al., 1993). Combinations of stereo and motion cues showed mixed results. For instance, completion times were shorter for motion cues combined with stereo than motion cues alone, but not necessarily better than stereo alone (Hubona et al., 1997; Naepflin & Menozzi, 2001; Ware et al., 1993). In fact, stereo alone resulted in faster task completion than a combination of stereo and MP (Naepflin & Menozzi, 2001), and stereo and OM (Ware & Mitchell, 2008). In terms of accuracy, some studies reported better results for the combination of cues than for either stereo or motion alone (Hubona et al., 1997; Sollenberger & Milgram, 1993; Ware et al., 1993; Ware & Mitchell, 2008) – all these studies employed OM; others showed that the combination of stereo with motion cues outperformed stereo in terms of accuracy, but equals accuracy with motion cues alone (Faubert, 2001; Naepflin & Menozzi, 2001 – these two studies employed MP). Table 1 shows that between the various types of motion, controlled OM produced more accurate responses than uncontrolled OM, whereas uncontrolled OM produced shorter completion times (Hubona et al., 1997; Ware et al., 1993).

This indicates that although the visual information is the same under user-controlled or uncontrolled object motion (assuming the same translations or rotations of an object) performance can be different. This result is in line with studies concerning estimations of slant (Boxtel, Wexler & Droulez, 2003) and sizes of objects (Combe & Wexler, 2010). These studies suggest that in addition to perceptive information, motor information also contributes to task performance (Wexler & Boxtel, 2005), which is in line with our embodied perspective described in Chapter 1. According to our definition of embodied interaction, we expect that head motion (MP) should reveal optimal performance, since the mapping between our own movements and the corresponding sensory changes on the screen are more direct than those based on changes produced by using for example a mouse.

Table 1: *Overview of studies investigating the effectiveness of controlled object motion (OM (c)), uncontrolled object motion (OM (uc)), movement parallax (MP) and stereo-vision for spatial task performance.*

Study	Task	N	Conditions				Best performance on top		Remarks
			OM	MP	stereo	stereo + motion	Completion times	Accuracy (% correct)	
Sollenberger and Milgram (1993)	Path tracing task 1 difficulty level	16	√	n/a	√	√	N/A	(stereo + OM(c)) OM (c) stereo	User had 12 seconds to complete the task
Naepflin and Menozzi (2001)	Path tracing task 3 difficulty levels	20	n/a	√	√	√	Stereo (stereo +MP) MP	(stereo + MP) and MP stereo	Difficulty levels were determined after the experiment
Faubert (2001)	Rod positioning task	5	n/a	√	√	√	N/A	(stereo + MP) and MP Stereo	Accuracy was expressed in positioning error
Hubona et al. (1997)	Mental rotation task	31	√	n/a	n/a	√	(stereo +OM (uc)) (stereo +OM (c)) OM (uc) OM(c)	(stereo +OM (c)) (stereo +OM (uc)) OM(c) OM (uc)	
Ware et al. (1993)	Node connection task. 1 difficulty level	11	√	√	√	√	stereo +OM (uc) stereo, (stereo + MP), (stereo + OM(c)), no cues OM (uc) MP OM (c)	(stereo +OM (c)) (stereo + MP) (stereo +OM (uc)) OM (uc) +MP OM (c) Stereo no cues	
Ware and Michell (2008)	Node connection task 4 difficulty levels	14	√	n/a	√	√	Stereo (stereo + OM (uc)) and OM (uc)	(stereo + OM(uc)) OM (uc) and stereo	The graphs were shown for 5s

Note. For each study we listed the task used in the experiment, number of participants (N), conditions used in the experiment and general remarks. In the heading 'best performance on top', the conditions are ranked based on the performance, where the best performance; i.e., lowest completion time or highest accuracy, is listed first.

The only direct comparison between OM (mouse-based) and MP (head-coupled) reported in the literature showed that MP indeed produced more efficient and more accurate responses than OM, however only in conditions without stereo (see Table 1). For stereo visualizations, results showed that OM led to more accurate performance than MP, but in terms of completion times, no difference emerged between MP and OM (Ware et al. 1993).

From the previous discussion, we can conclude that stereo enhances performance in terms of completion times, whereas motion is more effective in enhancing accuracy. Furthermore, the effectiveness when combining motion and stereo showed inconsistent findings. This may be explained by the use of different tasks, and therefore varying levels of complexities in various studies. Both Ware and Mitchell (2008) and Naepflin and Menozzi (2001) manipulated task complexity in their experiments. Ware and Mitchell (2008) showed that results in terms of completion times slightly varied between the easy and difficult tasks. For easy tasks, completion times were similar between conditions using static stereo and stereo combined with motion cues, whereas for difficult tasks stereo alone resulted in shorter completion times than when it was combined with motion. No difference was found between the two difficulty levels in terms of accuracy. Naepflin and Menozzi (2001) did not find a different pattern of results as a function of difficulty level, however, in their study, difficulty level was determined after the experiment, based on the percentage of correct answers for each task. Nevertheless, task complexity may be a relevant factor in understanding the effectiveness of stereo displays. Although a few studies hint at this possibility, the literature at this moment is too scarce to draw any final conclusions. Future studies should try to incorporate task difficulty as a factor in the design. In the following section we will discuss an additional factor that may explain part of the variance in findings. This factor pertains to the disparity level employed in studies using stereoscopic displays.

2.2.3 Task Performance and disparity levels

In Chapter 1, we explained that depth in 3D displays is created by providing a different view to each eye. The amount of depth perceived in these displays (disparity level) can be varied by changing the horizontal separation between these two views, i.e., by providing a larger difference in perspective between the left and right images. Technology allows us to vary disparity level over a wide range, yet levels that are too low or too high may negatively impact the performance on a task. A disparity level that is too small may not be effective since it is barely visible, whereas a disparity level that is too large may induce visual discomfort. Importantly, the majority of studies investigating performance on 3D tasks employed only one level of binocular disparity. Moreover, in many of the studies reported in earlier sections the disparity level used was not specified, making it difficult to estimate the effect of disparity level on task performance. Lastly, studies that did report the disparity level often used different expressions to quantify it (min of arc, camera-base distance, inter-ocular distance), again complicating the comparison of findings across studies. Disparity levels defined in terms of camera-base distance or inter-ocular distance both require viewing distance to estimate the disparity perceived by the user. A distance independent measure for disparity level is min of arc, that is the disparity that falls on the user eye and therefore better describes the level of disparity perceived by the observer.

Table 2: *Estimated optimal disparities in min of arc, based on the data and information available in the three studies listed below.*

	Original range	*Estimated range (min of arc)*	*Optimal disparity level (min of arc)*
De la Rosa et al. (2008)	0-14 min of arc	n/a	>4 (completion time) >3 (error rate)
Fishman et al. (2008)	0-73 mm (camera-base distance)	x**	-
Rosenberg (1993)	0-8 cm (camera-base distance)	0-80*	± 30

* the disparity levels were estimated by making a rough approximation of the screen disparity based on the information (e.g., viewing distance, camera-base distance) provided in the manuscripts, and should therefore be treated as such.
** The magnification of the setting was unknown, therefore we were not able to estimates the screen disparity.

To our knowledge, only a limited number of studies explicitly studied the effect of various disparity levels with respect to performance benefits of stereoscopic displays (De la Rosa, Moraglia & Schneider 2008; Fishman, Ellis, Hasser & Stern, 2008; Rosenberg, 1993). Unfortunately, these studies again employed different definitions of disparity; i.e., min of arc in De la Rosa and colleagues (2008), versus camera-base distance in cm in Fishman and colleagues (2008) and Rosenberg (1993). To get a feeling for the disparity level used in these experiments we roughly estimated the maximum disparity (in min of arc) used in the studies by Rosenberg (see Table 2). In the study by Fishman, we were not able to estimate the disparity levels, since an additional magnification factor was not reported in their paper. In the study by De la Rosa (2008), a visual search task was used with two search planes separated in depth. The results showed that when using disparity levels below 4 min of arc, the items of both depth planes intruded upon each other, whereas for disparity levels larger than 4 min of arc the depth planes were clearly separated, resulting in the fastest completion times. In terms of error rate, optimal performance was reached at a level above 3 min of arc, showing that - although depth planes could intrude upon each other - users were able to successfully complete the visual search task. For both completion times and error rates the performance remained constant for disparity levels up to 16 min of arc, which was the maximum disparity used in this study.

A similar pattern of results was found in the study by Rosenberg (1993), in which users horizontally aligned two pegs, yet with a different optimal disparity level. The results showed that the alignment error decreased with interocular distance up to 3 cm (± 30 min of arc) at a viewing distance of ± 80 cm. A further increase in disparity (interocular distances up to 8 cm corresponding to ± 80 min of arc) did not result in a further increase in performance. In the study by Fishman and colleagues (2008), users performed a ring placement task, placing a ring on a wire using the Da Vinci telesurgery system. Participants started with a camera-base distance of 72 mm, and the camera-base distance decreased in four steps to 0 mm (mono). Results showed that a decrease of 25-30 percent (to 55mm) did not decrease completion times, but further reductions in camera-base distance did. As indicated earlier, comparisons are complicated due to the differences in definitions and disparity indicators employed, as well as the different tasks used in these studies. Nevertheless, these studies showed a similar trend: a performance increase with increasing disparity up to a certain threshold level, after which performance did not improve with the introduction of larger disparities (in other words, a ceiling effect). According to the studies by De la Rosa (2008) and Fischman et al. (2008), disparity levels between 10-30 min of arc render optimal performance. Notably, none of the studies demonstrated a measurable decrease in performance with increasing disparity level, even though in the study of Rosenberg (1993) the maximum disparity was 80 min of arc, which is beyond the zone of comfortable viewing (approximately 60 min of arc, Lambooij et al., 2009).

2.2.4 Efficiency measures and workload

The studies reviewed above as well as those discussed in a recent review on stereoscopic displays in medicine (Beurden et al., 2012), showed that task performance is generally assessed in terms of completion times and percentage correct. Both are useful indicators of task efficiency (and effectiveness), however, additional indicators exist that might also prove relevant in various domains. For instance, in many professional contexts the availability of data is increasing, resulting in more detailed and complex visualizations of spatial structures (e.g., MRI and CT visualizations in medicine). More intuitive visualizations of these data structures might decrease cognitive load, and therefore increase performance. Cases in which the use of stereoscopic displays proves beneficial to task efficiency, one could also expect a decrease in experienced workload. However, we should not assume that primary task measures in themselves are sensitive as workload measures (O'Donnell & Eggermeier, 1986). For instance, if task complexity increases, adequate task performance could be attained by increasing attention or cognitive resources directed towards the task. In such cases, primary task performance indicators such as completion time and percent correct may remain constant, even though workload effectively increases. The cost of increasing workload would only become apparent on these indicators when no spare cognitive resources are left.

There are a number of approaches available to assess workload levels: task performance, subjective measures, and physiological measures. Although primary task performance sometimes provides information of the workload associated with a task, results should be interpreted with care as argued above. Adding a secondary task to the primary task can be a useful method to distinguish between differences in workload. The idea behind the use of a secondary task is that secondary task performance decreases when workload required for the primary task increases. One of the main drawbacks of secondary task measures is that they have an impact on the performance of the primary task (Williges & Wierwille, 1979). A good overview of the pros and the cons of a secondary task as indicator for workload is given in O'Donnell and Eggermeier (1986). Another method frequently applied to assess workload is subjective assessment: participants report the workload experienced during the task post-hoc. The most frequently used examples of self-report measures assessing workload are the Subjective Workload Assessment Technique (SWAT) and the NASA Task Load index (TLX) (O'Donnell & Eggermeier, 1986). The NASA-TLX is a multidimensional subjective workload rating measure, in which the experience of workload is operationalized as an integration of weighted subjective responses (e.g., emotional, cognitive, physical) and a weighted evaluation of behavior. Although the NASA-TLX has been validated in the field of aircrew task performance and workload assessments, it has also been applied to other fields such as on-road evaluation of a car radio (Jordan & Johnson, 1993), and performance on a visual vigilance task in the presence of ambient noise (Becker, Warm, Dember & Hancock, 1995). Hancock (1996) stated that the NASA-TLX and the SWAT are essentially equivalent in their sensitivity to manipulations in tracking tasks. Advantages of using questionnaires in the assessment of workload is that they are easy to administer and can be employed in a variety of domains. Their limitations pertain to the fact that users may under- or overestimate their own performance. As information processing in the central nervous system may also affect other bodily processes, physiological indicators have been used as an alternative means of measuring workload. Relevant indicators of workload are absolute heart rate, heart rate variability, eye movements, brain waves, and skin conductance. The advantage of this category of measures is that the data can be recorded continuously throughout the experiment; however, data are often confounded by many other factors influencing physiology, a problem of non-specificity of the psychophysiological inference (Fairclough, 2009). A discussion of the various physiological measures and their utility can be found in O'Donnell and Eggermeier (1986).

2.3 Rationale for the studies

In the previous sections, we discussed literature concerning task performance using stereoscopic displays. From these studies we learned that task performance can be enhanced by adding pictorial depth cues, motion-based depth, and stereo. Although stereoscopic displays are becoming less expensive, the fact that stereoscopy requires the purchase of specific devices still presents a barrier to its use. Motion, which can be used on any display, is often seen as a depth cue providing the same information as stereo (Rogers & Graham, 1982). Yet, previous studies have shown that stereo is more effective in terms of reducing completion times, whereas motion makes task performance more accurate. What remains relatively unexplored is the effect of different disparity levels on task performance. In addition, task complexity may influence the effectiveness of motion-based depth cues, stereo, and their combination. These questions will be addressed in the three experiments described in the current chapter. Furthermore, when studying the performance benefits of various depth cues, previous studies only focused on completion times and error rates. In addition to these performance criteria, we also measure the subjective experience of workload in this chapter. This provides us with insights into the cognitive demands users experience while performing tasks using different depth cues. It also shows how this measure relates to objective performance criteria such as completion time and error rate.

In the first experiment, we explore the effects of different disparity levels on task performance. Earlier studies have shown that a disparity level between 10-30 min of arc produced optimal performance (De la Rosa et al., 2008; Rosenberg, 1993). This presents quite a broad range which makes it difficult to establish an optimal disparity based on these studies. It should be noted, however, that in one of these two studies, we had to estimate the disparity (in min of arc) based on the (limited) information available in the reports. These estimated disparity levels may be different from those actually presented to the participants in the studies. Therefore, the first experiment studies the effects of different disparity levels on users' task performance. Furthermore, since it is unclear how task difficulty level influences task performance, we included three difficulty levels.

The effectiveness of OM, stereo and a combination of these cues was investigated in the second study. Empirical studies suggest that OM enhances accuracy but slows completion times in comparison to stereo cues (Sollenberger & Milgram, 1993; Ware et al., 1993; Ware & Mitchell, 2008). The majority of studies employed uncontrolled OM, revealing shorter completion times but lower accuracy compared to controlled OM. Since we are interested in interactive 3D displays, we explore the effectiveness of controlled OM, stereo, and their combination in terms of completion times, accuracy, workload, and discomfort. In addition, we again explore whether difficulty level influences the effectiveness of these depth cues.

The third experiment includes both controlled object motion (OM) and movement parallax (MP), in combination with stereo cues. We are interested if more embodied interaction such as head movements (MP) would result in a better performance than the mouse (OM). To our knowledge, only one study has made a direct comparison between MP and OM (Ware et al., 1993). This particular study observed differences in accuracy between the two types of motion cues, although it depends on the visualization method applied. In addition, previous literature comparing MP or OM as additional depth cue in stereoscopic visualizations, has reported no performance difference while adding MP, whereas a combination between OM and stereo produced more accurate responses. The third experiment reported in this chapter therefore directly compared the effectiveness of OM and MP, alone and in combination with stereo, both in terms of objective indictors (i.e., completion time, accuracy) as well as subjective indicators (i.e., workload, discomfort). Furthermore, two difficulty levels were included to study whether the effects of stereo and motion were moderated by difficulty level.

2.4 Experiment 1: Disparity level and performance

In Experiment 1, we studied the relative contribution of binocular disparity for tasks of varying levels of difficulty in terms of task performance (completion time, accuracy) as well as workload and perceived discomfort.

2.4.1 Method

Design

This study followed a 3x5 within-subjects design with Difficulty (easy, moderate, and difficult) and Disparity (0, 5, 10, 25 and 50 min of arc) as independent factors. Within each combination of difficulty and disparity level, five tasks were administered resulting in a total of 75 tasks, offered in fifteen blocks. Task was entered as repeated factor in the analysis. Dependent variables were accuracy and completion time, measured for each individual task, and perceived workload and discomfort measured for each block of five tasks.

Participants

Thirty participants, 24 males and 6 females with a mean age of 23 (*SD* = 3.4), all with normal or corrected-to-normal vision took part. All participants had stereo vision better than 40 sec of arc, measured with the Randot® stereo-acuity test. Participants were either students or employees at the Eindhoven University of Technology in Eindhoven, the Netherlands.

Setting and apparatus

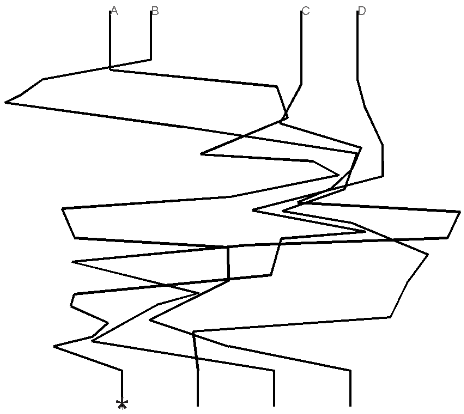
The experiment was carried out in the 3D/e lab of the Human-Technology Interaction group at Eindhoven University of Technology. The stimuli were displayed on a PLANAR SD2020 Stereo/3D monitor (20″ screen with a resolution of 1600x1200 pixels). This display consisted of two orthogonally placed polarized monitors and two half-silvered mirrors placed at a 45° angle to superimpose the left-eye and right-eye views onto each other. Subjects used polarized glasses, to separate the left and the right-eye views. The viewing distance was fixed at 73.5 cm in front of the screen, using a chin rest. Participants used a keyboard to provide their answers.

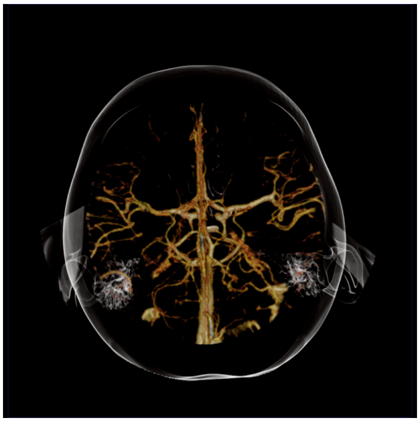
Stimulus generation

The task used in the current experiment is a path-tracing task, similar to the one used by Sollenberger and Milgram (1993). The task consisted of four lines randomly crossing each other (see Figure 5). Each line had the same number of line segments of the same length. The participant's task was to indicate which upper endpoint (a, b, c, or d) belonged to the line marked with an '*' at the lower end. With this task, we were able to test to what extent participants correctly perceived the spatial arrangement of the lines using different levels of stereo. As shown in Figure 5, the task serves as an abstract representation of an angiographic image, that - in line with the path-tracing task - also contains complex spatial structures. The difficulty level of the task was varied by changing the number of bends in all four lines.

A larger number of bends increased the number of line crossings and therefore increases the difficulty level of the task. The difficulty levels were selected based on a pilot study in which six difficulty levels were tested. Three difficulty levels were selected for the main experiment, consisting of 8 (easy), 14 (moderate), and 20 (difficult) line segments. In total 75 tasks were computed with Matlab and attributed to a specific condition.

It was assumed that with an increasing number of bends, the task became more difficult. However, we observed that several tasks were either easier or more difficult than intended. In some tasks, for example, line segments showed a large overlap, which rendered a task more difficult than one would predict based purely on the number of bends.





(a) (b)

Figure 5. *In the left panel an example of the wireframe stimuli used in this experiment (a), which serves as an abstract representation of complex spatial structures of blood vessels., such as the Circle of Willis located in the brain (b).*

Since for all participants the same set of tasks was assigned to a specific condition, this may have introduced a confound in the difficulty manipulation. We therefore assessed the actual difficulty of the 75 stimuli in a separate experiment. In this brief experiment, twenty participants between 19-30 years of age, all with normal or corrected-to-normal vision took part. In the study, all 75 stimuli were randomly presented on a standard 2D monitor and we recorded the answers as well as the associated completion times. We performed a repeated-measures ANOVA with Difficulty and Stimulus set (which in this case reflects the sets of tasks corresponding to one of the five disparity levels as used in Experiment 1) as predictors, and accuracy (percentage correct) and completion times as dependent variables. The analysis rendered the expected main effect of Difficulty, showing that with larger numbers of bends, accuracy was indeed lower [$F(2, 38) = 130.48$; $p < .001$, partial $\eta 2 = .87$] and completion time longer [$F(2, 38) = 68.87$; $p < .001$, partial $\eta 2 = .78$]. In addition, Stimuli set also showed a significant main effect on accuracy [$F(4, 76) = 17.14$; $p < .001$, partial $\eta 2 = .47$], indicating that the accuracy for the tasks in the stimulus set corresponding with a disparity of 5 was higher ($M = .82$, $SE = .02$; $p < .001$), and in the stimulus set corresponding with a disparity of 50 was lower ($M = .63$, $SE = .02$; $p < .001$) than for the stimulus sets of the other disparity levels (having an average of $M = .75$). Completion times were also significantly different between the Stimulus sets [$F(4, 76) = 4.01$; $p < .01$, partial $\eta 2 = .21$]. The set of tasks for a disparity of 5 ($M = 8.1$ $SE = 2.47$) revealed lower completion times than the set for a disparity level of 10 ($M = 9.6$, $SE = 2.47$; $p < .01$) and 50 ($M = 9.3$, $SE = 2.47$; $p < .05$).

Furthermore, the results showed a strong correlation between completion times and accuracy r = -.85, $p < .01$. Together, these results suggested that the difficulty of the sets of five tasks differed not only between the three difficulty levels, but also between the disparity levels. During the main analysis, we therefore corrected for these variations in difficulty by means of a covariate 'task complexity', reflecting the actual task complexity for each task based on this additional study (see the paragraph 'statistical analyses' for more detail how this covariate was used in the main analyses).

Measures

In the current study, we used both efficiency measures and subjective measures as indicators for performance. The efficiency measures used in this experiment are the time to complete a task (in seconds) and accuracy (percentage correct), assessed for each individual task. Workload was assessed for every block of five tasks using the NASA Task Load Index (NASA-TLX, Hart & Staveland, 1988). This questionnaire consists of six items (mental demand, physical demand, temporal demand, performance, effort, and frustration), all measured on 20-point scales ranging from (1) 'very low' to (20) 'very high'. At the end of the experiment, users indicated which items they thought contributed most to their perceived workload by means of 15 pair-wise comparisons. The resulting order of the items was used to weight the six items, and calculate the actual persons' workload score (for more details see Hart & Staveland, 1988). In addition to the workload questionnaire, we added a question concerning perceived discomfort while performing the task, asking: "Did you experience any discomfort during the performance of this task?" on a twenty-point scale ranging from (1) 'very low' to (20) 'very high'.

Procedure

Upon arrival at the 3D/e lab, participants were tested for their stereo acuity using the Randot® stereotest. Participants were then seated in front of the Planar display and received instructions explaining the procedure. Participants were instructed to follow the line, marked with a star (*), and indicate the corresponding endpoint by pressing one of four adjacent keys labeled 'a', 'b', 'c', and 'd'. During the experiment, participants rested their head on a chin rest to ensure the same viewing distance throughout the experiment. Participants were instructed to perform the tasks as fast and as accurately as possible. They performed five training trials to make sure they understood the procedure. The main experiment consisted of 15 blocks of tasks (5 disparity levels x 3 difficulty levels) with five tasks in each condition, resulting in a total of 75 tasks. Every condition had a specific set of tasks. The conditions and repetitions were randomized to counteract any learning effects. After each block of five tasks, participants filled in the NASA-TLX workload questionnaire and one additional question regarding perceived discomfort. The experiment took approximately 40 minutes and participants received 7.50 Euros for their participation.

Statistical analysis

Linear Mixed Model (LMM) analyses were performed to investigate the effect of Disparity and Difficulty on completion times, accuracy, workload and discomfort (separate analysis for each variable). Before the statistical analysis, within each participant we regarded completion times that departed from the mean with more than 3 SD as outliers and replaced this value with the completion time corresponding to the mean plus or minus 3 SD (1.7% of the data). The models used to analyze the results of completion time and accuracy differed slightly from those of workload and discomfort. In the analyses of completion time and accuracy, Participant was added as independent random intercept to group the data per participant to indicate that the same participant was measured multiple times. In addition, Repetition (order of the task) was added as a repeated random variable in the model, to indicate that in each condition five tasks were performed. In terms of workload and discomfort, Participant was again added as random variable, but we did not include Repetition as a repeated variable since workload and discomfort were measured at the end of each block of five tasks. For more details about LMM analyses, see e.g., Heck, Thomas and Tabata (2010).

In the analyses, we have two definitions of difficulty: one is based on the number of bends, and one is based on the results of additional study. The difficulty based on the number of bends is labeled 'Difficulty' and used as fixed factor in the analyses. The difficulty measured during the additional study is called 'Task complexity'. Note that more complex tasks have lower values, indicating a lower percentage of correct answers since both difficulty level and task complexity were highly correlated, we used the group mean centered score for task complexity in our model to avoid multicollinearity. This value represents the deviation of the complexity of each individual task from the average complexity score based on the number of bends (M_8 = .99, M_{14} = .74, M_{20} = .50)[4]. In other words, this variable specifies whether one task with a certain number of

[4] *These means are the average task complexities (in terms of percentage correct) per difficulty level based on the three difficulty levels used in this experiment*

bends is more or less complex than the average of all tasks with the same number of bends These centered values for task complexities, were added as covariates to the model to control for variations in task complexity between blocks of five tasks within one difficulty level. Since the centered task difficulty did not correlate ($r^2 = 0$) with difficulty level, both factors could be used in the analyses. Thus, we added Difficulty, Disparity, and the interaction between Disparity and Difficulty as fixed factors and task complexity as covariate to the model. In terms of workload and discomfort the centered task complexity did not have a significant impact on the model ($p = .49$ and $p = .70$ respectively) and therefore the analyses were performed without this variable as covariate.

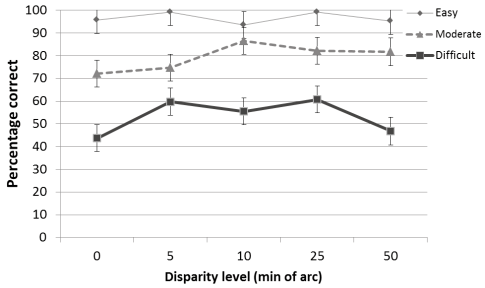
Please note that the data for completion times was positively skewed (>3.8), and therefore violated the assumption of homogeneity. Therefore we also ran the analysis with a Log10 transformed completion times (which did result in a normal distribution of the data) as dependent variable, to check whether this affected our results. This did not change the results of our study, and therefore we will report the data in terms of completion times. In the current experiment effect sizes are reported using Cohen's *d* which is calculated by: $d = M_1 - M_2 \,/ (\frac{SD_1 + SD_2}{2})$. For within-subject designs, statistics handbooks sometimes suggest a different calculation for Cohen's d, including the paired samples correlation. Following Cumming (2012), we will use Cohen's d as defined above, since this value can be used to compare both within and between effects of various studies. Throughout this thesis, we use Cohen's *d* to determine effect sizes found for main and post-hoc comparisons. A rule of thumb for interpreting these effect sizes suggests that values around .2 are interpreted as a small effect, values around .5 as a medium effect, and values exceeding .8 are interpreted as a large effect (Cohen, 1988).

2.4.2 Results

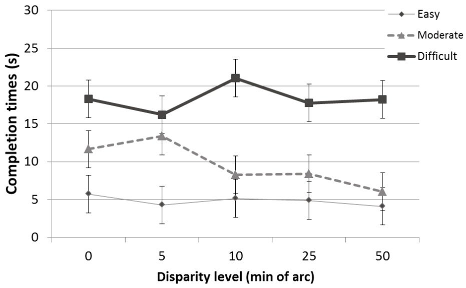
In this section, the effects of Disparity and Difficulty in terms of efficiency measures (completion time, accuracy) and subjective measures (workload and discomfort) will be reported.

Efficiency measures

Figure 6 shows the accuracy (a) and completion times (b) for the five disparity levels and three difficulty levels (estimated means for the main effects are reported in Table 3). The graph shows a different pattern of results depending on the difficulty level.



(a)



(b)

Figure 6. *Results in terms of efficiency measures with their 95% confidence intervals. The x-axis represent the five disparity levels, whereas the y-axis presents the estimated means in term of accuracy (a) and completion times (b). The three lines indicate the three difficulty levels used in this study. Results showed that for percentage correct (accuracy) both for medium and difficult tasks performance varied with disparity level, showing an optimum at approximately 10 min of arc (medium) and 25 (difficult). Completion times for moderately tasks decreased as disparity increased.*

The Linear Mixed Model (LMM) analysis with *accuracy* as the dependent variable rendered main effects for Difficulty and Disparity, as well as an interaction between Difficulty and Disparity. The main effect of Difficulty level [$F(2, 420) = 304.10$; $p < .001$] indicated that accuracy was lower for more difficult tasks (see Table 3), with effect sizes of $d = 1.64$ (easy vs. moderate), $d = 4.01$ (moderate vs. difficult) and $d = 4.01$ (easy vs. difficult). In addition results showed a main effect of Disparity [$F(4, 442) = 5.92$; $p < .001$]. Although the differences are small, Table 3 suggests an optimum in accuracy for disparity levels around 25 min of arc. Yet, the interaction between Difficulty and Disparity [$F(8, 430) = 3.36$; $p = .001$] further qualified this effect. Post-hoc comparisons with Bonferroni correction indicated that for the easiest task disparity did not impact performance (see Figure 6a). However, for the moderate and high difficulty levels, Disparity did affect accuracy. For tasks of moderate difficulty, a disparity of 10 ($M = .87$, $SE = .03$) rendered the highest accuracy, which was significantly different from 5 ($M = .75$, $SE = .03$; $p < .05$; $d = .81$) and 0 ($M = .72$, $SE = .03$; $p < .001$; $d = 1.02$), but not from disparities of 25 ($M = .82$, $SE = .03$; $p = 1$) and 50 min of arc ($M = .81$, $SE = .03$; $p = 1$). In tasks of high difficulty, a disparity level of 25 ($M = .87$, $SE = .03$) revealed the highest accuracy, which was significantly higher than at disparities 0 ($M = .44$, $SE = .03$; $p < .001$, $d = 2.92$) and 50 ($M = .47$, $SE = .03$; $p < .01$, $d = 2.72$), but not significantly different from disparities of 5 ($M = 0.60$, $SE = 0.03$) and 10 min of arc ($M = .56$, $SE = .03$; see Figure 6a). The centered

Task complexity also had a significant contribution to the model [F(1, 1209) = 484.22; p < .001], indicating that accuracy decreased as complexity increased (β = .79)[5].

Table 3: *Estimated means for the various difficulty levels and disparity levels in terms of accuracy, completion times, workload and discomfort*

Mean (SE)	*Disparity level (min of arc)*					*Difficulty level*		
	0	*5*	*10*	*25*	*50*	*easy*	*moderate*	*difficult*
Completion times (s)	11.9(1.06)	11.3(1.06)	11.5(1.06)	10.3(1.06)	9.46(1.07)	4.81(1.02)	9.56(1.02)	18.3(1.02)
Accuracy	.71(.02)	.78(.02)	.79(.02)	.81(.02)	.75(.02)	.97(.02)	.79(.02)	.53(.02)
Workload	9.71(.51)	9.98(.51)	9.23(.51)	8.90(.51)	10.1(.51)	6.44(.49)	9.27(.49)	12.4(.49)
Discomfort	6.40(.65)	6.35(.65)	6.84(.65)	7.30(.65)	9.81(.65)	5.30(.61)	7.27(.61)	9.44(.61)

The LMM analysis with *completion time* as dependent variable also showed significant main effects of Difficulty and Disparity, as well as a significant interaction. The main effect of Difficulty was significant [F(2, 420) = 332.20; p < .001], showing increasing completion times with increasing task difficulty (see Table 3), with effect sizes of d = .85 (easy vs. moderate), d = 1.56 (moderate vs. difficult) and d = 2.41 (easy vs. difficult). As shown in Figure 6b, the main effect of Disparity suggested a gradual decrease in completion times with increasing disparity levels [F(4, 428) = 4.05; p = .003]. However, the interaction between Difficulty and Disparity [F(8, 424) = 6.49; p < .001] indicated that there was no consistent pattern in completion times across tasks of varying difficulty. For the easy tasks, no difference was found between the different levels of disparity. But for tasks with moderate task difficulty, completion times decreased with increasing disparity, revealing the highest completion times for a disparity level of 5 (M = 13.41, SE = 1.27), which was significantly larger than 10 (M = 8.30, SE = 1.27; d = .82), 25 (M = 8.38, SE = 1.27; d = .8) and 50 min of arc (M = 6.07, SE = 1.28; d = 1.27) (all having a large effect), but not different from the 0 disparity level (M = 11.65, SE = 1.27). Quite unexpectedly, for difficult tasks completion times were similar between the different

[5] *Note that task complexity was expressed in terms of percentage correct measured during the additional study, suggesting that higher scores correspond to less complex tasks.*

disparities except for a disparity level of 10 min of arc (*M* = 21.04; SD = 1.27), at which completion times were higher than for a disparity of 5 (*M* = 16.21; SD = 1.27; $p = .001$; $d = .78$) and marginally higher than a disparity of 25 (*M* = 17.7; SD = 1.27;$p = .06$; $d = .53$) (see Figure 6b). Centered task complexity also had a significant effect on completion time [$F(1, 1187) = 253.83$; $p < .001$] showing increasing completion times with a higher task complexity (β = 11.23).

Subjective measures

Figure 7 presents the scores for perceived workload and perceived discomfort as a function of the five disparity levels and three difficulty levels. As in the earlier analyses, the LMM analysis with *workload* as dependent variable rendered significant main and interaction effects of Difficulty and Disparity. Difficulty revealed a significant main effect [$F(2, 406) = 282.32$; $p < .001$], showing that perceived workload increased with increasing task difficulty (see Table 3), with effect sizes of $d = 1.05$ (easy vs. moderate), $d = 1.17$ (moderate vs. difficult) and $d = 2.22$ (easy vs. difficult). The main effect of Disparity was significant [$F(4, 406) = 4{,}65$; $p < .001$], but the interaction with Difficulty [$F(8, 406) = 3{,}81$; $p < .001$] qualified this effect. Post-hoc tests with Bonferroni correction indicated that in the easy tasks no significant difference in workload emerged between the disparity levels. For both moderate and difficult tasks significant differences in workload emerged, although the effect sizes were smaller than found for completion times and percentage correct For moderate task difficulty a disparity of 25 min of arc (*M* = 8.81, *SE* = .60) resulted in the lowest workload. This was significantly lower than workload at disparity of 5 min of arc (*M* = 9.90, *SE* = .60; $p < .05$, $d = .37$) and showed a non-significant trend for a disparity of 0 min of arc (*M* = 9.70, *SE* = .60; $p = .07$, $d = .30$). However, a disparity of 25 min of arc did not reveal lower levels of workload compared to disparities of 10 and 50 min of arc. For the difficult tasks a mixed picture emerged, with even smaller effect sizes, in which the lowest workload appeared at disparities 5 (*M* = 11.35, *SE* = .60) and 25 min of arc (*M* = 11.43, *SE* = .60), being only significantly different from workload at a disparity level of 50 min of arc (*M* = 13.93, *SE* = .60; both $p < .001$, d < .17).

Difficulty had also a significant main effect on Perceived discomfort, indicating higher discomfort with increasing difficulty levels [$F(4, 406) = 58.08$; $p < .001$] (see Table 3), with effect sizes of $d = .59$ (easy vs. moderate), $d = .65$ (moderate vs. difficult) and $d = 1.24$ (easy vs. difficult). In addition, as shown in Table 3, Disparity had a main effect [$F(4, 406) = 16.66$; $p < .001$], showing that perceived discomfort at a disparity of 50 min of arc was significantly larger than that at the other disparity levels (all $p < .001$; $d > .78$).

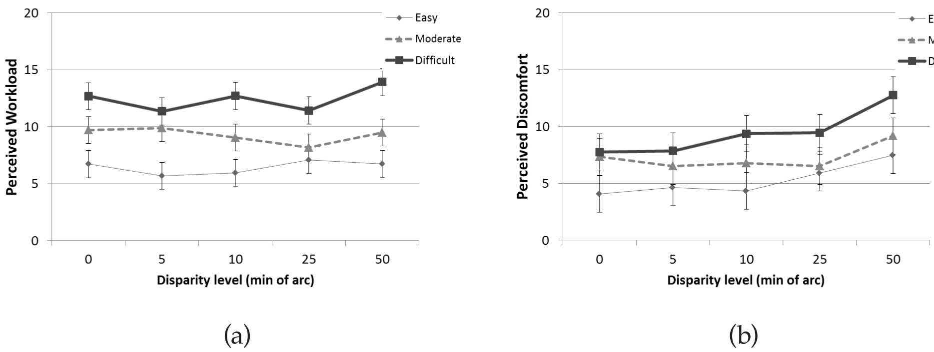


Figure 7. *Results on subjective measures with their 95% confidence intervals. The x-as represent the 5 disparity levels, whereas on the y-axis the estimated means in term of perceived workload (a) and perceived discomfort (b) are given. The three lines are the three difficulty levels used in this study.*

No interaction effect emerged between Disparity and Difficulty [$F(8, 406) = 1.58$; $p = .13$] suggesting that the effect of disparity on perceived discomfort was not moderated by difficulty level (see Figure 7b).

2.4.3 Summary of results

The results of Experiment 1 showed that performance is dependent on both difficulty and disparity level. The effect of disparity was moderated by difficulty, showing that stereo improved performance more for the moderate and difficult tasks, than for the easy tasks. Earlier work (De la Rosa et al., 2008; Fishman et al., 2008; Rosenberg, 1993) demonstrated a performance increase with increasing disparity, up to a certain threshold level after which performance did not improve any further. However, the current study suggested that these effects differ with task difficulty. The easiest tasks showed no sensitivity to disparity level, probably due to a ceiling (accuracy) and floor (completion time) effect. Tasks with moderate difficulty levels revealed a similar pattern as found in previous literature (De la Rosa et al., 2008; Fishman et al., 2008; Rosenberg, 1993), showing a gradual increase in accuracy and a decline in completion times with disparity until 10-25 min of arc, after which these performance indicators remained constant. Cohen's d effect sizes showed that the effect of disparity on both completion times and accuracy was large ($d > .80$). For difficult tasks the effects size of disparity on accuracy was even larger ($d \approx 2.70$), revealing an optimum in accuracy in the midrange of disparities (i.e., between 5 and 25 min of arc), and a lower performance for disparities both below and above this range. The

Chapter 2

effect sizes for completion times were smaller, and showed a less clear pattern for the difficult tasks (*d* between .55 - .78). However, for tasks with moderate difficulty level effect sizes were larger (*d* between .8 – 1.27), and showed lower completion times when disparity level increases.

Perceived workload showed a similar response pattern, with low sensitivity for disparity during the easy tasks. An inverse relation between workload and disparity for moderate tasks was found, balancing out for the highest disparity levels. For the difficult tasks a U-shape relationship was found, with an optimum around 10-25 min of arc. This illustrates that perceived workload is also sensitive to variations in task difficulty as well as disparity level. Nevertheless, the relatively small effect sizes ($d < .37$) found for the effect of disparity level on perceived workload suggest that more research is needed before we can draw any conclusions. In contrast, visual comfort showed no moderation of disparity effects by task difficulty. However, participants consistently perceived higher levels of discomfort for a disparity of 50 min of arc ($d > .78$).

In this experiment, we showed that stereo has an effect on completion times, accuracy, workload and discomfort. In the current experiment, users did not have the ability to interact with the content. However, in many applications in which understanding spatial structures is vital, users are able to manipulate the content (e.g., rotating, zooming). In the second experiment we are therefore interested in how object-motion (i.e., rotating an object with the mouse) and stereo facilitate task performance using a similar task as in Experiment 1.

2.5 Experiment 2: Effectiveness of object motion and stereo for easy and difficult tasks

In this second experiment, we studied the effectiveness of object motion (OM) and stereo on task performance in terms of completion time and accuracy as well as on perceived workload and perceived discomfort.

2.5.1 Method

Design

The study followed a 2x2x2 repeated-measures design, with Difficulty (easy vs. difficult), Visualization method (mono vs. stereo) and Motion (static vs. OM) as independent factors. The dependent variables are completion time, accuracy, perceived workload, and perceived discomfort. The order of the experimental conditions was randomized for each participant. Each condition consisted of five unique tasks randomly selected from 20 tasks per difficulty level, which were generated with Matlab (see stimuli section).

Participants

Twenty participants took part in this experiment (11 male, 9 female), all with normal or corrected-to-normal vision. All participants had stereo vision better than 60 seconds of arc, tested with the Randot® stereotest. Participants were either students or employees at the Eindhoven University of Technology, the Netherlands.

Stimuli

The type of task used in this experiment, i.e., the path-tracing task, was the same as discussed in Experiment 1. However, for this experiment new sets of lines were computed due to a different experimental set up. Since the previous experiment showed that difficulty increased with the number of line segments in the task, we again used this criterion for manipulating task difficulty. However, in this experiment, the tasks were randomly distributed over the eight experimental conditions for each participant. The difficulty levels were selected based on a pilot study in which six difficulty levels were tested. Since users were now able to rotate the task, the pilot showed that the difficulty levels should be higher than the ones used in Experiment 1. The difficulty levels selected for this experiment contained 20 (Difficulty: easy) and 26 (Difficulty: difficult) line segments. In total 40 unique tasks were computed with Matlab. The files used in this experiment were in voxel format; files were read by volume rendering software6 able to display voxel files on a stereoscopic display. The disparity level used for stereo was 30 min of arc.

Setting and apparatus

The experiment was carried out at the 3D/e lab of the Human-Technology Interaction group at Eindhoven University of Technology. The stimuli were displayed on a Heinrich Hertz Free2C autostereoscopic 3D Display, which was 21.3 inches and used in portrait format. The resolution of the display was 1200 x 1600 pixels. The stereoview on this display is created using a moving lenticular which steers the exit pupils to the user's current eye position. The eye position was determined with a stereo video

[6] *This software was developed within the European Funded FP7 HELIUM3D project, used to display volumetric images on a stereoscopic display*

head-tracking device mounted on top of the display. In contrast with Experiment 1, no chinrest was used in Experiment 2. Participants were seated in front of the display at approximately 65 cm. For OM the participants used a mouse to rotate the object; rotation was fixed to the vertical axis only. In the static condition users were not able to rotate the object.

Measures

Similar to Experiment 1, the dependent variables measured for each individual task were accuracy (percentage correct) and completion time (in seconds). At the end of each block both perceived workload and perceived discomfort were measured. Perceived workload was measured using the NASA Task Load Index (Hart & Staveland, 1988). Additionally, visual discomfort was addressed with the question: "Did you experience any discomfort performing this task?" on a twenty-point scale ranging from (1) 'very low' to (20) 'very high'.

Procedure

On arrival at the 3D/e lab, participants were tested for their stereo acuity using the Randot® stereotest. When participants completed the test with a score of at least 60 arc seconds, they were seated in front of the display and received written instructions explaining the procedure. Participants were instructed to perform the task as fast and accurately as possible. Before the start of the experiment, participants performed four training tasks to make sure they understood the procedure. The experiment consisted of eight blocks with five trials, and after each block participants completed the NASA-TLX workload questionnaire and the question regarding visual discomfort. The experiment lasted around 30 minutes and students received a compensation of 5 Euro for their participation.

Statistical Analysis

A repeated-measures ANOVA was performed to investigate the effect of Difficulty (easy vs. difficult), Visualization method (mono vs. stereo), and Motion (static vs. OM) on completion times, accuracy, perceived workload and discomfort (separate analyses for each variable). In addition to the main effects, all 2-way and 3-way interactions were added to the model. Before the statistical analysis, within each participant we regarded completion times exceeding ± 3 SD as outlier and replaced this value with the completion time corresponding with the mean plus or minus 3 SD (1.6% of the data). Please note that the data for completion times is positively skewed (skewness = 1.42), and therefore violating the assumption of homogeneity. Therefore, we also run the analysis with a Log_{10} transformed completion times (which did result in a normal distribution of the data), to check whether this affected our results. This did not change the results of our study, and therefore we will report the data in terms of completion times. For accuracy, we first calculated the percentage of correct responses for each of the 8 conditions

In the current experiment, two effect size measures will be reported. For the interpretation of main and interaction effects, partial η2 is reported, which is a measure of variance, like R-squared. It indicates the variance explained by the dependent variable, excluding the variance of other variables in the experimental design. Therefore, the sum of the effect sizes for all dependent variables in one study often exceeds one. As a rule of thumb we use the interpretation used for R squared, where .10 represents a small effect, .30 a medium effect, and .50 a large effect. In addition to partial η2, we will report Cohen's *d* for main effects and post-hoc comparisons using the formula given in paragraph 2.4.2. This measure of effect size was used to compare effect sizes within and between the different experiments discussed in this thesis

2.5.2 Results

First, the results of accuracy and completion times will be discussed, followed by the results of perceived workload and perceived discomfort.

Efficiency measures

The repeated-measures ANOVA with accuracy as dependent variable rendered a significant main effect of Difficulty [$F(1,19) = 42.1$; $p < .001$, partial η2 = .69], showing higher accuracy for easier tasks ($M = .64$, $SE = .03$) compared to the difficult tasks ($M = .48$, $SE = .04$; $d = 1.02$). Although Figure 8 shows that accuracy is slightly higher in the stereo condition, results did not reveal a difference between mono and stereo visualizations [$F(1,19) = 3.45$; $p = .08$, partial η2 = .15]. However, as shown in Figure 8, the number of correct responses was higher (with a large effect size) for conditions where participants could rotate the object using OM ($M = .72$, $SE = .05$) compared to static images ($M = .41$, $SE = .03$; $d = 1.73$), with [$F(1,19) = 55.20$; $p = <0.001$, partial η2 = .74]. Figure 9a suggests that the effect of Motion is more pronounced for the difficult tasks, which was confirmed by the significant interaction between Difficulty and Motion [$F(1,19) = 5.21$; $p = .03$, partial η2 = .22]. A post hoc test with

Bonferroni correction revealed that for the difficult tasks, the difference between OM (M = .76, SE = .05) and static images (M = .52, SE = .03; d = 1.88) was slightly larger than the difference between OM (M = .67, SE = .06) and static images (M = .29, SE = .03; d = 1.34) for the easy tasks.

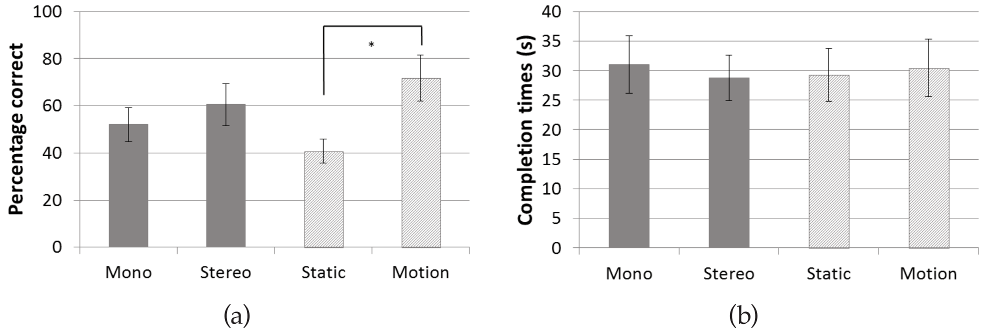


Figure 8. *Main effects of Experiment 2 in terms of percentage correct (a) and completion times (b) with their 95% confidence intervals. This figure represents the main effect of Mono vs. Stereo and Static vs. Motion (OM). Significant difference are indicated with a * (p<0.05). Results showed that Motion significantly increased accuracy (a), without a change in completion times (b). No difference emerged between stereo and mono visualizations in terms of accuracy (a) and completion times (b).*

The repeated-measures ANOVA with completion times as a dependent variable indicated a main effect of Difficulty [F(1,19) = 97.1; p < .001, partial η2 = .84], showing longer completion times for the difficult tasks (M = 35.7, SE = 2.7) compared to the easy tasks (M = 24.10, SE = 1.70; d = 1.17). As illustrated in Figure 8, Motion itself did not have an effect on completion times [F(1,19) = .43; p = .5, partial η2 = .02]. Although Figure 8 shows a small decrease in completion times when using stereo, this results was not statistically significant [F(1,19) = 4.0; p = .06, partial η2 = .17]. The significant interaction between Visualization method and Motion [F(1,19) = 6.98; p < .01, partial η2 = .27], showed that stereo only speeded up task performance in conditions with motion (stereo: M = 27.80, SE = 1.80; mono: M = 33.20, SE = 3.30; p < .05; d = .47). No difference was found between mono (M = 28.90, SE = 2.28) and stereo visualization (M = 29.70, SE = 2.40; p = .50) in the static condition (see Figure 9b).Figure 9b further suggests that this pattern only existed for difficult task, but the three-way interaction was not significant [F(1,19) = .60; p = .45, partial η2 = .31].

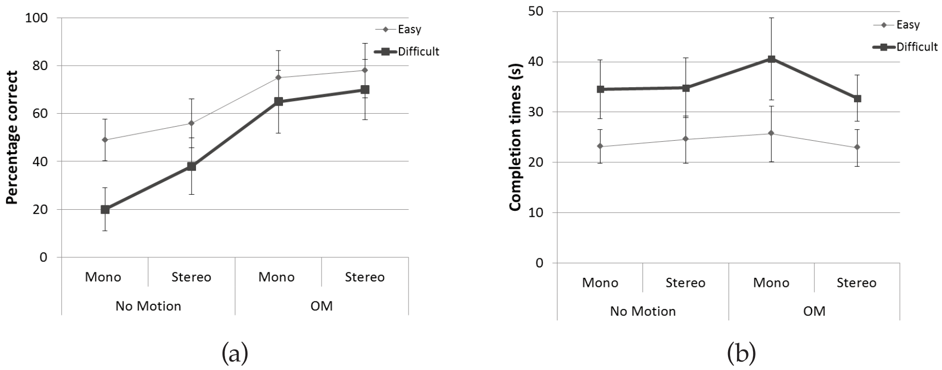
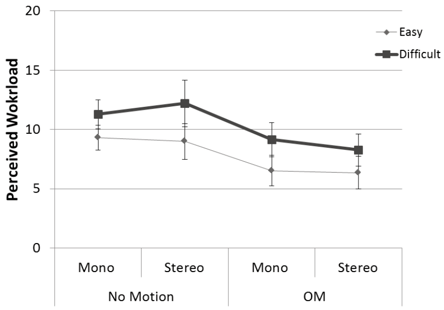


Figure 9. *Results on efficiency measures with their 95% confidence intervals. The x-axis represents the four conditions of the experiment, whereas on the y-axis the mean scores in term of accuracy (a) and completion times (b) are given. The two lines are the two difficulty levels used in this study. This figure illustrate the effect of difficulty on both accuracy and completion times. In addition, it shows that motion increases percentage of correct answers more than stereo. However, a combination between stereo and motion decreased completion times compared to motion without stereo.*

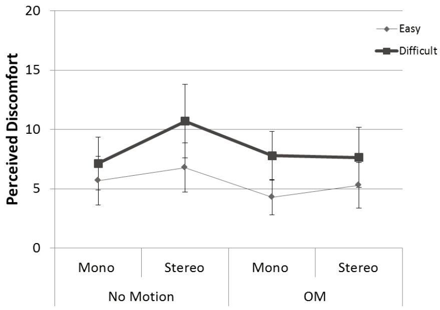
Subjective measures

The results of the repeated-measures ANOVA with workload as dependent variable showed a significant main effect of Difficulty [$F(1,19) = 33.60$, $p < .001$, partial η2 = .64]. Difficult tasks induced higher levels of perceived workload ($M = 10.20$, $SE = .49$) compared to easy tasks ($M = 7.80$, $SE = .53$; $d = .97$). As shown in Figure 10, perceived workload decreased when participants were able to rotate the object using OM ($M = 7.58$, $SE = .59$) compared to conditions without motion ($M = 10.40$, $SE = .60$; F(1,19) = 42.0; $p < .001$, $d = 1.06$, partial η2 = .69).In addition, no difference emerged between mono ($M = 9.07$, $SE = .51$) and stereo ($M = 8.96$, $SE = .67$) visualizations [$F < 1$; ns] or any interaction between Visualization method, Motion, and Difficulty in terms of workload (all p > .10). Figure 10b shows the results in terms of perceived discomfort. The repeated-measures ANOVA showed a significant main effect of Difficulty, Visualization method, Motion and an interaction between Motion and Visualization method. As shown in Figure 10, perceived discomfort was lower for easy tasks ($M = 5.53$, $SE = .85$) than for tasks with a high difficulty level ($M = 8.33$, $SE = 1.10$; $d = .64$) with [$F(1,19) = 15.0$; $p = .001$; partial η2 = .44]. The main effect of Visualization method [$F(1,19) = 7.29$; $p < .05$, partial η2 = .28] indicated higher levels of discomfort in the stereo ($M = 7.61$, $SE = 1.0$) than in the mono condition ($M = 6.24$, $SE = .88$; $d = .32$). Object motion (OM) on the other hand decreased perceived discomfort ($M = 6.26$, $SE = .91$) compared to the static condition

(M = 7.60, *SE* = 1.0; *d* = .31) with [F(1,19) = 4.96; $p < .05$; partial η2 = .21]. The interaction between Motion and Visualization method was also significant [$F(1,19) = 6.23$; $p < .05$, partial η2 = .25]. Post-hoc tests with Bonferroni correction indicated that for the static condition, stereo ($M = 8.70$, $SE = 1.13$) increased perceived discomfort compared to a monoscopic visualization ($M = 6.40$, $SE = 1.0$; $p < .01$; $d = .48$). During the condition with OM, no difference emerged between stereo ($M = 6.48$, $SE = 1.0$) and mono visualizations ($M = 6.05$, $SE = .80$; $p = .30$). Figure 10b suggest that this effect was more pronounced for the difficult tasks, however the three way interaction between Motion, Visualization method, and Difficulty only showed a non-significant trend [$F(1,19) = 4.0$; $p = .06$, partial η2 = .18].



(a)



(b)

Figure 10. *Results in terms of perceived workload (a) and perceived discomfort (b) with error bars representing the 95% confidence intervals. The x-axis represents the four conditions of the experiment whereas on the y-axis the mean scores in term of perceived workload (a) and perceived discomfort (b) are given. The two lines are the two difficulty levels used in this study. Results in terms of workload showed that Motion (both OM and MP) decreased workload compared to the condition without motion. In terms of discomfort, the results showed that without motion stereo increased discomfort, whereas with motion the increase in discomfort was smaller.*

2.5.3 *Summary of results*

The results presented in the second study showed a strong increase in accuracy ($d = 1.73$) when using OM. Results showed that participants answered only 41% correct in conditions without motion and 72% correct in conditions with motion. Also stereo increased accuracy (from 52% for mono to 61% for stereo), however this effect was not significant. Combining motion and stereo did not reveal an additional increase in accuracy compared to OM without stereo. This result is not in line with findings in previous literature concerning controlled OM (Hubona et al., 1997; Sollenberger & Milgram, 1993; Ware & Mitchell, 2008), but is in line with studies using MP (Faubert, 2001; Naepflin & Menozzi, 2001). The difference in tasks (Hubona et al., 1997) or the time limit employed by Sollenberger and Milgram (1993), might explain the different findings. Nevertheless, in terms of completion times, the result of the current experiment is more in line with previous literature (Hubona et al., 1997; Naepflin & Menozzi, 2001 Ware & Mitchell, 2008). Motion did not improve completion times, yet a combination of motion and stereo decreased completion times with approximately eight seconds, resulting in a moderately strong effect size (d = .47). In terms of workload, motion had a large effect (d = 1.06), showing lower levels of workload when users were able to rotate the stimuli. Performing tasks in stereo did not reveal a difference in perceived workload. In addition, object motion reduced perceived discomfort, while stereo did increase discomfort, but only in conditions without motion.

Overall, this study showed that OM has the largest effect on task performance, since it enhanced accuracy, and reduced perceived workload and perceived discomfort. Combining stereo and motion reduced completion times compared to motion without stereo, and did not result in more discomfort. Another potential depth cue that can be implemented on displays is movement parallax, using the user's head position to change perspective on the screen. The effectiveness of OM and MP will be compared in Experiment 3.

2.6 Experiment 3: Effectiveness of Motion based depth cues and stereo

In Experiment 2, we showed that OM increased accuracy compared to the static conditions and decreased participants perceived workload. In this previous experiment, participants rotated the stimulus with the mouse, the current standard for computer interaction. Another way to interact with such volumes is the use of movement parallax (MP), where the perspective of the object is changed according to the position of the user's head. As discussed in the introduction of this chapter, only one study performed a direct comparison of OM and MP (Ware et al., 1993), whereas other studies used either OM or MP. In the current study, we therefore explore whether controlling an object using our head or a mouse elicits performance differences, measured with both subjective and objective indicators. In Experiment 2, we found that adding stereo did not significantly increased percentage correct, but decreased completion times in the motion condition. The relatively high levels of discomfort associated with stereo might have influenced task performance. Therefore, in the current experiment, we decided to reduce the amount of disparity from 30 min of arc to 10 min of arc. This level of disparity should reduce the visual complaints and still lead to increased performance compared to the non-stereo conditions (see Experiment 1). In addition we slightly reduced the difficulty level of the most difficult tasks used in Experiment 2, from 26 to 24 line segments, since users only scored at chance level in the static conditions with 26 line segments. In sum, Experiment 3 serves to both extend (by adding MP) and replicate findings of Experiment 2, studying the effect of motion, stereo and difficulty level on completion times, percentage correct, perceived workload and perceived discomfort.

2.6.1 Method

Design

The study followed a 2x2x3 repeated-measure design, with Difficulty (easy vs. difficult), Visualization method (mono vs. stereo) and Motion (static vs. OM vs. MP) as independent factors. The order of the experimental conditions was randomized over participants. Each condition consisted of four unique tasks randomly selected from 24 tasks per difficulty level, which were generated with Matlab (see Stimuli section).

Participants

Twenty participants took part in this experiment (14 male, 6 female), all with normal to corrected-to-normal vision. All participants had stereo vision better than 40 seconds of arc, tested with the Randot® stereotest. Participants were either students or employees at the Eindhoven University of Technology, the Netherlands.

Stimuli

The tasks used in this experiment were similar to those used in Experiments 1 and 2, although new sets of lines were computed for this experiment. The software displaying the images was modified such that it displayed lines instead of voxels, which increased the quality of the lines. This experiment consisted of two difficulty levels containing 20 (Difficulty: easy) and 24 (Difficulty: difficult) line segments. In total 48 unique tasks were computed with Matlab, which were randomly distributed over the 12 experimental conditions for each participant. The maximum disparity used in this experiment was 10 min of arc. In the OM condition, the participants used a mouse to rotate the object. For MP, the orientation of the object was calculated according to the position of the user's head. For both OM and MP the rotation was limited to the vertical axis only.

Setting and apparatus

The setting of the experiment was identical to that described in Experiment 2. To present the stimuli, the Heinrich Hertz Free2C autostereoscopic 3D Display was again used (see Experiment 2). The eye positions retrieved from the video head-tracking device were used to calculate the appropriate view for to the current head position.

Measures

In this experiment, we used the same dependent variables as used in Experiments 1 and 2; i.e., completion times (in seconds), accuracy (percentage correct), perceived workload and perceived discomfort. Completion times and accuracy were recorded for every task, perceived workload and discomfort were scored after each block of five tasks. Perceived workload was again measured using the NASA Task Load Index (Hart & Staveland, 1988), and visual discomfort was measured with the question: "Did you experience any visual complaints." on a twenty-point scale ranging from (1) 'very low' to (20) 'very high'.

Procedure

Upon arrival at the 3D/e lab, participants were tested for their stereo acuity using the Randot® stereotest. Participants with an acuity better than 40 min of arc took part in the experiment. Participants were seated in front of the display and received instructions explaining the procedure. Participants were instructed to perform the task as rapidly and accurately as possible. Before the start of the experiment, participants performed four training tasks to make sure they understood the procedure. The experiment consisted of 12 blocks, of four tasks, and after each block participants filled in the NASA-TLX workload questionnaire, and the question regarding visual discomfort. The experiment took approximately 40 minutes and users received a compensation of 7.50 Euros for their participation.

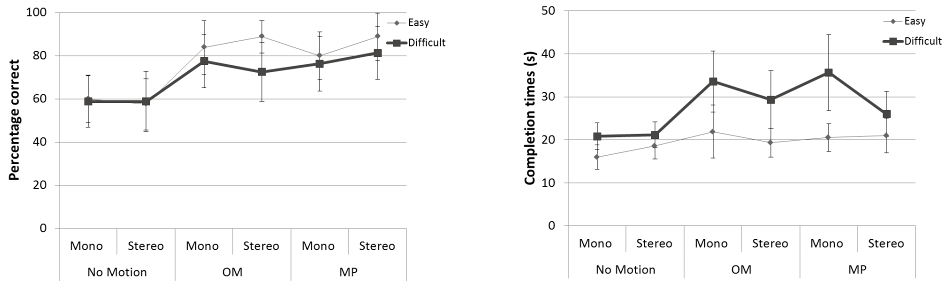
Statistical analysis

Comparable to Experiment 2 we used repeated-measures ANOVAs to test effects of Motion, Visualization, Difficulty, including the 2-way interaction and 3-way interaction between Motion, Visualization, and Difficulty. In contrast to Experiment 2, three levels of Motion were tested (static, OM, and MP) in the current study. Before the statistical analysis, within each participant we regarded completion times exceeding ± 3 SD as outlier and replaced this value with the completion time corresponding with the mean plus or minus 3 SD (1.8% of the data). The data for completion times was positively skewed, and therefore violating the assumption of homogeneity. Therefore, we also run the analysis with a Log_{10} transformed completion time (which did result in a normal distribution of the data), to check whether this affected our results. Generally, the results of the transformed data did not change the results and therefore we will report the data in terms of the original completion times. In cases of differences between these two analyses, we report this in the text. For accuracy, we first calculated the percentage of correct responses for each of the 12 conditions. Under the Motion conditions, the assumption of sphericity was violated regarding the results of workload, accuracy, and completion times. To correct for this, the Greenhouse-Geisser correction was applied. In the current experiment, effect sizes were again reported in terms of partial η^2 and Cohen's *d*, as previously discussed in paragraph 2.5.1.

2.6.2 Results

Efficiency measures

As shown in Figure 11a, the repeated-measures ANOVA with accuracy as dependent variable revealed a significant main effect of Difficulty [$F(1,19) = 6.27$; $p < .05$, partial $\eta2 = .25$], indicating that users made more errors during the difficult tasks ($M = .71$, $SE = .05$) than the easy tasks ($M = .77$, $SE = .04$; $d = .29$). As shown in Figure 12a, Visualization mode did not reveal a significant main effect, so stereo images produced accuracies similar to those of mono images ($F < 1$, *ns*).

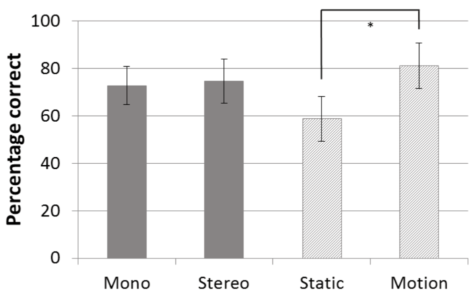


(a) (b)

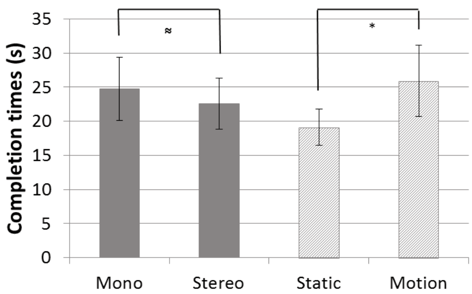
Figure 11. *Results in terms of percentage correct (a) and completion times (b) with their 95% confidence intervals. The x-axis represent the six conditions of the experiment whereas on the y-axis the mean scores in term of accuracy (a) and completion times (b) are given. The two lines are the two difficulty levels used in Experiment 3. This figure illustrate the effect of difficulty on both accuracy and completion times. In addition, it shows that the effects of motion and stereo are more pronounced for the more difficult tasks.*

However, also illustrated in Figure 12a, Motion significantly increased the percentage of correct answers [$F(1.37,25.98) = 19.14$; $p < .001$, partial $\eta 2 = .50$]. Post-hoc comparisons with Bonferroni correction for the three levels of Motion showed that both OM ($M = .81$, $SE = .05$) and MP ($M = .82$, $SE = .05$) significantly increased accuracy compared to the static condition ($M = .59$, $SE = .05$; both $p < .01$, $d = .98$ and $d = 1.02$ respectively). No difference was found between OM and MP (p = 1). The analysis further revealed no interaction between Difficulty and Visualization method (F < 1, *ns*), Difficulty and Motion [$F(1,19) = 2.56$; $p = .1$, partial $\eta 2 = .12$], Motion and Visualization method [$F(1,19) = 1.20$; $p = .30$, partial $\eta 2 = .06$], nor a three-way interaction between Difficulty, Visualization method and Motion (F < 1, *ns*).

The repeated-measures ANOVA with completion time as dependent variable indicated a main effect of Difficulty [$F(1,19) = 66.20$; $p < .001$, partial $\eta 2 = .77$]; completion times were longer for the difficult tasks ($M = 27.80$, $SE = 2.60$) compared to the easy tasks ($M = 19.50$, $SE = 1.70$; $d = .86$). In addition, main effects emerged for Visualization method and for Motion, as well as interactions between Visualization method and Motion, Visualization method and Difficulty, and between Motion and Difficulty. As shown in Figure 12b, Motion significantly increased completion times [$F(1.2, 22.8) = 11.95$; $p < .001$, partial $\eta 2 = .39$]. Post hoc tests with Bonferroni correction showed that participants used more time to complete the tasks using both OM ($M = 26.0$, $SE = 2.80$) and MP ($M = 25.8$, $SE = 2.5$) than in the static conditions ($M = 19.10$, $SE = 1.30$; both $p < .01$; $d = .75$ and $d = .79$ respectively). No statistical difference was found between OM and MP (p = 1).



(a)



(b)

Figure 12. *Main effects of Experiment 3 in terms of percentage correct (a) and completion times (b) with their 95% confidence intervals. This figure represents the main effect of mono vs. stereo and static vs. motion (both MP & OM). Significant difference (p < .05) are indicated with a * and ≈ presents a non-significant trend (p < .10). Results showed that Motion significantly increased accuracy (a), yet at the same time completion times also increased (b). Stereo did not demonstrate such a speed-accuracy trade-off, with a non-significant trend towards faster completion times, without a drop in accuracy.*

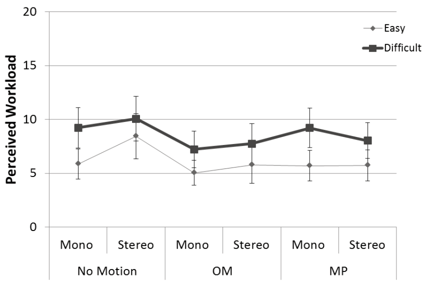
Visualization method showed a non-significant trend, suggesting that stereo decreased completion times[7] [$F(1,19) = 6.34$; $p < .05$, $d = .22$ partial $\eta 2 = .25$]. Figure 11b shows that stereo only decreased completion times in the motion conditions, which is in line with the significant interaction between Motion and Visualization method [$F(1,19) = 13.19$; $p = .001$, partial $\eta 2 = .41$]. Post hoc tests with Bonferroni correction revealed that, similar to the results in Experiment 2, stereo significantly decreased completion times in conditions with OM (ΔM= -3.4, $SE = 1.30$; $p < .01$; $d = .27$) and MP ($\Delta M = -4.60$, $SE = 1.20$; $p < .01$, d = .39), but only showed a non-significant trend towards longer completion times for stereo in the static condition ($\Delta M = 1.50$, $SE = .70$; $p = .05$, $d = .23$). Figure 11b suggests that the effect of stereo on completion times emerged particularly for the difficult tasks, which is confirmed by the interaction between Difficulty and Visualization method [$F(1,19) = 12.2$; $p < .01$, partial $\eta 2 = .39$]. A post-hoc test with Bonferroni correction revealed that for the difficult tasks, stereo ($M = 25.50$, $SE = 2.34$) decreased completion times compared to a monoscopic presentation ($M = 30.0$, $SE = 2.90$; $p < .01$, $d = .38$), whereas for the easy tasks no difference existed in completion times between stereo ($M = 19.60$, $SE = 1.55$) and monoscopic visualization ($M = 19.4$, $SE = 1.89$; $p = .90$). In addition, Figure 11b also shows that the increase in completion times when using OM or MP only exists for the difficult tasks. The significant interaction between Difficulty and Motion [$F(2,38) = 6.56$; $p < .01$, partial $\eta 2 = .26$] confirms this observation. Post-hoc test with Bonferroni correction revealed that for easy tasks the static condition ($M = 17.30$, $SE = 1.34$) produced shorter completion times than MP ($M = 20.8$, $SE = 1.77$; $p < .05$, $d = .50$), but not significantly shorter than OM ($M = 20.60$, $SE = 2.30$; $p = .10$, $d = .41$). For difficult tasks, this effect was much larger and visible for both OM and MP: completion times in the static condition ($M = 20.98$, $SE = 1.46$) were shorter than those with MP ($M = 30.80$, $SE = 3.47$; $p < .01$, d = .89) and OM ($M = 31.50$, $SE = 3.30$; $p < .01$, d = .99). The results did not reveal a significant 3-way interaction between Difficulty, Motion and Visualization method [$F(2,38) = 2.06$; $p = .14$, partial $\eta 2 = .10$].

Subjective measures

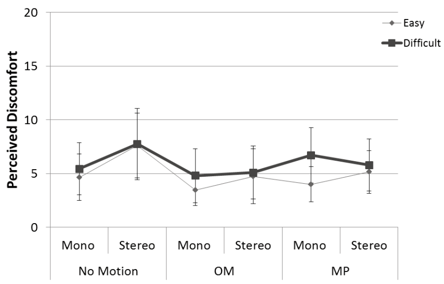
The repeated-measures ANOVA with workload as dependent variable showed a significant main effect of Difficulty [$F(1,18) = 69.20$, $p < .001$, partial $\eta 2 = .79$], revealing higher workload for the difficult tasks ($M = 8.60$, $SE = .49$) compared to the easy tasks ($M = 6.10$, $SE = .53$; $d = 1.09$) . As shown in Figure 13a the main effect of Motion [F(1.48,26.6)

[7] *Analyzing this with the Log10 transformed completion times was not significant ($F(1,19) = 4.28$; $p = 0.05$, $d = .12$, partial $\eta 2 = .18$), and therefore this effect should be interpreted as such.*

= 6.23; $p < .05$, partial η2 = .26] indicated that both OM and MP significantly decreased perceived workload compared to static conditions, although the effect of OM is larger ($d = .75$) than the effect of MP ($d = .44$).



(a)



(b)

Figure 13. *Results in terms of subjective measures with their 95% confidence intervals. The x-axis represent the six conditions of the experiment whereas on the y-axis the mean scores in term of perceived workload (a) and perceived discomfort (b) are given. The two lines are the two difficulty levels used in Experiment 3. Results in terms of workload showed that Motion (both OM and MP) decreased workload compared to the condition without motion. In terms of discomfort the results showed that without motion stereo increased discomfort, whereas with motion no such increase was observed.*

No differences in workload emerged between stereo ($M = 7.64$, $SE = .74$) and mono visualizations ($M = 7.06$, $SE =.58$) [$F < 1$]. The results further revealed a non-significant trend for an interaction between Visualization and Motion [$F(2,36) = 2.99$; $p = .06$, partial η2 = .14], suggesting that stereo increased workload in the static, but not in the motion conditions. Post-hoc tests with Bonferroni correction indeed revealed that in the stereo condition, both OM ($M = 6.76$, $SE = .87$) and MP ($M = 6.89$, $SE = .68$) had significantly lower levels of workload than the static condition. ($M = 9.27$, $SE = 1.0$; both $p < .05$, $d = .60$ and $d = .63$ respectively). In the mono condition no difference in workload emerged between OM ($M = 6.14$, $SE = .62$), MP ($M = 7.47$, $SE = .73$) and static conditions ($M = 7.56$, $SE = .73$; $p > .13$). No significant interaction was found between Motion and Difficulty (F<1, *ns*) and Visualization method and Difficulty [$F(1,18) = 2.35$, $p = .14$, partial η2 = .12]

The repeated-measures ANOVA with perceived discomfort as the dependent variable showed main effects of Difficulty, Motion as well as an interaction between Visualization method and Motion. The main effect of Difficulty level revealed a small effect [$F(1,19) = 7.39$, $p < .05$, $d = .27$ partial $\eta 2 = .28$] with higher levels of discomfort for the difficulty tasks (see Figure 13). The main effect of Motion [$F(1.41,26.8) = 4.72$, $p<0.05$, partial $\eta 2 = .20$] showed that compared to the static condition, both OM and MP rendered lower levels of discomfort, with a higher effect size for the OM condition ($d = .45$ and $d = .26$ respectively). The results did not reveal a main effect of Visualization method [$F(1,19) = 1.49$, $p = .24$, partial $\eta 2 = .07$]. As illustrated in Figure 13b, the main effect of Motion was mainly caused by the static stereo condition, which is confirmed by the interaction between Visualization method and Motion [$F(2,38) = 3.99$, $p < .05$, partial $\eta 2 = .17$]. Post-hoc tests with Bonferroni correction showed that no difference in perceived discomfort between mono and stereo in the OM ($\Delta M = .80$, $SE = 1.1$; $p = .50$) or MP ($\Delta M = .13$, $SE = .70$; $p = .90$) conditions, whereas in the static condition stereo showed a non-significant trend for increased discomfort ($\Delta M = 2.63$, $SE = 1.37$; $p = .07$, $d = .46$). Furthermore, the results did not reveal significant interactions between Difficulty and Visualization method [$F(1,19) = 1.98$, $p = .18$, partial $\eta 2 = .1$] and Difficulty and Motion [$F(2,38) = 1.29$, $p = .29$, partial $\eta 2 = .07$].

2.6.3 Summary of results

In line with Experiment 2, the results of this third study revealed a large increase in accuracy when participants interacted with the content using OM ($d = .98$) or ($d = 1.02$), compared to conditions without motion. On average participants responded correctly in 81% and 82% of the trials when using respectively OM and MP, compared to 59% correct responses without motion. In contrast to Experiment 2, completion times significantly increased when using OM (average of 6.9 seconds) or using MP (average of 6.7 seconds). This suggests that rotating the image took more time, but increased the number of correct answers, which is in line with the traditional thoughts on speed-accuracy trade-off. The use of stereo visualizations did not significantly improve accuracy. However, the significant interaction between motion and visualization method showed that stereo reduced completion times in the OM and MP conditions compared to the static condition. Although effect sizes were modest ($d = .39$ for MP and $d = .27$ for OM), results showed that participants completed the tasks faster in these conditions when combined with stereo, with on average, 3.4 and 4.6 seconds for OM and MP respectively. Interestingly, this reduction of completion times did not resulted in a change in accuracy, showing that the speed-accuracy trade-off did not play a role here. This finding is in line with results found in Experiment 2 and previous literature concerning MP effects (Naepflin & Menozzi, 2001). In contrast with Experiment 2, difficulty did not moderate effects of motion on accuracy, but showed to be an important parameter in understanding the potential effects of both motion and visualization method on completion times. For the easy tasks, visualization method did not have an effect on completion times, whereas for the more difficult tasks completion times decreased when images appeared in stereo. As mentioned before, using either OM or MP resulted in longer completion times, however also here the effect was more prominent for the difficult tasks as shown by the larger effect sizes for the difficult tasks (d = .99 for OM and d = .89 for MP) compared to the easy tasks (d = .41 for OM and d = .50 for MP). The results for workload were in line with those reported in Experiment 2; revealing lower workload when users rotated the images via OM or MP compared to static images. Effect sizes revealed that compared to the static condition, OM (d = .75) reduced workload more than MP (d = .44). Stereo did however not impact workload. Perceived discomfort slightly decreased using OM or MP, compared to the static condition. In addition, a combination of motion and stereo did not increase discomfort, whereas in the static condition stereo slightly increased perceived discomfort, which replicates the findings reported in Experiment 2. Nevertheless, participants reported relatively low perceived discomfort in the current study: averages ranged from 5 (mono) to 7 (stereo) on a 20-point response scale.

2.7 Discussion

In this chapter, we focused on the potential benefits of stereoscopic visualization, user-controlled object motion, and (head-controlled) movement parallax as means of displaying and interacting with data in depth. We studied effects on traditional task performance indicators (percentage correct; time needed to complete a task) as well as perceived workload, and perceived visual discomfort. Perceived workload has not yet been considered in the context of stereoscopic display evaluations, but we theorized that if data are processed more efficiently using stereo and/or motion cues, this should be reflected in individuals' perceptions of cognitive load, which should improve under display conditions that are optimal for executing the task at hand. Importantly, task difficulty was considered as a critical factor moderating the added value of stereo and/or motions cues, since we assumed that only for tasks of sufficient complexity an advantage of visualizing the third dimension would be evident.

To our knowledge, only one study investigated the combination of object motion, movement parallax and stereo in an integrated fashion (Ware et al., 1993). In addition, much of the available research on stereo has deployed a binary 'stereo-on-off' approach, without an attempt to identify an optimal disparity level. A more parametric approach to stereoscopic disparity levels is likely to be important, however, as the informational value of spatial representations in 3D will trade off against the potential cost of increasing visual discomfort with increasing stereo disparities (Lambooij et al., 2009).

The three studies reported in this chapter provide more insight into the added value of stereoscopic visualizations, both for settings in which users are able to control content using OM and MP, as well as for non-interactive settings. The results show consistent results across the three experiments: An optimal performance in terms of completion times, accuracy, and workload depends on task difficulty, visualization method (mono vs. stereo) as well as the availability of object motion and movement parallax.

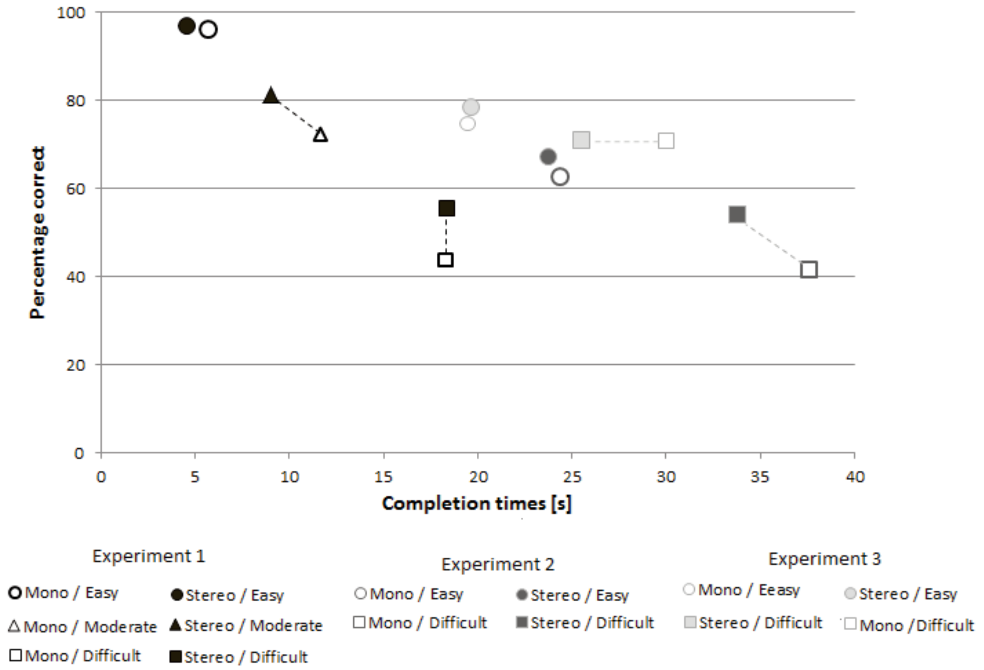


Figure 14. *Results of the three Experiments presented in Chapter 2, presenting the effects of visualization method and difficulty on accuracy (y-axis, in percentage correct) and completion times (x-axis, in seconds). The symbols and colors in this figure present the three experiments (Experiment 1: black; Experiment 2: grey; Experiment 3: light grey). In addition, the open symbols represent mono visualizations and the filled symbols stereo visualizations. The various symbols are the difficulty levels; squares for difficult tasks, triangles for tasks with medium difficulty and circles for easy tasks. The dotted lines between the various symbols are differences between mono and stereo visualizations for each difficulty level and experiment. We can see that across the three experiments, for both accuracy and completion times, stereo either yielded an advantage or resulted in no difference. In no case did it yield a disadvantage when compared to the monoscopic conditions.*

In Figure 14 and Figure 15, we visualized the results of these studies in terms of completion times and percentage correct. Figure 14 shows the effects of both visualization method and difficulty on completion times and accuracy in all three experiments presented in this chapter.

Generally, this figure shows that relatively easy tasks (circles) were performed faster and more accurate than difficult tasks (squares). This suggests that by manipulating the number of line segments in the wire frame stimulus the complexity of the task changes, i.e., participants were less accurate on tasks with more line segments.

The results presented in this chapter showed that for completion times task complexity had a large effect ($d > .8$) in all three experiments (see also Figure 14). As can be seen in Figure 14, the effect of difficulty level on accuracy was smaller in Experiment 3 (light grey symbols) compared to the Experiment 1 (black symbols) and Experiment 2 (grey symbols). This is in line with the reported effect sizes in these three experiments, showing a small effect of difficulty on accuracy in Experiment 3 ($d = .29$) and large effects in Experiment 2 ($d = 1.02$) and Experiment 1 ($d > 1.64$). These findings are probably due to the smaller difference in complexity level between the easy and difficult tasks used in Experiment 3 compared to Experiments 1 and 2.

Figure 14 also shows the effects of stereo (filled squares and circles) and mono visualizations (open squares and circles) on completion times and accuracy. Experiment 1 (black symbols) revealed a main effect of visualization method, suggesting an increase in percentage correct and a decrease in completion times for stereoscopic compared to monoscopic visualizations, which is in line with previous literature (Hu et al., 2002; Hubona & Shirah, 2005; McWhorter et al., 1991; Yeh & Silverstein, 1992). In addition, in Experiment 1, where no additional depth cues were available, the effect of disparity level was larger for accuracy (with effect sizes between $d = 2.70$ and $d = 2.90$), than for completion times (having effect sizes between $d = .55$ and $d = .78$). However, in Experiments 2 and 3, where users were able to rotate the images, results did not reveal a significant main effect of visualization method on accuracy or completion times. Previous literature showed that the use of effective monocular depth cues already increase performance to a level beyond which stereo does not further increases performance (Barfield & Rosenberg, 1995; Hendrix & Barfield, 1995). Although in terms of accuracy these findings are in line with Experiment 2 and 3, completion times were lower when motion and stereo were combined. In Experiment 1 we learned that the level of disparity is an important factor determining potential benefits of stereoscopic depth, which might explain why in Experiment 2 and 3 no main effect was found for visualization method. The higher disparity level employed in Experiment 2 induced more perceived discomfort, which may explain why completion times and accuracy did not benefit from stereo in this second study, even in conditions where motion was not present. Moreover, a relatively low level of disparity was used in Experiment 3 to avoid discomfort. This could, however, also explain why differences in performance between monoscopic and stereoscopic visualization method were more subtle.

Results in this chapter showed that difficulty level is also an important factor revealing performance benefits when using stereoscopic displays. As shown in Figure 14, the effect of visualization method on performance in terms of completion times was dependent on difficulty level: the effect of stereo was more pronounced for the difficult tasks, whereas for the easy tasks no difference emerged between stereo and mono visualizations. This interaction was, however, only significant in Experiment 1 and 3 (39% of variance explained[8]). In Experiment 1, the effect of visualization method on accuracy was also significantly moderated by difficulty level. Overall, the results suggest that both difficulty level and disparity level are important factors determining the added value of stereo presentations on task performance. Across the three experiments, for both accuracy and completion times, stereo either yielded an advantage or resulted in no difference (depending on whether motion cues were present – see next paragraph). In no case stereo did yield a disadvantage when compared to the monoscopic conditions.

In Experiments 2 and 3, we investigated the effect of motion, in addition to stereo, on task performance in terms of accuracy and completion times. Figure 15 shows the effect of motion (diamond symbols) compared to static visualizations (circles) on completion times and accuracy. In addition, the filled symbols represent stereo visualizations and the open symbols mono visualizations. As Experiment 3 did not reveal different results between OM and MP, we averaged these findings in Figure 15. Note that the overall higher levels of accuracy and lower completion times in Experiment 3 are probably due to the easier tasks applied in Experiment 3. Results of both studies showed that accuracy improved when users were able to rotate the task (see Figure 15); showing large effects in Experiment 2 (d = 1.73) and Experiment 3 (d = .98 for OM and d = 1.02 for MP). In addition to improvements in accuracy, results of Experiment 3 revealed an increase in completion times when using motion compared to conditions without motion ($d \approx .50$). In contrast, in Experiment 2 completion times did not differ between static visualizations and OM. The results of Experiment 3 are in line with findings from literature (Naepflin & Menozzzi, 2001; Faubert, 2001; Sollenberger & Milgram, 1993;

[8] *Note that we could not establish partial η2 values in Experiment 1 due to the hierarchical structure of the data.*

Ware et al., 1992, Hubona et al., 1997; Ware & Mitchell, 2008), suggesting that motion is an effective cue for disambiguating complex spatial structures, but that this process does require extra time since an additional interaction is required in order to reveal the spatial structure.

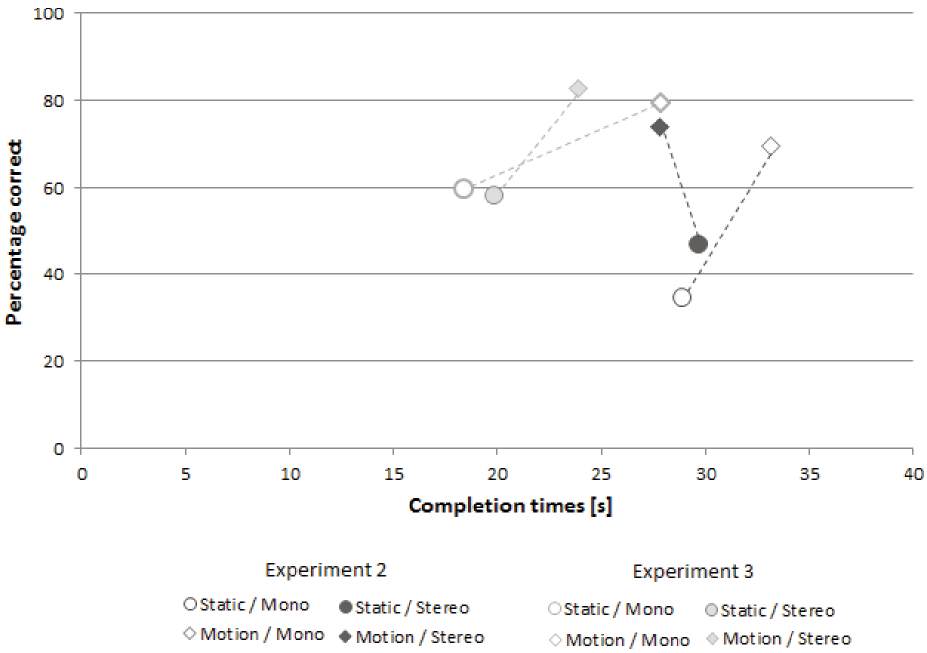


Figure 15. *Results of Experiments 2 and 3 presenting the effects of Visualization method and Motion on accuracy (y-axis) and completion times (x-axis). The symbols and grayscale colors in this figure present the two experiments (Experiment 2: grey; Experiment 3: light grey). In addition, the open symbols represents mono visualizations and the filled symbols stereo visualizations. The diamond symbol represent the conditions with motion and the circles those without motion. The dotted lines between the various symbols are differences between the motion and no motion conditions.*

Figure 15 also shows a beneficial effect in terms of completion times when combining motion and stereo cues. In both Experiment 2 and 3 the effect sizes found for the interaction between motion and visualization method (e.g., partial $\eta 2 = .27$ in Experiment 2 and partial $\eta 2 = .41$ in Experiment 3), were larger than the main effect sizes of visualization method (e.g., partial $\eta 2 = .12$ in Experiment 2 and partial $\eta 2 = .25$ in Experiment 3). This illustrates that Stereo most effectively decreased completion times when combined with Motion. Thus, even though motion (i.e., temporally integrated successive views) offers more or less the same geometric object information as stereo (i.e., spatially integrated views), seeing depth instantaneously (i.e., using stereo) combined with the ability to rotate the images can speed up task performance. These results are in line with findings reported by Naepflin and Menozzi (2001), Hubona et al., (1997) and Ware and colleagues (1993), all revealing the most efficient performance when combining OM or MP with stereo. However, in contrast to our findings, some studies also revealed an increase in accuracy for stereo visualizations combined with motion (Naepflin & Menozzi (2001), Hubona et al., (1997), Sollenberger & Milgram (1993), and Ware & Mitchell (2008)). These differences cannot be explained by the absence or presence of user control, since both Sollenberger and Milgram (1993), and Ware et al., (1993) used controlled OM and still noted added benefits of stereo and motion on accuracy. Potentially, the time limit employed by Sollenberger and Milgram (1993) and Ware and Mitchell (2008) played a part in this. Our findings showed that users require more time solving the task with only motion. Limiting the time users have to accomplish the task, could therefore decrease their accuracy; in such cases adding stereo cues may help to quickly disambiguate the image.

In the introduction of this chapter, we argued that interaction via head movements (i.e., MP) is an embodied method of interaction, in which both motor and perceptual information are used during the task. Mouse-based interaction (i.e., OM), on the other hand, is less embodied since the relation between our own movement and the changes on the screen is less direct. Therefore, we hypothesized that tasks would be performed better in the MP condition than in the OM condition. Results, however, showed that whether motion is controlled by head movement or via the mouse did not affect performance in terms of either accuracy, completion times, or workload. One explanation can be that the mouse is also an embodied method of interaction in this task, since the direction of the hand movement corresponds with the rotation of the volume. Furthermore, since we use the mouse for computer work, it has become a natural method of interaction, and people can use it effortlessly. In the following chapters of this thesis we will further explore the concept of embodied interaction, extending currently applied performance-based measures towards a broader perspective of user experience (Chapters 3 and 4).

Perceived workload has been shown to be a useful concept in studying task performance as a complementary measure to objective indicators such as completion times and accuracy. Generally speaking, when cognitive resources are not yet depleted, subjective workload measures could be sensitive to an increase in cognitive load even when primary task measures do not yet yield any measurable effect. In the experiments reported in this chapter, workload was consistent with the findings reported for the primary task-performance measures (i.e., accuracy and completion times) for difficulty (Experiments 1, 2, and 3), motion (Experiments 2 and 3), stereoscopy (Experiment 1). Motion decreased perceived workload in both Experiment 2 and 3. However, the workload reduction was larger in Experiment 2 (d = 1.06) than in Experiment 3 (d = .75 for OM and d = .44 for MP). An explanation why the effect of workload is larger in Experiment 2, than in Experiment 3, could be the higher disparity level for the stereo visualizations employed in Experiment 2. Another explanation could be the higher task complexity for the difficult tasks, which potentially requires motion more for a successful task completion. Stereo visualizations did not affect perceived workload in both Experiments 2 and 3, but workload was slightly lower in Experiment 1. Moreover, although the effects of difficulty level and motion on completion times were moderated by stereo in Experiments 2 and 3, workload did not follow this result. This showed that, in line with O'Donnell and Eggermeier (1986), primary task measures are not always sensitive in measuring participant's workload. An explanation for why the introduction of stereo did not decrease workload might be the relatively large disparity level used in Experiment 2, which may have induced discomfort and thus negated any potential positive effects of stereo on perceived workload (as shown for the difficult tasks in Experiment 1). On the other hand, in Experiment 3 the level of disparity may have been too small, and therefore insufficient to extract additional depth information above what was already available when rotating the object. This hypothesis will be tested in Chapter 4, in which motion will be combined with stereo, using a disparity level that lies between the levels used in Experiments 2 and 3.

Another interesting finding, reported both in Experiment 2 and 3, was the effect of Motion on perceived discomfort. In line with what we expected, stereo increased perceived discomfort compared to monoscopic presentation in two of the three experiments. The larger disparity level used in Experiment 2, resulted in a significant but small increase in discomfort (d = .32), whereas in Experiment 3, where we used a smaller disparity level, stereo did not significantly increase discomfort. Nevertheless, the interaction between visualization method and motion in both experiments showed that compared to the static stereo condition, combining motion and stereo decreased discomfort with comparable effects sizes (d = .48 in experiment 2 and $d \approx .6$ in Experiment 3). This suggests that when an object is moving, using either OM or MP, stereo leads to less discomfort compared to the static conditions. An explanation can be that when an object is rotating, the eye is fixating less towards a fixed point in the image, thereby potentially ameliorating the accommodation/

vergence conflict. This is a potentially valuable result, however, more research is clearly needed into this topic before we can confirm or discard this assertion. Such research would need to include a more extended set of questions, a broader set of visual stimuli, and a set of objective, optometric indicators of the visual state of the participants' eyes, which falls outside the scope of this thesis.

In sum, results in this chapter showed that both difficulty level and disparity level are important factors determining the added value of stereo presentations on task performance. Results showed that stereo either yielded an advantage or resulted in no difference in performance, whereas stereo never decreases performance. Motion showed to be the most important factor to increase accuracy, however combined with stereo tasks were performed faster than motion without stereo.

2.7.1 Practical implications

From this chapter we learned that stereo is most effective in reducing completion times. Results showed that using 3D displays do not always increase performance, since the effectiveness of stereo depends on both disparity level employed and difficulty level. Stereoscopic 3D displays have been shown to be most effective for tasks with higher levels of complexity. Tasks that are too easy reveal a floor effect, as other depth cues (pictorial or motion) may already yield sufficient depth information for optimal performance. Not only task complexity is important, also the level of disparity used in the experiments affects the performance benefits of stereo displays. Using disparity levels that are either too large or too small will not result in improved performance levels over and above those found for a monocular presentation of the task. Lastly, results showed that when users were able to interact with the content (e.g., rotate the stimuli), performance increased and workload decreased. Although task performance already increased when participants were able to rotate the image, a stereoscopic presentation of the content decreased completion times and revealed lower levels of perceived discomfort compared to rotating the image in 2D. All these considerations can make it difficult for a designer to decide whether or not stereoscopic displays will benefit task performance. Overall, stereo speeds up tasks performance without decreasing or increasing accuracy. Nevertheless, the effects of stereo found in this study were not as large as the effects for motion that had a larger effect on improving accuracy, however with increasing completion times. In these circumstances, a combination between motion and stereo seems to combine the best of both worlds and revealed the most optimal performance. One caveat should be noted though: since the wire frame stimuli used in this experiment offered little pictorial depth cues, it is unclear if similar results will be found when using stimuli that contain additional monocular depth cues. However, for tasks that contain only a small or degraded set of monoscopic depth cues (e.g., due to the specific imaging technique, such as x-ray or sonar), or tasks that utilise imaging to reveal structures that are inherently ambiguous in 2D, such as complicated vessel structures as found in angiography, 3D displays (ideally in combination with object rotation) will offer improved task performance. In the following chapters, we will explore new and innovative ways of interacting with 3D displays, and study the potential use of gesture-based interaction in terms of performance and user experience.

CHAPTER 3

A user-centered perspective on embodied interaction[9]

Technology is not far away and impersonal. It's here, it's intensely personal, and it's great fun. (Sutherland, 1996, p. 31)

[9] *Experiment 4 has been reported in: Beurden, van M.H.P.H., & IJsselsteijn, W.A. (2010). Range and variability in gesture-based interactions with medical images: Do non-stereo versus stereo visualizations elicit different types of gestures? IEEE Virtual Reality: workshop on medical virtual environments, Waltham, MA, USA.*

Experiments 5 and 6 have been reported in: Beurden, van M.H.P.H., IJsselsteijn, W.A., & de Kort, Y.A.W. (2012). User experience of gesture based interfaces: A comparison with traditional interaction methods on pragmatic and hedonic quality. In: E. Efthimiou, G. Kouroupetroglou & S.-E. Fotinea eds Gesture and Sign Language in Human-Computer Interaction and Embodied Communication. 9th International Gesture Workshop - GW 2011, Athens: Revised selected papers. LNCS/LNAI Vol. 7206, Springer.

3.1 Introduction

In Chapter 2, we focused on task performance and demonstrated that actively interacting with stereoscopic content, enabled through controlled object motion or movement parallax, increased performance and decreased perceived workload of a task compared to passively watching these images. In Chapter 1, we discussed that for embodied interaction both body representations and sensory-motor contingencies are important for users' direct and intuitive engagement with virtual content. As technologies for sensing and processing are advancing rapidly, and are becoming cheaper and more commonplace, these interfaces are reaching a state of development in which their performance level and user experience can be evaluated against more traditional device-based interfaces. For instance in the game domain recent studies showed that the Wii controller was perceived as more natural than a standard controller, and the sense of spatial presence and game enjoyment increased compared to standard consoles (McGloin, Farrar & Krcmar, 2011; Skalski, Tamborini, Shelton, Buncher & Lindmark, 2011). Yet, motion-based controllers do not necessarily present the most accurate interface for every task. In a racing game, McMahan and colleagues (2010) revealed that the Wii console used as steering wheel, resulted in lower performance than a standard console, although the Wii was more fun to use. In this study, latency and less accurate steering performance using the Wii might explain the decreased performance. Although the Wii enabled a more natural interaction compared to traditional controllers, for some applications (i.e., medicine) a tool might not be the most practical solution due to sterility requirements. In Chapter 2, we showed that head movements revealed similar performance as mouse-based interacting when solving a complex task. Another potential interaction method that does not require a device is gesture-based interaction in which participants can interact with technology using hand and arm movements.

3.1.1 Gesture-based interaction

Gesture-based interaction allows users to interact with a computer or technology via hand and arm movements. According to the Oxford dictionary a gesture is defined as "a significant movement of a limb or the body as an expression of thought and feelings". In daily life, we use gestures while we talk, think, communicate with each other, and manipulate objects around us. To structure the various gestures, several taxonomies have been developed based on human communication and linguistics (Cassell, 1998; Kipp, 2004; McNeill, 2005; Quek et al., 2002). These taxonomies describe the gestures performed during spoken human-human communication or communication without speech. Based on these classifications, various researchers have developed taxonomies that can be used for human-computer interaction (Karam & Schraefel, 2005; Pavlovic, Sharma & Huang, 1997; Quek, 2004). Comparing the various taxonomies is difficult, since different terminology has been used to refer to the same gesture (Wexelblat, 1998). In addition, taxonomies developed for HCI are based on taxonomies that originate from human communication, but not all of these gestures are suitable for gesture-based interaction. To clarify this point, we will discuss the various gesture classes in terms of naturalness and embodiment by using the taxonomy described by Karam and Schraefel (2005). In this taxonomy five gesture classes are defined that can be used when interacting with computers – deictic, gesticulation, semaphore, manipulation, and sign language (see Figure 16).

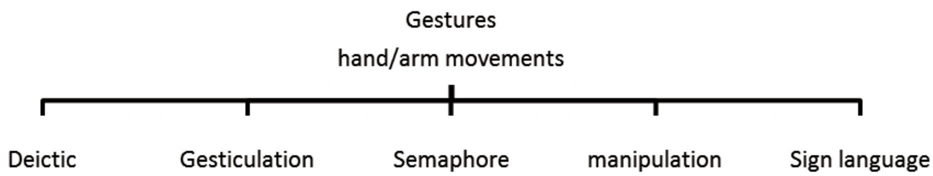


Figure 16. *Gesture classes identified by Karam and Schraefel (2005) that can be applied during human-computer interaction.*

Deictic gestures are pointing gestures to identify objects in the environment. One of the first gesture-based interfaces that used deictic gestures was the 'put that there' interface developed by Bolt (1980), combining speech with object identification using gestures. These gestures are natural and embodied since we use them in daily life when pointing to objects around us.

Gesticulations are gestures used to accompany speech, and are the gestures we use most frequently in daily life (McNeill, 2005). These gestures are spontaneous and can be used to clarify or emphasize speech. Examples of gestures accompanying speech are metaphors (for example to say 'on and on' making hand rolling gesture), icons (describing a square while drawing it with one's hand), emblems[10] (gestures which have an often culturally base conventional meaning such as thumb up (good) or beat (rhythmic movements with no relations to speech content). Although some of these

[10] *According to McNeill's taxonomy (McNeill, 2005) emblems are a different category, since speech is not always present.*

gestures can be applied in human-computer interaction (e.g., metaphors, icons, emblems) some are not likely to be used during the interaction (e.g., beat). In addition, since some gestures accompany speech, not all of these gestures will be meaningful without speech, and might therefore be less suitable for interaction with computers without speech recognition. An example of a gesture falling under gesticulation and applied in current touch-based interfaces is zooming in and out using a pinch or reverse-pinch gesture. This gesture relates well to the metaphor of enlarging an elastic object (that might be used during speech emphasizing the growth of something).

Semaphores refer to gestures in which hand and arm movements represent signs or signals to communicate information. This gesture class does not exist in any taxonomy based on human communication or linguistics and is an interaction method representing abstract gestures (i.e., signals) to communicate information. An example is the flag system in which specific flag positions represent an alphabetic code. Specific hand and finger configurations make it is easier to detect the gesture by a tracker, and are therefore frequently applied when developing gesture interaction. Typically, this interaction is not very natural and embodied since it does not have any relation with previously learned skills, and the gestures do not carry intrinsic meaning for the naive user.

Manipulative gestures are gestures aimed to control an object, such as steering a cursor or resizing objects on the screen. This can be performed on a computer screen, or during tangible interaction in which the user manipulates real world objects presented virtually (Fitzmaurice, Ishii & Buxton,1995). These manipulative gestures are well fitted for tasks on computer screens that involve manipulation of objects, using skills acquired in the real world.

Sign language is closely related to semaphores and used by the deaf to communicate with each other. This class of gestures can be used for a specific group in computer applications such as teaching children sign language, however it is not likely to be applied in everyday life interaction due to the complexity of the sign language for everyday users. To our view sign language is a special type of gesturing, which is natural and possibly embodied for deaf people, however not for the general hearing human population.

The taxonomy presented in Figure 16 shows the various gesture classes that can be used when interacting with virtual environments or technical artefacts. An implementation of gesture-based interaction will most likely be a combination of different gesture classes, as shown by Karam and Schraefel (2005). However, in order to arrive at embodied interaction, the most promising gestures are deictic, manipulation, and gesticulation. These gestures most frequently rely on previously learned sensory-motor couplings (i.e., grab an object, rotate an object) and use representations of our body and the environment when gesturing (i.e., rolling-hand gesture meaning go-on; pinching by means of stretching an image). Semaphores, although frequently applied when designing gesture-based interactions, do not have a meaning related to our body, other than pre-defined codes which we have to learn before using them. Therefore, semaphores are less embodied than the previously mentioned gestures.

3.1.2 Challenges and advantages of gesture-based interaction

In the previous paragraphs, we explained the concept of gesture-based interaction, and discussed the different gesture classes that can be applied in human-computer interaction. In this section, we will discuss the challenges and advantages of gesture-based interaction.

Deviceless interaction

Since gesture-based interaction does not require any device, it is advantageous to use it in settings in which handhelds are not available or desired. Such practical needs arise not only close to home - where the number of remote controls appears to be ever increasing and the right one always appears to be lost - but also in more advanced contexts such as operating rooms. Operating rooms have stringent sterility requirements and require fast and intuitive access to volumetric medical data without the need for a controller. Several gesture-based interaction technologies are currently being explored in this context. Examples include the FAce MOUSe (Nishikawa et al., 2003), a laparoscopic positioning system controlled by the surgeon using face gestures, the Non-Contact Mouse (Graetzel, Fong, Grange & Baur, 2004), with which surgeons interact with endoscopic images while using a well-defined set of gestures to perform standard mouse functions (pointer movement and button presses), and Gestix (Wachs et al., 2008) which is another hand gesture-based system for browsing medical images from an EMR data base. Another potential advantage when interacting with one's hands is that the degrees of freedom is larger than when holding a device (Sturman, 1991). In addition, with our hands we are more flexible in expressing ourselves and therefore a potentially large variety of gestures can be used when interacting with technology. However, as already discussed in Chapter 1, in daily life we also use tools and devices for many tasks. Haans and IJsselsteijn (2012) already described that the human body is able to incorporate tools into our body schema, to the extent that we are no longer aware of holding a tool (see also Clark, 2003). This is in line with Winograd and Flores (1988) arguing that tools become transparent when users are unaware of the interaction device and fully focused on the task at hand. This suggests that when similar movements are used in both tool-based and gesture-based interaction, the experience might not be different, which will be addressed in Experiment 6.

Gesture set

The development of deviceless interaction still presents a challenge for technicians developing gesture-based interaction systems, that can detect arm and hand movements accurately. To increase recognition accuracy (e.g., ability to recognize the gesture correctly), for instance for surgeons in the examples presented earlier, a relatively well-specified and limited set of gestures need to be learned (i.e., semaphores). This will increase the detection of gestures by the software, however it makes the interaction less embodied and natural. Similarly, Quek (1996) and Graetzel et al. (2004) developed a gesture-set based on what could be detected by their tracking technology. However, such gestures might again not be the most natural ones when interacting with computers. Although an accurate detection of gestures is an important factor when developing interaction technology, gestures should also be intuitive and map naturally on the task at hand. In the last few years, technological developments in gesture-based interaction systems have progressed rapidly (e.g., Microsoft Kinect, Leap Motion), and hand and finger movements can be detected more accurately than before. Yet, an additional challenge, related to recognition accuracy, is to detect when a gesture starts and when it ends (Wexelblat, 1998). This is a particular challenge when gestures are less reliant on a well-specified, limited set of gestures, but invite the use of broader, more natural, and less well-specified sets of gestures. Nevertheless, new and more accurate detection algorithms opens new opportunities for developing gesture sets that correspond to our previous real world experience. A few studies have looked at the range and variability of gestures users naturally make when interacting with interfaces such as surface computing (Wobbrock, Morris & Wilson, 2009), computer displays (Hauptman, 1989), or large projection screens (Fikkert, 2010). Hauptman (1989) analysed the use of gestures and speech for graphic manipulation. Results showed that users moved their hands in all three dimensions, users preferred a combination of speech and gestures, and users used both hands and multiple fingers when manipulating objects. In this paper, Hauptman did not explain how the gestures were performed but discussed objective criteria such as the number of hands and fingers used, motion trajectories of the hands, etc. Wobbrock and colleagues (2009) studied users' gesturing using a table-top environment and asked them to perform tasks such as rotating, minimizing, zooming, and deleting. Results showed that participants used one hand more frequently than both hands, and the performed gestures were often based on existing desktop metaphors. Furthermore, some tasks (e.g., zooming, deleting, minimizing items) did not show much agreement across participants. Fikkert (2009) used a Wizard of Oz experiment determining gestures for panning and zooming. Results of this experiment also revealed variability in gesturing performance; for zooming the variability across participants was larger (six different gestures) than for panning tasks (three different gestures). These results suggest that arriving at a gesture set that is intuitive to all users will not be easy, if not impossible, although one should strive for a set that feels natural and intuitive to the majority. The studies reported here used different displays (touch screen, desktop computing, large screens), and different sets of tasks, and therefore the results cannot easily be generalized. In addition, from previous studies it is currently unknown whether monoscopic visualization would elicit the same set of gestures compared to stereoscopic presentation. Nevertheless, when developing gesture-based interaction, knowledge of these studies and the methodology can be used to understand the range of gestures that come naturally when interacting with displays.

Body fatigue

Using our whole body when interacting with technology can induce fatigue much like any physical exercise. This issue should be taken seriously when developing gesture-based interaction. High levels of discomfort might dissatisfy users, degrade performance and even injure users (Hinckley, Pausch, Globe & Kassel, 1994). Discomfort can be decreased when gestures and arm movements are performed closer to the body (Kölsch, Beall & Turk, 2003). In well-designed gesture interfaces fatigue should only occur after prolonged uninterrupted use of the interface, and therefore Hinckley et al. (1994) suggested that some time-outs should be built in, in which users can rest their arms.

Evaluation methods

In Chapter 1, we already discussed the model-based approaches (e.g., Fitt's law, Steering Law) used when evaluating interaction technologies. Performance-based measures (e.g., completion times and accuracy) are the most important attributes used during these evaluations. As discussed in Hornbaek (2006), Nielsen (1994) and the Usability standard (ISO 9241-11, 1997), when evaluating products one should include both performance-based measures as well as users' attitudes and experiences, frequently measured in terms of satisfaction, ease of use and learnability. Few authors have taken into account user characteristics when designing and evaluating interaction devices (e.g., Buxton, 1983; Card, Mackinlay & Robertson, 1990), however these efforts contrast sharply with efforts in the domain of Graphical User Interfaces (Bowman et al. 2005). In line with the embodied interaction perspective described in Chapter 1, interaction technologies may also change experiences of users. Recent studies have shown that interaction technologies that support hand and arm movements increase users' experiences of fun and engagement in game environments (McGloin, Farrar & Krcmar, 2011; Skalski, Tamborini, Shelton, Buncher & Lindmark, 2011). An explanation can be that embodied interaction increases bodily engagement during the interaction, which gives rise to a more visceral experience, affecting user's emotional state as discussed in the James-Lange theory of emotion (James, 1884). According to James, our emotions are formed through our bodily activity, and therefore emotions are embodied. In line with this thought, Riskind and Gotay (1982) showed that posture affects emotional experience and behaviour. In addition, when participants were able to mimic facial expressions, they detected a change in facial expression earlier than when participants were prevented from mimicking facial expressions (i.e., by holding a pencil between their lips; Niedenthal, Brauer, Halberstadt & Innes-Ker, 2001). These studies have shown that information provided through our body is used in everyday tasks and influences experiences. Within the Human Computer Interaction (HCI) community, there is an increasing interest in incorporating users' feelings and emotions when evaluating and designing products. The term used to cover both performance (usability) and affective information is 'user experience' (Hartson & Pyla, 2012). User experience includes both pragmatic and hedonic quality. Pragmatic quality is the extent to which a system allows for effective and efficient goal-achievement and is thus closely related to the notion of usability. Hedonic quality is the extent to which a system allows for stimulation by its challenging and novel character, or for identification by communicating important personal values (Hassenzahl, 2004). Although the current usability measures (which includes satisfaction) already gives some information on how users feel about the technology, it typically refers to how users experience usability and usefulness and can therefore also be considered as a component of pragmatic quality (Hassenzahl, 2004). Hassenzahl (2004) argued that in addition to these pragmatic indicators, hedonic indicators are also important when evaluating products (see also, Jordan, 2000; Norman, 2004). The hedonic qualities represent users' personal values, users' emotional states and pleasure experienced using technologies. Higher hedonic qualities are often associated with more pleasure (Hassenzahl, 2004), which in turn may enhance creativity and cognitive flexibility (Baas, de Dreu & Nijstad, 2008), and is therefore relevant to both entertainment as well as professional applications. In addition, time is an important factor in defining how users experience and evaluate products (Karapanos, Zimmerman, Forlizzi, & Martens, 2009). In this paper, the authors showed that early experiences relate mostly to ease of use and hedonic aspects that are concerned with being stimulated, whereas when participants used the product longer (up to 4 weeks), experiences of how the product becomes meaningful in one's life became more important (e.g., hedonic quality-identification) (Karapanos et al., 2009). In the current thesis, we define user experience as the totality of experience of users when using a product/system/device (Hartson & Pyla, 2012). To our knowledge gesture-based interaction have not yet been evaluated in terms of their broader user experience. Therefore, Experiments 5 and 6 extend currently widely applied performance-based measures, and include affective and hedonic qualities as an inextricable part of the user experience (Norman, 2004; Tractinski, Katz & Ikar, 2000).

3.1.3 Rationale for the studies

Frequently mentioned advantages of embodied interaction, both with and without a device, are its potential for more natural expression, and greater ease of learning. Gesture-based interaction has practical advantages for environments in which a controller is not desired, and users can make a larger number of expressions without a controller. In spite of the restrictions and limitations, such as decreased accuracy and body discomfort, we expect that gesture-based interaction – once developed and designed to a sufficient level of accuracy and reliability – has the potential to compete with device-based interaction on relevant tasks. In addition, we hypothesize that embodied interaction may change the user experience, resulting in an enhanced personal identification, and enjoyment using systems that respond to personal, expressive movements. In the current chapter, we present three studies on gesture-based interaction.

In order for gesture-based interfaces to be natural and embodied, we need an understanding of the kinds of gestures that come naturally when interacting with a screen. Previous studies have often based their gesture set on what can be accurately detected (e.g., Quek, 1996; Graetzel et al., 2004), hence gestures were not always natural to use or easy to remember and are therefore less embodied. Only a few studies took a user-centered perspective, studying gestures that users produced spontaneously while performing a variety of tasks (Fikkert, 2010; Hauptman, 1989; Wobbrock, Morris & Wilson, 2009). In Experiment 4, we explore which gestures users perform naturally when manipulating a 3D object on a computer screen. Moreover, we investigate whether the production of gestures is different depending on whether users interact with a traditional non-stereo display as compared to a stereoscopic 3D display. We expect that images that are not displayed in stereo will elicit gestures that are in line with traditional interactions using a desktop metaphor (e.g., point and click, double-click). In contrast, we expect that stereoscopic 3D images will elicit gestures that are more spatial in nature, which will include more movement in the directions of the three-dimensional action space and an increase in two-handed interactions.

The gestures found in Experiment 4 are used as input for a working prototype developed for Experiment 5. In the fifth experiment, we compare the user experience of embodied interaction (i.e., gesture-based interaction) with controller-based interaction that can be seen as less embodied (i.e., mouse-based interaction). In section 3.1.2, we hypothesized that interaction that is embodied will give rise to a more visceral experience measured in terms of user experience. We therefore compared gesture-based interaction with mouse-based interaction in terms of both pragmatic and hedonic qualities. With pragmatic quality, we address factors related to the traditional notion of usability, focusing on effective and efficient goal-achievement. The evaluation of hedonic qualities allows us to explore the extent to which the system is experienced as fun, original, interesting, engaging, and personally relevant. Since body discomfort is often considered as potential side effect of gesture-based interaction, we also include subjective experience of body discomfort.

In Experiment 6, we aimed to compare the user experience of two embodied methods of interaction; - i.e., gesture-based interaction with interacting using the Wii (offering controller-based yet embodied interaction), improving limitations of the setup used in Experiment 5. Again, both pragmatic and hedonic qualities are measured, as well as body discomfort. Results of Experiment 5 and 6 will provide a better understanding of the effects of embodied interaction on the user experience in terms of both pragmatic and hedonic qualities.

3.2 Experiment 4: Range and variability of gesture-based interaction

In the first study of this chapter, we presented images to the participants and asked them to use gestures to generate specific actions. We explored which gestures they used, and whether these gestures differed depending on whether images were displayed in 2D or 3D.

3.2.1 Method

Design

We manipulated Visualization method (non-stereo vs. stereo) in a between-subjects design. Within each presentation mode participants performed seven different tasks; positioning, selecting, activating, rotating, zooming in, zooming out, and deactivating. The tasks were performed using four types of content (three images and one overview of these images). The gestures were recorded and later categorized.

Participants

Twenty-four participants between 20 and 34 years, all with normal or corrected to normal vision took part in this study. All participants had stereovision better than 40 seconds of arc, measured with the Randot® stereotest. Participants were either students or employees at the Eindhoven University of Technology with no or little knowledge of using gestures as interaction technology.

Setting and Apparatus

The stimuli were displayed on an HHI Free2C 3D Display, which was a 21.3 inch screen used in portrait format. The resolution of the display was 1200 x 1600 pixels. The stereo-view on this display was created using a moving lenticular which steers the exit pupils to the user's current eye position. All gestures were recorded with three cameras, one from the left, one from the right and one from above (see Figure 17), to ensure that the gestures would be clearly visible for later analysis. Users were seated approximately 60-70 cm from the screen.

Stimuli

The participants performed seven tasks as described above. Three medical images were used: a scan of a heart, a hip with blood vessels, and an image of a spine. In a fourth image, the three medical images were arranged vertically. The images were obtained from a public domain website (Fovia, 2010), and did not contain any identifiable patient information. The maximum disparity was approximately 60 min of arc.

Measures

User-generated gestures were recorded and the video streams were later analyzed using Noldus Observer XT 9. Two observers performed the classification of the gestures (inter-observer reliability was 94%). Gestures were classified based on the number of hands and fingers used during the interaction. In addition, we classified the gestures into functional categories, which emerged after analysis of the videos.

Procedure

On arrival at the lab, users were made aware that their behavior was recorded during the experiment. After a stereo acuity test, we explained the procedure and explained that there would be no restrictions and no right or wrong answers, in other words any kind of gesture would be acceptable in this experiment. To familiarize participants with the task of gesturing in relation to what was being displayed, an image with four colored squares was presented and the participant was asked to point to the colors mentioned by the experimenter. The actual experiment then started with an overview of the three medical images. Participants first gestured to move one of the images to the bottom or the top of the screen, and then gestured to select a specific image. This image was subsequently shown and users performed five tasks.

Figure 17. *the three camera settings used to analyses the gestures performed by the participants*

The first was activating the volume (i.e., such that it can be manipulated using gestures), followed by rotating, and zooming in and out of the volume. The last task was to deactivate the volume, such that it would no longer respond to hand movements. After completing these tasks, the overview image was presented again and users were asked to select the next image. This procedure was repeated for all three medical images.

The seven tasks were explained with short scenarios without the use of technical terms, such as rotation or zoom-in or zoom-out, to avoid a priori associations with desktop metaphors, mobile phones with touch screens, or other technical products users might have been familiar with. We formulated rotation as: “If you want to see the back or the side of the volume, how would you do that?”, zooming in: “If you would like to see that structure in more detail, how would you do that?”, zooming out: “If you want to go back to the original size, how would you do that?”, selecting: “If you would like to select one of the images, how would you do that?”, positioning: “If you would like to move the object on the screen, how would you do that?”, activate: “How would you make the system aware that you want to interact with the content?”, deactivate: “How would you deactivate the system such that it does not respond to your gestures anymore?”. The order of the images was counterbalanced for each participant. We manipulated 2D and 3D between participants; i.e., participants saw the images either in monoscopic or in stereoscopic viewing mode. The gestures performed by the participants did not result in changes in the image, i.e., the images remained static during the experiment. The duration of this experiment was around 15 minutes and users received a compensation of 5 Euros for their participation.

Statistical Analysis

We used Chi-square tests to investigate the effects of action performed (selection, activating, rotation, zoom-in, zoom-out, and deactivating) and visualization method (mono vs. stereo) on the number of hands and fingers used during the interaction, followed by a discussion based on the qualitative assessment of the functional categories.

3.2.2 Results

First, we will discuss the results in terms of the number hands and fingers used when performing the gestures. Subsequently, we will discuss the results of the video analyses, classifying the range of gestures users performed when interacting with the content.

Number of hands and fingers

Videos of the sessions were used to analyze whether participants used one or two hands in gesturing specific user-actions. The results of this analysis are reported in Table 4. Overall, one-handed interaction was used in 85 percent of the cases whereas in 15 percent of the sessions users performed gestures using both hands. A Chi-square test with presentation mode and number of hands showed that the number of hands differed between the 2D and 3D conditions [χ^2 (1, N = 456) = 16.41, $p < .001$]. Unexpectedly, the 2D condition triggered two-handed interaction more frequently (22%) than the 3D condition did (8%). Results of a second Chi-square test showed that the use of one or two hands also depended on the type of task performed [χ^2 (5, N = 456) = 16.79, $p < .01$].

In all tasks the majority of users performed the gesture using one hand, however for zooming in (17%), zooming-out (24%), rotating (18%) and deactivating (21%) a larger proportion of gestures was performed with two hands, than for positioning (0%), selecting (6%) and activating (11%). Table 4 further shows that for 3D visualizations only zooming-in, zooming-out, deactivating and selecting were performed using two-hands, whereas for 2D visualizations also activating and rotating were performed using two hands.

Table 4: *Percentages of sessions in which users used one vs. two hands when performing the six tasks, for 2D and 3D presentation*

Percentage (per action)		*Action*						
		Positioning	*Selecting*	*Activating*	*Rotating*	*Zooming-in*	*Zooming-out*	*Deactivating*
2D	One hand	100%	97%	78%	64%	75%	81%	67%
	Two hands	0%	3%	22%	36%	25%	19%	33%
3D	One hand	100%	92%	100%	100%	92%	72%	92%
	Two hands	0%	8%	0%	0%	8%	28%	8%

In addition to the number of hands used, we also analyzed the number of fingers used in each gesture. Results indicate that users used one-finger interaction more frequently in the 3D setting (49%) than in the 2D setting (21%), whereas the whole hand was used more frequently in the 2D (68%) than during the 3D visualization (42%). A Chi-square test showed that this association was significant [χ^2 (2, N = 456) = 40.30, $p < .001$]. A Chi-square test also showed an association between the number of fingers and the type of task performed [χ^2 (10, N = 456) = 82.30, $p < .001$]. As shown in Table 5 for selecting, 61% of the gestures were performed with one finger, whereas for rotating, zooming (in/out), activating, and deactivating the whole hand was used more frequently.

Table 5: *Percentages of sessions in which users used one finger, two fingers or the whole hand when performing the six tasks.*

Percentage (per action)	*Action*						
	Positioning	*Selecting*	*Activating*	*Rotating*	*Zooming-in*	*Zooming-out*	*Deactivating*
One finger	54%	61%	46%	23.5%	28%	24%	19%
Two fingers	8%	10%	1%	5.5%	25%	28%	3%
Whole hand	38%	29%	53%	71%	47%	58%	78%

Functional categories

During video analyses the gestures were classified using functional categories. In discussing these gestures we limit ourselves to the most frequently used categories presented in Figure 18 and .Pointing to a certain color was not one of the main tasks, but used to familiarize users with the task. Therefore, this condition was performed only in the 2D condition. Results showed that for pointing all participants used a pointing finger to accomplish this task. For both the mono and stereo visualizations, similar gestures were performed for positioning and selecting an object.

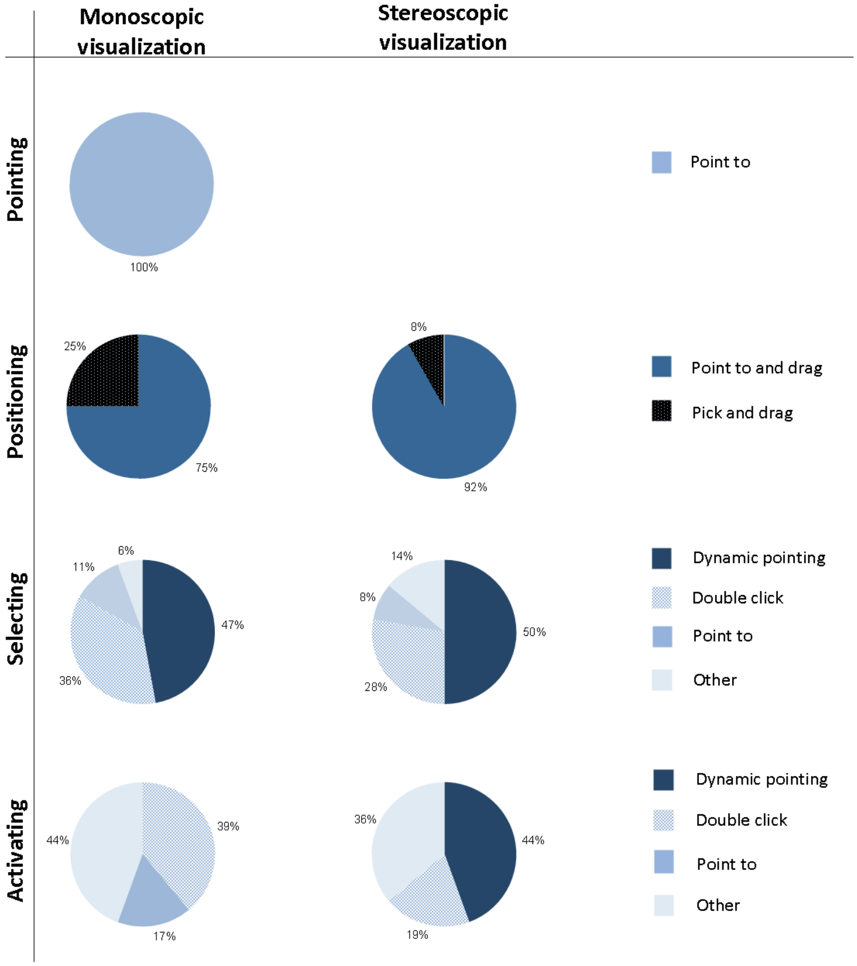


Figure 18. *Gestures users performed interacting with monoscopic and stereoscopic visualizations for the tasks pointing, positioning, selecting and activating an object.*

When users were asked to change the position of an object, users used 'point to and drag' in most of the cases in both stereo (91%) and mono (75%) visualizations. The 'point and drag' gesture was performed by pointing at an object and moving it towards the new position. An alternative gesture used instead of 'point and drag' was a 'pick and drag' gesture (tapping two fingers and moving the object to the new position), however this gesture was performed less frequently (see Figure 18). As shown in Figure 18, when selecting an object, the majority of participants used 'dynamic pointing' in both stereo and mono visualizations. This gesture was performed by moving a pointing finger toward the screen and back. The second most frequently chosen gesture was selecting an object by using a 'double click' gesture in, which is typically performed with a finger making two small movements toward the screen. Additionally some users used a pointing finger to select an object.

Activating an object revealed more variability between mono and stereo visualizations (see Figure 18). For stereo visualizations participants used 'dynamic pointing' in 44 percent of the cases, whereas in the mono condition a 'double click' gesture was used most frequently. The second most frequently used gesture in the stereo condition was a double click gesture (19%), whereas in mono visualizations a pointing finger (17%) was used most frequently. The results further revealed a high percentage in the category 'other', showing gestures such as clapping of hands, flat hand pushes, wiping movements (mono) or waving, grabbing movements or a stop sign, e.g., a full hand in front of the display, (stereo) to activate the volume.

When participants were asked to rotate the object, 58 percent of the users used a swiping gesture (i.e. a horizontal movement of the hand or finger in front of the screen) in the stereo condition, whereas during mono visualizations this was used in only 28 percent of the cases (see Figure 19). The most frequently occurring gesture during mono visualizations was a 'turn arm' (36%), which looks similar to holding a paper in front of you with both hands while rotating it along the Cartesian z-axis. A turn of the wrist was used in 19 percent of the cases - users only rotated their hand in the direction in which they wanted to rotate the object. In the stereo visualization, 'point to and drag' was performed in 14 percent of the cases to rotate an object. In the category 'other', gestures such as, grasping an object on the left or right side, or pointing to one of the sides were used by some participants.

For zooming-in on an image, the results showed a different ordering in gestures for the mono vs. stereo conditions (see Figure 19). In addition, the stereo condition showed more variability in gesture behaviour than the mono condition. In the mono condition, users most frequently used a 'reverse pinch' gesture (42%). The reverse pinch was performed by opening two fingers as if one were interacting with a touch screen.

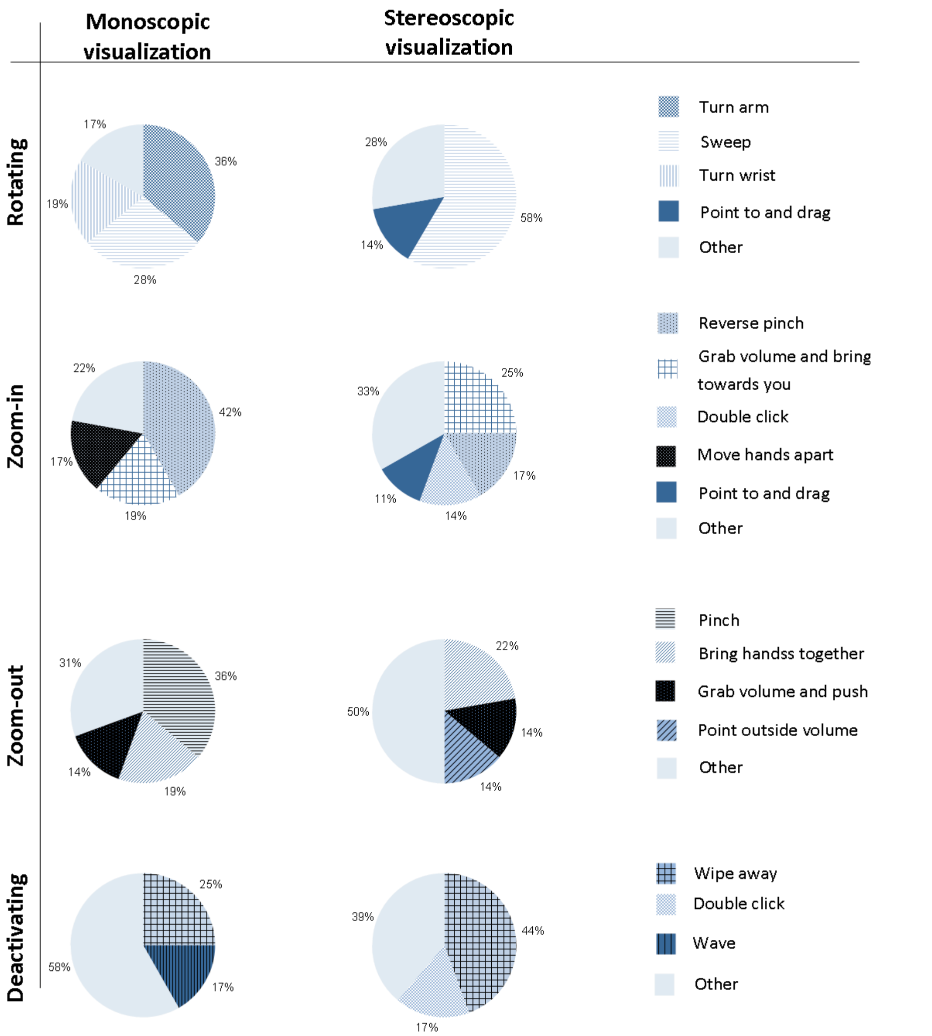


Figure 19. *Gestures users performed interacting with monoscopic and stereoscopic visualizations performing gestures for the tasks rotating, zoom-in, zoom-out and deactivating*

The second most frequently chosen alternative was grabbing the volume and bringing it towards oneself' (19%), and 'moving hands apart' (17%). For stereo visualizations participants used 'grabbing the volume and bringing it towards oneself' (25%) most frequently, followed by 'reverse pinch' (17%), 'double click' (14%), and 'dynamic pointing' (11%). Other gestures performed in the mono condition included pushing while opening the hand, or pinching while moving the arm from the screen. In stereo some users performed gestures such as moving the hands apart or pointing a finger to zoom into an object.

Zooming-out also produced more variability in the stereo condition than without stereo. For zooming out the 'pinch' gesture was used most frequently (36%) in the mono condition, which is performed by moving two fingers towards each other. Other popular alternatives were 'bring hands together' (19%) and 'grab volume and push' (14%). In the stereo condition, users most frequently used 'bring hands together' (22%), 'grab volume and push' (14%), and 'point outside volume' (14%). Less frequently occurring gestures in stereo were the 'double click', 'point' and 'pinch' gestures.

The last category was deactivating the volume, such that it does not respond to any gestures. Results showed that in both mono and stereo visualizations participants most frequently used the 'wipe away' gesture (see Figure 19), which is performed by moving the hand horizontally in front of the screen, or from the top to the bottom of the screen. The difference between the wipe away and the sweep gesture is that the wipe away gesture is performed in one single movement; i.e., the hand does not return to the starting position. The sweep gesture, on the other hand, returned to the position from which the gesture started. In the stereo condition, the 'double click' gesture was used in 17 percent of the cases, whereas in the mono condition participants used a 'wave' gesture in 17 percent of the cases. Other gestures performed by participants were 'point to', 'stop sign', 'double click' in the mono condition, and 'wave', 'stop sign', and 'dynamic pointing' in the stereo condition.

This functional classification of gestures revealed both variability as well as considerable similarity in the type of gestures used by participants. However, it should be noted that we quite broadly defined gesture categories, thus the gesturing behavior within one category still varied somewhat per participant and per visualization mode. Moreover, the number of hands or fingers, as well as the execution of the gestures varied between participants. For example, 'dynamic pointing' and the 'double click' gesture can be performed with one finger, or using the full hand to point or double click. For positioning, selecting, and activating the majority of users performed the gesture using one finger. Typical gestures that were performed with one full hand were: 'wave', 'sweep', 'grasp object and bring towards you', and 'wipe away' gestures. The 'pinch' and 'reverse pinch' gestures were performed either with two fingers of one hand or one full hand.

3.2.3 Summary of results

Embodied interaction assumes that the interaction makes use of our body representations and sensory-motor coupling used in daily life. Developing gesture-based interaction systems in which gestures are based on what can be tracked by the current state-of-the-art gesture trackers, is therefore not necessarily embodied. Therefore, in the current experiment we studied the type and style of gestures from a user-centered perspective, exploring what gestures participants naturally make when they were unconstrained by technology. Tasks used in this study were all related to manipulating 3D volumes, e.g., rotating, zooming-in and out, positioning, selecting, activating, pointing and deactivating. Since stereoscopic displays are increasingly commonplace for various applications, we were interested to learn whether stereoscopic visualizations elicited similar gestures compared to monoscopic visualizations. The first thing we observed was that a number of basic gesture types have become quite familiar to people, as part of the conventions of interacting with touch-based interfaces such as used in smartphones and tablets. For example, many participants used a "sweeping" movement to rotate images around the vertical axis – in line with the horizontal wiping motions of the fingers to go to a new page on a touch-sensitive device. Similarly, the "reverse pinch" that could be observed when zooming in, resonates with the convention on many touch-sensitive devices to use pinching, i.e., expanding the placement of two fingers, to contract or expand a displayed image. Despite such apparent conventions, the actual execution of the gestures varied substantially per participant, even when participants appeared to perform basically the same gesture (e.g., pinching or sweeping). For example, some users used their thumb and pointing finger while pinching, whereas others used their thumb and all other fingers.

In line with our hypothesis, we found a number of differences in gestures between mono and stereo modes of visualization. For visualizations without stereo, the gestures used for activating, zooming-in and zooming-out were comparable with traditional interaction methods using a desktop metaphor (e.g., double click, pinch). For stereo visualizations, however, those gestures were more spatial in nature, such as 'sweeping' a volume or 'grabbing' a volume and pulling it towards oneself or pushing it away. In addition, for zooming-in and zooming-out, we observed more variability in gestures during the stereo visualization than in mono visualization. However, for rotation and deactivation, the effect was reversed; users were more consistent in the stereo condition, eliciting gestures such as sweeping and wiping away. On the other hand, tasks such as positioning and selecting did not show much variation between mono and stereo modes of visualization.

Contrary to our expectations, gestures in stereo were performed less often with two hands compared to gestures in the non-stereo visualization. In addition, the whole hand was used less frequently in the stereo condition compared to non-stereo visualizations. In the 3D setting participants interacted more frequently with one finger than in the 2D setting.

Current findings were used to inform the design of the gesture tracker employed in Experiment 5. To this end, the results concerning commonly used gestures as well as how the gestures were executed by the users (based on video-recordings) during this experiment were communicated to Fraunhofer HHI11. To limit the number of gestures that could be tracked by the prototype, the HELIUM3D project team decided to implement five tasks: point, rotate, zoom-in and zoom-out a 3D volume and deactivate the gesture tracker. For each of these five tasks, we selected the two most frequently occurring gestures to interact with a 3D volume (see Figure 18 and Figure 19). These results were communicated to Fraunhofer HHI, that inspected the video-recordings to gain more detailed insights in the actual performance of the gestures and selected the gestures that could be accurately detected by the prototype gesture-tracker. From this selection, the 'grab volume and bring towards you' (for zooming-in), 'grab volume and push' (for zooming-out) and 'point to' (for pointing) followed directly from the results of the current study. As discussed by Beurden, IJsselsteijn & Hopf (2011), due to technological limitations the 'sweep' gesture (for rotating) could not be implemented as such. In addition, the second most frequently performed alternative for rotating 'point to and drag' was very similar to the gesture used for pointing and could therefore not be used to accurately distinguish between these tasks. Consequently, for rotating a different gesture (i.e., moving two spread fingers to the left and right) was implemented, which could be accurately recognized by the gesture-tracker. Based on a similar reasoning, for deactivating it was decided to use the reverse of grabbing (i.e., open hand), which was also different than the two most frequently occurring gestures in the current study (see Figure 19). In Experiment 5, this prototype gesture tracker was used to compare the user experience of gesture-based interaction with mouse-based interaction.

3.3 Experiment 5: User experience comparing gesture- and mouse-based interaction

Gesture-based, embodied interaction is assumed to have several benefits over controller-based, non-embodied interaction. Potential advantages relate to the naturalness of interaction, ease of learning new interactions, more visceral experience and hedonic qualities. In the current experiment, we wanted to test these expectations. Naturally, a test

[11] *Fraunhofer HHI was one of the partners in the FP7 HELIUM3D project and responsible for the technical implementation of the gesture tracker.*

such as this one requires a broader perspective on user experience than that employed in previous literature. We therefore incorporated indicators of both pragmatic and hedonic qualities in the evaluation of interaction methods. Pragmatic quality addresses factors related to effective and efficient goal-achievement, in other words, usability in the traditional sense; Hedonic quality refers to the extent to which a system is experienced as fun, original, interesting, engaging, and personally relevant. In order to create a fair comparison between mouse and gesture-based interaction, the gestures should be intuitive and easy to learn. Based on findings in Experiment 4, we defined a gesture set that was implemented in a tracker developed by Fraunhofer HHI (Beurden et al. 2011), with which users could rotate, zoom in and out, and point towards objects. In the current experiment the performance and experience of gesture-based interaction was compared to mouse-based interaction, focusing on pragmatic and/or performance aspects as well as hedonic qualities. In addition, we also measured body discomfort, since interacting using body movement can increase body fatigue.

3.3.1 Method

Design

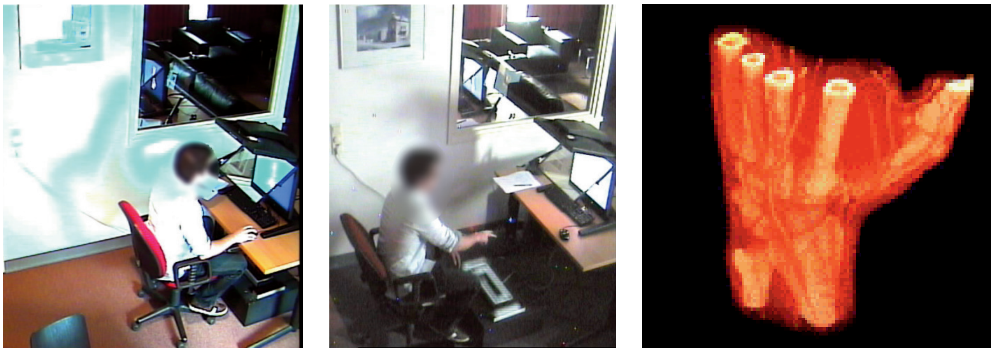
The experiment followed a one-factor (Interaction method: mouse vs. gestures), within-groups design, with indicators of usability, discomfort, and both pragmatic and hedonic quality as dependent variables. Interaction method was counterbalanced to avoid order effects.

Participants

Nineteen participants, (11 males and 8 females), between 19 and 35 years of age, all with normal or corrected to normal vision, took part in this study. All participants had stereo acuity better than 40 seconds of arc, tested with the Randot® stereotest. Participants were recruited from a database containing both students as well as individuals unrelated to the university.

Stimuli

The stimulus presented to the participants was a stereoscopic image of the internal structures of a hand, see Figure 20a. Participants freely explored this 3D object presented on the screen using rotation, zooming in and out, or pointing to a specific part of the volume. The same task and stimuli were used in both the mouse and gesture condition.



(a) (b) (c)

Figure 20. *Example of the content used in this experiment (a); Setup in which users interact with the mouse (b); Setup in which users interact via gestures (c).*

Setting and apparatus

The experiment was carried out at the Uselab of the Human-Technology-Interaction group at Eindhoven University of Technology. The stimuli were displayed on a Planar SD2020 stereoscopic display, with a resolution of 1200 x 1600 pixels. The stereo view on this display was created by a half-silvered mirror and participants wore polarized glasses to separate the left and the right eye views. The disparity level in the mouse condition was approximately 45 min of arc, whereas in the gesture condition the disparity level was 23 min of arc. The disparity level varied between these two conditions because the viewing distance was larger in the gesture condition due to the technical set up as shown in Figure 20b and Figure 20c. The gesture tracker set-up had two cameras (stereo approach) providing high accuracy in the three Cartesian coordinates x, y and z. The cameras were equipped with infrared filters eliminating visible light. Infrared light sources transmitted synchronized light pulses illuminating the captured objects. The cameras and infrared light sources of the gesture tracker were placed on the floor and detected hand movements from below. The software implementation used a set of modules suitable for detecting basic hand characteristics and identifying specific shapes such as fingers and the center of the palm. The gesture detection software uses a combination of shape identification and real-time position measurements. In the mouse condition, participants were seated approximately 75 cm from the display, whereas in the gesture condition the viewing distance was around 150 cm. The distance was larger because the gesture technology was positioned on the floor in front of the display, to avoid unwanted reflections from the table surface.

A disadvantage of this set up is that the viewing distance is larger than the ideal viewing distance for this type of display. Furthermore, the ambient light condition during the gesture evaluation was slightly dimmed, to avoid reflections of the lighting from luminaires mounted on the ceiling.

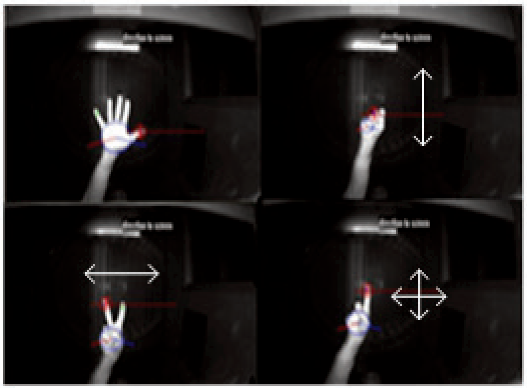


Figure 21. *Gestures implemented in the gesture tracker. An open hand (upper left panel) stops the interaction, moving a closed hand towards or away from the display (upper right panel) signifies zooming in and out the volume. Moving two fingers spread out in a v-shape (lower left panel) rotated the volume, and a pointed finger (lower right panel) moved the cursor in three dimensions.*

The set of gestures had been determined during the user requirement study reported in Experiment 4. The gestures implemented for the gesture tracker are shown in Figure 21. An open hand gesture (Figure 21, upper left panel) stopped the interaction. Zooming in and out was gestured with a fist, moving towards the display (zooming in) and away from the display (zooming out; Figure 21, upper right panel). Moving two spread fingers left or right (Figure 21, lower left panel; limited to the horizontal direction) rotated the 3D volume around the y-axis in the corresponding direction. With a pointed finger (Figure 21, lower right panel) the user could move the cursor in all three dimensions.

In the mouse condition, participants zoomed in or out by pressing the right mouse button and moving the mouse up or down. Pressing the left mouse button and moving the mouse left and right rotated the 3D volume around the y-axis (again, limited to the horizontal direction). Pointing was performed using the standard mouse cursor, which was shown in the application.

Measures

In the current study, user experience was measured with self-reports probing body discomfort (Corlett & Bishop, 1976), usability (Hornbaek, 2006), and pragmatic and hedonic quality (Hassenzahl, 2004). Participants evaluated physical discomfort indicating their perceived fatigue, perceived exertion, and perceived pain for various upper body parts (e.g., shoulder, upper arm, lower arm, hand) using the Borg scale (Borg, 1982). Current usability questionnaires focus on user interfaces (i.e., the interface you see on the screen, like windows) and not on interaction methods (i.e., the method that is used to interact with the interface like gesture or mouse-based interaction). Although in daily routine both the interaction technology and the interface determine persons' overall experience, we focused on participants' experience of the interaction technology when manipulating a 3D object. Therefore, we selected ten relevant items from the QUIS questionnaire (Chin, Diehl & Norman, 1988) to assess the usability of the interaction technology. A Principle Axis Factoring analysis with Oblimin rotation revealed four factors: perceived performance, ease of learning, fun and perceived experience. The perceived performance subscale consisted of three items (efficiency, speed and accuracy) and was internally consistent with $\alpha = .80$. Ease of learning consisted of two items (memorability and learnability) with an internal consistency of $\alpha = .61$. Fun was assessed with one item (fun). Perceived experience consisted of four items (impression, practicality, naturalness, satisfaction) with an internal consistency of $\alpha = .82$. Scores were computed using the average of all items for each subscale. These four values were used as indicators of usability. In addition, overall usability of the interaction methods was measured with 10 items ($\alpha = .88$) based on the SUS questionnaire (Brooke, 1996). Note that we slightly adjusted the questions to better fit to the interaction method used in the current study. Examples of the items are: 'I think that I would like to use this interaction method frequently' and I found the interaction method very cumbersome to use'. Hedonic and pragmatic qualities were assessed using the AttrakDiff questionnaire consisting of 21 semantic differential items, e.g., bad–good and easy-hard, rated on 7-point response scales (Hassenzahl, 2004). This questionnaire consists of four subscales; pragmatic quality, hedonic quality-identification, hedonic quality-stimulation and attractiveness, each containing seven items. Pragmatic quality (PQ; items: technical-human, complicated-simple, impractical-practical, cumbersome-direct, unpredictable-predictable, confusing-clear structured, and unruly-manageable) was internally consistent with $\alpha = .86$, hedonic quality-stimulation (HQS; items: typical-original, standard-creative, cautious-courageous, conservative-innovative, lame-exciting, easy-challenging, ordinary-new) was internally consistent with $\alpha = .95$, hedonic quality-identification (HQI; items: isolating-connective, amateurish-professional, styleless-stylish/classy, cheap-valuable, noninclusive-inclusive, takes me distant from people-brings me closer to people, unpresentable-presentable) was internally consistent with $\alpha = .70$, and attractiveness (ATT; items: ugly-beautiful, bad-good, rejecting-inviting, repulsive-appealing, discouraging-motivating, disagreeable-sympathic, unpleasant-pleasant) was internally consistent with $\alpha = .86$.

Procedure

On arrival at the Uselab, participants signed a consent form, and were informed that their actions were recorded during the experiment. Subsequently, participants were tested for their stereo acuity using the Randot® stereotest, followed by instructions regarding the experimental procedure and the questionnaires used during experiment. After participants were seated in front of the display, they were instructed how to perform the task and offered time to practice the technology they would start with. Half of the participants started with the gesture-based interface, the other half with the mouse. Once users were familiar with the interaction technology they carried out the experimental task for 5 minutes, followed by questionnaires. After participants completed the questionnaires, they evaluated the other interface technology following the same procedure. Users had sufficient time to practice these interaction methods. After completing the evaluation of both the gesture-based interaction and mouse-based interaction, a short interview was administered in which participants further elaborated on their experiences with the interaction methods. At the end of the experiment, participants were thanked for their participation. The experimental procedure took between 45-60 minutes and participants received a compensation of ten euros for their time.

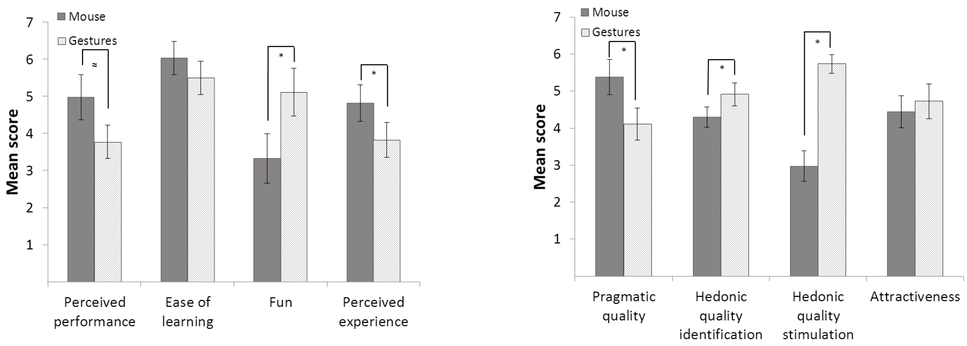
Statistical analysis

Paired-samples t-tests were performed to investigate the effects of interaction technology (mouse/gestures) on hedonic and pragmatic quality, usability, and body fatigue. In the current experiment, effect sizes were again reported in terms of partial η^2 and Cohen's *d*, (see section 2.5.1).

3.3.2 Results

Figure 22a presents the results of the usability indicators, assessed with items adopted from the QUIS; perceived performance, ease of learning, fun and perceived experience. Paired-samples t-tests revealed main effects of Interaction method on perceived performance, fun, and perceived experience. In terms of fun, gestures

($M = 5.11$, $SE = .33$) were evaluated better than the mouse ($M = 3.32$, $SE = .34$; $t(18) = 3.67$, $p < .01$, $d = 1.23$, partial $\eta2 = .43$).



(a) (b)

Figure 22. *Mean scores of various attributes using the mouse or gesture interaction in terms perceived performance, ease of learning, fun, and perceived experience (a) and pragmatic and hedonic quality (b). The bars indicated with a * are statistically significant (p < .01), and ≈ presents a non-significant trend (p < .10) Error bars depict 95% confidence intervals. Results showed that in term of pragmatic quality, perceived performance and perceived experience the mouse rendered higher scores. However, in terms of hedonic quality and fun, users preferred gesture-based interaction.*

However, in terms of perceived experience the mouse ($M = 4.82$, $SE = .25$) revealed better scores than gestures ($M = 3.82$, $SE = .24$; $t(18) = 2.61$, $p < .05$, $d = .94$, partial $\eta2 = .27$). Also perceived performance revealed higher scores for the mouse ($M = 4.97$, $SE = .31$) than for gestures ($M = 3.77$, $SE = .23$; $t(18) = -3.10$, $p < .01$, $d = 1.02$, partial $\eta2 = .35$). No significant difference was found on ease of learning $t(18) = 1.79$, $p = .09$, partial $\eta2 = .15$. Following the results of perceived performance and perceived experience, the SUS questionnaire showed higher scores for the mouse ($M = 83.3$, $SE = 3.88$) than for gestures ($M = 62.9$, $SE = 3.5$; $t(18) = -4.95$, $p < .001$, $d = 1.27$, partial $\eta2 = .58$). In Figure 22b, the results of pragmatic and hedonic quality indicators are presented. A paired-samples t-test showed that for pragmatic quality, the mouse ($M = 5.38$, $SE = .24$) had a significantly higher score than gestures ($M = 4.11$, $SE = .22$; $t(18) = -4.64$, $p < .001$, $d = 1.26$, partial $\eta2 = .54$). However, as shown in Figure 22, gestures were preferred in terms of hedonic quality. For hedonic quality-identification, gestures ($M = 4.91$, $SE = .16$) revealed significantly higher scores than the mouse ($M = 4.29$, $SE = .14$; $t(18) = 3.61$, $p < .001$, $d = .95$, partial $\eta2 = .42$). For hedonic quality-stimulation, this effect was even more pronounced; gestures

(M = 5.73, SE = .21) were rated significantly better than the mouse (M = 2.97, SE = .13; $t(18) = 11.87$, $p < .001$, d = 3.72, partial η2 = .89). In terms of attractiveness, the results did not reveal any difference between the mouse (M = 4.72, SE = .24) and gestures (M = 4.44, SE = .22; $t(18) = .80$, p = .40. partial η2 = .03).

Effects of Interaction method on subscales of body discomfort were also analyzed with paired samples t-tests. No significant differences in perceived pain emerged between the mouse (M = .39, SE = .18) and gestures (M = .65, SE = .30; $t(18) = 1.67$, p = .11, partial η^2 = .13). However, gestures (M = 2.15, SE = .32) resulted in significantly higher levels of perceived exertion than the mouse (M = 1.32, SE = .29; $t(18) = 4.50$, $p < .001$, d = .62, partial η2 = .53). A similar effect was found in terms of fatigue, gestures (M = 1.52, SE = .34) revealed higher scores than the mouse (M = .76, SE = .21; $t(18) = 4.82$, $p < .001$, d = .63, partial η2 = .56). In the shoulder and upper arm, high levels of fatigue and exertion were reported when using gestures. Although the mouse revealed less discomfort than gesture-based interaction, participants experienced moderate levels of fatigue and exertion in their hand when using the mouse. Results of post-test interviews highlighted fatigue, recognition errors and accuracy as the most important disadvantages of the gesture tracker. Perceived advantages of the gesture tracker included its fun, ease of use, naturalness, and greater involvement in the task using gestures than using the mouse.

3.3.3 *Summary of results*

In this experiment, we compared users' experience of interacting via gestures (embodied) with that using the mouse (non-embodied). The experience of users was measured in terms of body discomfort, usability, pragmatic quality and hedonic qualities. In this study, the mouse outperformed gesture-based interaction on perceived performance, pragmatic quality, perceived experience, and SUS score with large effect sizes (cohen's d) ranging from d = .94 to d = 1.27. This is in line with post-test interviews, where participants perceived the interaction using the gesture tracker as less accurate and slower than the mouse. Although all users were unfamiliar with gesture-based interaction, in terms of ease of learning both the mouse and gestures revealed similar scores. In addition, the results in our study showed that interacting through gestures was experienced as more fun, original, interesting, engaging, and personally relevant revealing higher scores in terms of hedonic quality identification and stimulation (having a large effect; d = .95 and d = 3.72 respectively). This effect may be attributed to several factors. The first potential explanation for the difference in evaluation between gesture-based interaction and mouse-based interaction is the embodied nature of gesture-based interaction. A second explanation is that people interacted with technology without the need to press buttons and hold devices. A third, alternative explanation for this result might be the novelty effect, since – at the time of this study - interacting with digital technology without any device was certainly unusual and not commercially available (note: the experiment was performed before Microsoft's introduction of the Kinect interaction technology). As discussed by Karapanos et al. (2009), hedonic aspects that are concerned with being stimulated are especially important in the early stages of technology use, and will become less important when the product is used for a longer period of time. However, aspects related to hedonic quality identification showed to increase with prolonged use. The future will reveal how experiences evolve when gesture interaction becomes more commonly available. In addition, due to the technical limitations the viewing distance and consequently the disparity level were different between the mouse and gesture conditions. Therefore in Experiment 6 we will improve the experimental setup and compare gesture-based interaction with interacting using the Wii (i.e., the Wii is device-based like the mouse, but embodied, in contrast to the mouse).

3.4 Experiment 6: User experience of device and deviceless embodied interaction

In this experiment, we compared gesture-based interaction with interaction using the Wii. In Experiment 5, we hypothesized that the higher levels of hedonic qualities reported for gesture-based interaction may have emerged due to the embodied nature of the interaction, or the fact that people did not hold a device. In the current experiment, we isolate the effect of 'devicelessness', by contrasting two embodied interaction methods; with a device (Wii) and without a device (Gestures). In addition, we improved the experimental setup compared to Experiment 5, using the same viewing distance and disparity levels for both interaction methods. Results of the current experiment will provide a better understanding of the effects of device and deviceless embodied interaction in terms user experience.

3.4.1 Method

Design

The current study investigated user experiences following a one-factor (Interaction method: Wii vs. gestures) within-groups design, with subjective indicators of usability, discomfort, and both pragmatic and hedonic quality as well as objective task performance as dependent variables. In this experiment, objective performance was measured in terms of selected icons was also recorded. Interaction method was counterbalanced to avoid order effects.

Participants

Nineteen participants, (12 males and 7 females), between 19 and 32 years of age, all with normal or corrected to normal vision, took part in this experiment. All participants had a stereo acuity better than 40 seconds of arc, tested with the Randot® stereotest. Participants were recruited from a database containing both students as well as individuals unrelated to the university.

Task

Participants evaluated both interaction technologies performing an icon selection task for five minutes (see Figure 23a). During this task users navigated to a flickering icon within a set of icons presented on the screen, and selected this icon using the Wii (Figure 23b) or gestures (Figure 23c). The icons were arranged in a spherical shape with a selection square located in the center of the screen. To select an icon located on left side of the selection square, participants moved their hand or the Wii controller towards the left; to select an icon located above the selection square, they moved up, etc. The icon was selected by a circular movement of the hand (gesture condition), or pressing the shoot button (Wii condition).

(a) (b) (c)

Figure 23. *A screenshot from the selection task (a); a participant interacting using the Wii (b); a participant interacting using gestures (c)*

Chapter 3

Setting and apparatus

The experiment was carried out at the Uselab of the Human-Technology Interaction group at Eindhoven University of Technology. The task was displayed on a 55″ Samsung 7700 3D display, with a resolution of 1920 x 1080. The stereo view on this display was created wearing shutter glasses. In both conditions, the distance from the screen was around two meters, and the disparity of the task was approximately 40 min of arc. The application was programmed in Java and used Flash to run the application. To optimize the performance of the gesture tracker the room was brightly illuminated, such that the frame rate of the webcam ran around 25 fps. The gesture tracker used in the current experiment consisted of a Trust 1.3 megapixel camera with a resolution of 1280 x 1024. The software allowed detectionof hand motion, responding both to translations along the horizontal and vertical axis, and circular gestures). Participants navigated through the icons by moving their arm to the left, right, up and down. For each gesture, the hand had to return to the start position in order to complete the gesture. By making a circular movement with the hand, participants selected an icon (see Figure 24). For interacting with the Wii, we implemented movements comparable with the gesture technology using the position sensor of the Wii console. Participants navigated through the icons by making left, right, up and down movements relative to a central area. The central area was relative and defined according to the initial position in which the user held the Wii controller. By moving the controller away from this central area (up, down, left or right), corresponding actions resulted on the screen. The shoot button on the Wii was used to select an item.

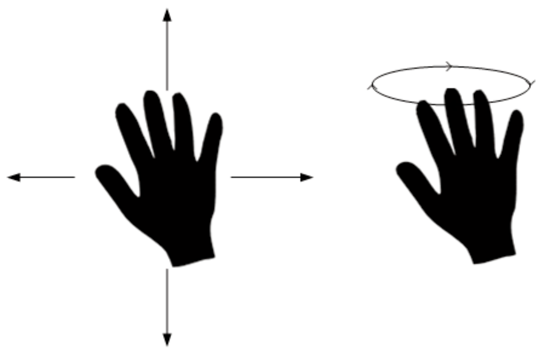


Figure 24. *Gestures implemented in the gesture tracker. Users controlled a highlighted icon by moving their hand in 4 directions. To select an icon users made a circle in the air*

Measures

The dependent measures were identical to the ones described in Experiment 5: body discomfort (fatigue, pain and exertion), usability (QUIS: perceived performance (α = .85), fun (1 item), ease of learning (α = .57), perceived experience (α = .92), and general usability score with the SUS (α = .77), pragmatic quality (PQ, α = .67), hedonic quality (hedonic quality -stimulation (HQS, α = .86); hedonic quality-identification (HQI, α = .72), and attractiveness (ATT, α =.83)). In addition, the number of selected icons during the five minutes (after the practice trials) was recorded as an objective measure of performance. At the end of the experiment, we asked participants about their experience using a short interview.

Procedure

The procedure was similar as described in Experiment 5, with fifty percent of the participants starting with the Wii and the remaining fifty percent with the gesture-based interaction. Before the experiment participants had sufficient time to practice the interaction. The experimental procedure took between 45-60 minutes and participants received a compensation of ten euros for their time.

Statistical analysis

Paired-samples t-tests were performed to investigate the effects of interaction technology (Wii vs. gestures) on hedonic and pragmatic quality, usability, and body fatigue. In the current experiment effect sizes were again reported in terms of partial η^2 and Cohen's *d*, (see paragraph 2.5.1)

3.4.2 Results

Figure 25a presents the results of the usability indicators for perceived performance, ease of learning, fun and perceived experience for the comparisons between the Wii and gestures. A paired samples t-test showed that the scores in terms of perceived performance were significantly higher for the Wii (M = 3.88, SE = .35) than for the gestures (M = 2.83, SE = .21; $t(18)$ = -3.09, $p < .01$, d = .86, partial $\eta2$ = .35). This is in line with the number of selected icons (even with larger effect sizes), showing that interacting with the Wii console (M = 45.26, SE = 5.31) resulted in a better performance than using gestures (M = 14.94, SE = 1.10); $t(18)$ = -6.01, $p < .001$, d = 2.17, partial $\eta2$ = .67). Perceived performance and objective performance were correlated r = .35, ($p < .05$). In addition, as shown in Figure 25a, results revealed higher scores in terms of perceived experience for the Wii (M = 4.49, SE = .27) than for gestures (M = 3.47, SE = .32; $t(18) = 2.78$, $p < .05$, d = .79, partial η2 = .30).

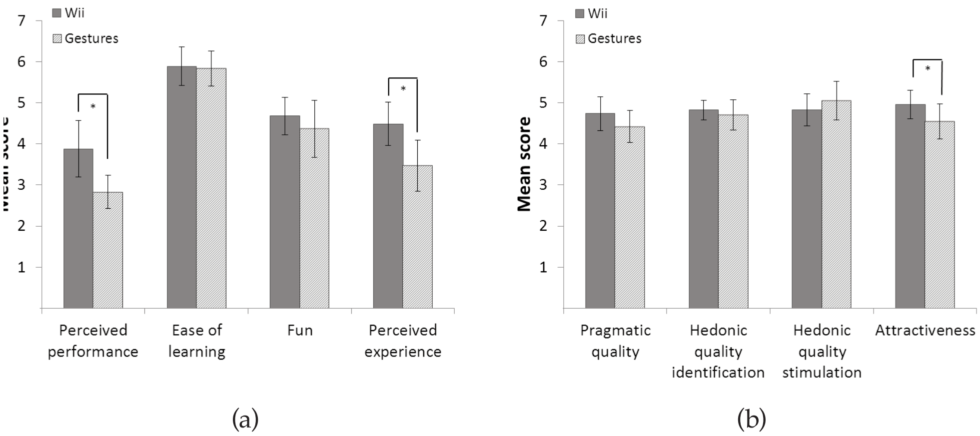


Figure 25. *Mean scores of various attributes using the Wii or gesture interaction. The bars indicated with a * are statistically significant (p < .01). The left panel shows the results in terms of perceived performance, ease of learning fun and perceived experience (a) and in the right panel the results in terms of pragmatic and hedonic quality and attractiveness (b). Error bars depict 95% confidence intervals.*

The Wii and gestures revealed similar scores in terms of ease of learning ($t(18) = .21$, $p = .80$, partial η2 < .01), and fun ($t(18) = .90$, $p = .38$, partial η2 = .04). Also the SUS revealed no significant different scores between the Wii (M = 72.11, SE = 3.83) and gestures (M = 66.18, SE = 3.27; $t(18) = -1.58$, $p = .13$, partial η2 = .12). Figure 25b shows the results in terms of pragmatic quality, hedonic quality and attractiveness for the Wii vs. gestures. A paired samples t-test showed that the Wii (M = 4.96, SE = .18) was evaluated as more attractive than gesture-based interaction (M = 4.55, SE = .22; $t(18) = -2.38$, $p < .05$, d = .47, partial η2 = .24). The results did not reveal a difference between the Wii and gestures in terms of pragmatic quality ($t(18) = -1.20$, $p = .24$, partial η2 = .07), hedonic quality-stimulation ($t(18) = 1.66$, $p = .11$, partial η2 = .13) and hedonic quality-identification ($t(18) = -0.81$, $p = .43$, partial η2 = .03).

In terms of body discomfort all three indicators (e.g. fatigue, exertion and pain) revealed a significant main effect of interaction device. A paired samples t-test, with exertion as dependent variable, using the average scores of all body parts, showed a significantly higher level of exertion using gestures (M = 2.36, SE = .38) compared to the

Wii (M = 1.7, SE = .29; $t(18) = 2.42$, $p < .05$, d = .45, partial $\eta2$ = .25). Furthermore in terms of fatigue, the gesture condition (M = 1.89, SE = .27) revealed higher levels of fatigue than the Wii (M = 1.28, SE = .22; $t(18) = 2.96$, $p < .01$, d = .57, partial $\eta2$ = .33). In terms of perceived pain scores were also higher using gestures (M = 1.14, SE = .14) compared to the Wii (M = .48, SE = .14; $t(18) = 3.17$, $p < .01$, d = 1.08, partial $\eta2$ = .36). A closer look at the scores of the individual body parts indicated, that overall discomfort was the largest in the shoulder and upper arm when using gestures. Although the Wii revealed lower levels of fatigue, users experienced "moderate" to "somewhat strong" levels of fatigue in their wrist and lower arm when using this interaction method.

Results of post-test interviews revealed similar results as reported in Experiment 5, reporting fatigue, recognition errors, and low accuracy as disadvantage for interacting with gestures compared to the Wii. In addition, users mentioned unwanted interactions (system responses to gestures that were not intended as input, e.g., scratching one's nose) and slow gesture interpretation as additional disadvantages. Perceived advantages of the gesture tracker related to fun, ease of use, naturalness and involvement, similar to the findings reported in Experiment 5.

Chapter 3

3.4.3 Summary of results

In Experiment 5, we observed that gestures rendered higher scores in terms of hedonic qualities compared to mouse-based interaction, however it remained unclear whether these differences should be attributed to the embodied character of the interaction, or instead to the fact that no device was needed. In addition, limitations of the technical set-up resulted in different viewing distances between the mouse and gestures-based interaction, and consequently a difference in perceived disparity levels could be confounding. Therefore, in Experiment 6 we studied the effect of holding a handheld device vs. user experience of deviceless interaction in a similar set-up. The results revealed no differences between gesture- and Wii-based interaction in terms of fun and hedonic quality. Apparently, holding a device or being able to interact without a device did not change the experience of users in terms of fun and hedonic quality. Nevertheless, the Wii was perceived as more attractive than gestures, with medium effect sizes (d = .47). Moreover, using the Wii resulted in better objective performance than did gesture interaction (d = .49). This result was in line with the subjective indicators concerning perceived performance and perceived experience, although with larger effect sizes (d = .86 and d = .79 respectively). Although, both pragmatic quality and SUS scores appeared slightly higher when interacting with the Wii, this result was not statistically significant. Pragmatic quality as a factor consists of items representing users' impression of the performance (e.g., practicality, complexity manageability) as well as items, which are less dependent on the actual performance (e.g., technical-human, directness of the interaction). This may explain why gesture and Wii-based interaction did not reveal a difference in terms of pragmatic quality. A similar reasoning can be used for the SUS questionnaire of which items such as ease of use, ease of learning, and cumbersomeness are less dependent on the actual performance of the system. Results further showed that both gesture and Wii-based interaction were equally easy to learn, which is not surprising since both the Wii and gestures used similar movements. In addition, the number of gestures that users had to memorize was low and therefore did not cause any problems. Although Wii- and gesture-based interaction are both embodied methods of interaction, the Wii revealed significantly less body discomfort than gesture-based interaction with effect sizes between $d = .45$ for exertion and $d = 1.08$ for pain. Although perceived pain resulted in a large effect size, the average score was 1.14 on a scale from 0 to 10, indicating only minor levels of perceived pain. The smaller body movements while interacting with the Wii compared to gesture-based interactions can explain this result. In addition, users interacting with the Wii often rested their elbows on their knees while interacting, whereas in the gesture condition the whole arm was used during the interaction.

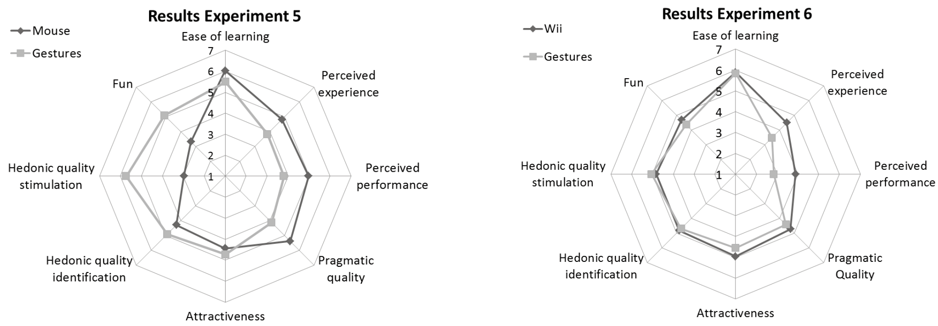
3.5 Discussion

Within HCI, there is increasing interest in extending the bandwidth of human-machine interaction and moving away from the constraints of the traditional keyboard/mouse interfaces. Two recent developments support the change of how we interact with technology. First, the increasing popularity of stereoscopic displays demands interaction methods in which users can intuitively interact with spatial content on the screen. Second, new sensing technologies are offering opportunities to engage the body during the interaction to a greater extent (such as the Nintendo Wii and Microsoft Kinect) than during traditional interaction. In addition, the domain of human-computer interaction has developed from productivity-oriented technologies in which performance was a key objective towards applications meant for entertainment, leisure and play. Therefore various authors have stressed the importance of incorporating measures of user experience that go beyond traditional usability measures, such as fun, hedonic qualities and emotions (Hassenzahl, 2004; Norman, 2004; Tractinski et al., 2000). Although in the area of Graphical User Interfaces these measures are now being accepted, for interaction methods and technology the main perspective is still very much performance-driven.

In the current chapter, a user-centered perspective was taken instead of focusing on what is technologically feasible.

In this chapter we gave special attention to gesture-based interaction, in which participants interact with mediated environments without being constrained by a controller. Gesture-based interaction is often seen as more natural than controller-based interaction, however current implementations are often driven by technological constraints. Therefore, in the first experiment we studied the movements users make naturally; i.e., without being constrained by technology, when manipulating objects on a display. In addition, we explored whether users used different gestures when interacting with stereoscopic displays compared to monoscopic displays. The gestures most frequently used in this first experiment were implemented in the gesture tracker used in Experiment 5. This experiment tested effects of gesture-based interaction in comparison to mouse-based interaction on user experience. Lastly, Experiment 6 compared gesture-based interaction to the Wii, an embodied yet device-based interaction method.

Results of Experiment 4 showed that for stereo visualizations, the gestures used for activating, zooming-in and zooming-out were more spatial in nature (e.g., 'sweeping' or ' grabbing a volume') than for non-stereo visualizations, in which gestures were comparable to traditional interaction methods using a desktop metaphor (e.g., 'double click', 'pinching'). Although in both mono and stereo conditions, two-handed gestures were performed less frequently than one-handed interaction, bimanual interaction was used more frequently in mono than in the stereo condition. Based on the current study we cannot draw firm conclusions, and future research should reveal if there is a difference between 2D and 3D visualizations on preference of bimanual interaction. However, as mentioned in Chapter 1, two-handed interaction might be used more frequently when users are asked to perform multiple tasks at the same time, such as changing the position and rotating an object at the same time. In Experiment 4 we asked users to perform one task at a time, which might explain the overall preference of participants for one-handed interaction. When designing and implementing gesture-based interactions we can take advantage of the fact that for some interactions (positioning and selecting), the gestures are relatively uniform. However, other, more complicated actions, such as rotating and zooming, showed more variability, thus making a "one-size-fits-all" implementation of such actions less intuitive for at least some of the users. Of course, in limited, well-specified tasks (e.g., browsing an EMR, without zooming or transforming the image, volume, or channel switching), the natural set of gestures may be more limited than in our study. Alternatively, gesture recognition software could incorporate a learning algorithm making it more robust to some of the between- and within-user variability. Moreover, as mentioned above, our results also demonstrated some differences in gesture-based interactions in relation to non-stereo versus stereo visualizations of the same content. Although this may partly be due to individual variation, it should be taken into account as a potentially relevant parameter in the design of future interaction systems utilizing 3D displays.



(a) (b)

Figure 26. *Results of both Experiment 5 and Experiment 6 showing differences in experience between mouse-based interaction and gesture-based interaction (a), and Wii-based interaction vs. gesture-based interaction (b). Results showed that the mouse performed better on the usability factors such as perceived experience, perceived performance, and pragmatic quality. However, in terms of fun and hedonic qualities, gestures revealed higher scores (a). Comparing the Wii and gestures showed similar scores in terms of fun and hedonic qualities, whereas in terms of perceived experience and perceived performance the Wii was preferred (b).*

In Experiments 5 and 6 we studied gesture-based interaction in terms of user experience, by comparing it to the mouse (Experiment 5) and the Wii (Experiment 6). The results of these two experiments are graphically presented in Figure 26. In both Experiments 5 and 6 the mouse and the Wii outperformed current gesture-based technology in terms of perceived performance and perceived experience (Cohen's d between .86 and 1.03). In line with these results, Experiment 5 revealed that pragmatic quality and the overall usability score (SUS) were larger for the mouse than for gesture-based interaction, with similar effects sizes (d = 1.26 and d = 1.27 respectively). In Experiment 6, no differences emerged between the Wii and gesture-based interaction in terms of pragmatic quality and overall usability (SUS). However, as shown in Figure 26a, Experiment 5 demonstrated that hedonic quality and fun were higher for embodied (i.e., gesture-based interaction) than non-embodied interaction technologies (i.e., mouse).

The largest effect was found for hedonic quality stimulation (d = 3.72) followed by fun (d = 1.23) and hedonic quality identification (d = .95). Figure 26b shows that the scores for hedonic quality and fun did not differ between the Wii and gesture-based interaction. The fact that in Experiment 6 the Wii scores were similar to those of the gesture-based interaction, indicates that holding a device or interacting with bare hands elicited similar experiences. This suggests that the differences between gesture-based and mouse-based interaction in Experiment 5, may not be attributed to the fact that gesture-based interaction does not require a device. Instead, it is more likely that the embodied nature of the interaction was responsible for the more positive experience of both gesture and Wii-based interaction. As shown in Figure 26 no difference emerged between gesture-based interaction and device-based interaction in terms of ease of learning. This result is not surprising, since participants only had to remember three gestures in the current set-up. We also asked users to reflect on their experience with the interaction, involving indicators such as practicality, naturalness and satisfaction (i.e., perceived experience), that are often used in usability evaluations. The results showed that, in line with Hassenzahl (2004), the score in terms of perceived experience closely matched the findings in terms of perceived performance and pragmatic quality (see Figure 26). This confirms our idea that for a better understanding of users' experience both pragmatic and hedonic qualities are needed. The higher scores in terms of hedonic quality and fun suggest that embodied interaction also affects users' visceral experience expressed in terms of enjoyment, personal identifications and stimulation. This is in line with studies comparing two game consoles (embodied vs., non-embodied), revealing more fun and higher level of engagement when gaming with embodied interaction devices such as the Wii (McGloin et al., 2011; Skalski et al., 2011).

From post-test interviews, we further learned that another potential benefit of gesture-based interaction is the higher feeling of involvement in the task, since the interaction felt more direct using gestures than when holding a device. A disadvantage of gesture-based interaction was the higher levels of perceived body discomfort, showing more body fatigue, body exertion, and pain when interacting with gestures. Although the Wii revealed lower levels of fatigue than gestures, users experienced "moderate" to "somewhat strong" levels of fatigue in their wrist and lower arm. This finding may be inherent to embodied interaction, which indeed does require more energy and induces more muscle strain – although surely interaction with the mouse also has its known repercussions on the body. Other disadvantages mentioned frequently were related to its accuracy and processing speed, and the occurrence of unintended actions due to misinterpretation of random movements. These issues are frequently discussed in studies on gesture-interaction (e.g., Wexelblat, 1998), and need considerable attention when further developing gesture-based interaction. Nevertheless, the increasing computing power and new developments in sensor technology will increase both accuracy and processing speed such that these problems may fade away as gesture-based interaction systems are developed further. In addition, due to increasing body fatigue, gestures will probably not replace current hand-held technologies completely, but will serve as an additional interaction method used for short term interaction such as changing the volume or rotating images.

3.5.1 Practical Implications

The results in this chapter showed that gesture based interaction positively affects user experiences in terms hedonic aspects compared to more traditional interaction styles such as the mouse. On the other hand, also embodied device-based interaction, such as the Wii, showed similar hedonic experiences as for gestures, although with higher scores in terms of pragmatic aspects. Indeed, there are still many challenges for gesture-based interaction, such as improving detection accuracy, and determining when a gesture starts and ends. Nevertheless interacting without the need of a controller has several potential advantages, such as flexibility for users interacting with displays or ambient technology as, for example in public or other multi-user environments. Additionally environments that have strict hygienic requirements, may be better suited for gestures-based than for device-based interaction.

Judging from the speed of innovations in gesture-based gaming, hedonic qualities are clearly relevant to designers in entertainment contexts. Professional applications appear to be slower in adopting such technologies, most likely indicating greater interest in performance and accuracy of interaction devices. Yet clearly, both categories need to be considered in any context, as performance obviously is relevant during play, and hedonic experiences may be more important in professional contexts than we currently realize (cf. Norman, 2002; Tractinsky, Katz & Ikar, 2000). It is clear that the broader perspective on user experiences employed in evaluating embodied interaction technologies can inform the design of such technologies, in terms of their strengths and weaknesses in comparison to more traditional interaction methods. Moreover, it allows designers to balance the full gamut of qualities of different interface alternatives, and offers them better-informed ways to optimize and tailor their design decisions to the specific context of the application.

CHAPTER 4

User experience of gesture-based interaction in a performance-oriented context

4.1 Introduction

Interaction in 3D concerns both the visualization of 3D content and the manipulation of it in natural/embodied ways. The previous chapters have addressed 3D visualization effects in relation to user control (Chapter 2), and embodied manipulation of 3D content (Chapter 3). Both chapters have also demonstrated the mutual dependency of visualization and interaction method. In daily life, we naturally interact with objects in three dimensions using our hands and body. Interacting with stereoscopic displays similarly requires interaction methods through which users can easily and intuitively manipulate objects in three dimensions. Embodied interaction (gesture or device-based) provides a promising alternative to the interaction methods commonly employed to interact with 2D content on monoscopic displays. In Chapter 3, we learned that hedonic aspects of user experience increased when users interacted in more embodied ways. Hedonic quality and enjoyment are important aspects for entertainment purposes, such as gaming and leisure applications. Also in professional contexts, embodied interaction may prove advantageous, since a positive affect facilitates creativity and cognitive flexibility (Ashbly, 1999; Davis, 2009; Isen, 2001). However, this effect only emerged as long as the task was interesting or important to the user (Ashbly, 1999; Davis, 2009; Isen, 2001). When tasks are dull, unpleasant or unimportant, positive affect might lead to impaired performance (Isen, 2001). Nevertheless, in tasks important for the user, a positive mood induced by embodied interaction might increase performance.

In Experiment 6, gesturing did not result in an increase in performance, whereas previous literature showed that gestures can facilitate learning and memory and increase performance on mental tasks (Goldin-Meadow, 2010; Goldin-Meadow, Nusbaum, Kelly & Wagner, 2001). For instance, Chu and Kita (2011) showed that when users were encouraged to use gestures while solving mental rotation tasks, users performed better compared to groups that were merely allowed, or even prohibited from using gestures. As another example, Wexler, Kosslyn and Berthoz (1998) demonstrated facilitation of mental rotation tasks through congruent movements. They showed that when (manual) rotation of a joystick was congruent with the direction of the Cooper-Shepard mental rotation task, the task was performed more accurately and rapidly compared to incongruent joystick control. Gesturing can also promote math learning among children and reflect the readiness to learn a task (Goldin-Meadow, 2010). In addition, gesturing may also decrease cognitive load. Participants who were able to gesture while explaining a math problem performed better on a secondary task (remembering words and letters) than participants who were not allowed to gesture (Golden-Meadow et al. 2001). To nuance these findings, one might argue that forcing users not to gesture may have added cognitive load and that potentially this was responsible for the decrease in performance. However, results reported by Chu and Kita (2011) showed that participants who were allowed to use gestures, but did not choose to do so, also performed less well on the secondary task.

It is important to note that the gestures investigated in these studies are of the type that occur naturally with speech and thought, and are therefore different from gestures used in human-computer interaction, which might not always be meaningful, intuitive, and congruent with the users' thoughts. A study performed by Cook & Colleagues (2011) showed that users' cognitive load was only lower when movements produced by them actually conveyed meaning. Meaningless arm movements did not result in better performance on a secondary task. This again illustrates the importance of taking a user-centered perspective when developing gesture-based interaction as discussed in Chapter 3. Therefore using gestures that are meaningful are not only easier to learn, but may also decrease cognitive load and support learning and thinking. All these aspects are at least as interesting and relevant in performance-oriented contexts (e.g., professional settings, education), as they are in a leisure context.

4.1.1 Rationale for the studies

In the current chapter, we will study embodied (gesture-based) interaction in a performance-oriented context by combining methodologies used in Chapters 2 and 3. This allows us to answer various research questions concerning stereoscopic displays as well as gesture-based interaction.

First, in the studies reported in Chapter 2 we tested effects of stereo visualization and user control, but only employed performance-related measures (completion times, accuracy and workload) to assess interaction, not taking the broader user experience perspective we advocated in Chapter 3. The current study therefore investigates the effects of stereo-visualization on both pragmatic and hedonic qualities.

Second, in Chapter 3 we studied user experience of gesture-based interfaces and compared it with the use of the mouse and the Wii. Results showed that both gestures and the Wii revealed high scores in terms of hedonic quality and enjoyment as compared to the non-embodied interaction with the mouse. Importantly, however, the context of this interaction was not explicitly performance-driven. The experience of users might be radically different when users are requested to complete tasks as rapidly and accurately as possible. In such contexts, the relevance of pragmatic qualities may start to outweigh that of hedonic quality. Moreover, from previous studies it was unclear which attributes of a user's experience (hedonic or pragmatic qualities) would impact user preference the most. The current study therefore included preference elicitation at the end of the experiment, to shed some light on the pragmatism-hedonism trade-off in a performance-oriented context. Furthermore, the focus on task performance also allowed us to investigate workload effects of gesture-based interaction. Since gestures decrease cognitive load when used in natural conversation (Goldin-Meadow et al. 2001), we hypothesized that gesture-based interaction might reveal similar results.

Third, in addition to pragmatic and hedonic qualities we also asked users to reflect on their positive and negative emotions. Since hedonic qualities are closely related to users' emotions we expected higher levels of positive affect and lower levels of negative affect for users interacting via gestures. Moreover, in addition to measures related to the experience of users while interacting with technology we also added a few questions regarding image quality. Previous studies (Lambooij et al., 2010) revealed that stereo visualization increases naturalness of images and increases perceived depth. In the current experiment, we were interested to see whether similar results would emerge in a performance-oriented context. In addition, Beerends and De Caluwe (1999) demonstrated a cross-modal interaction effect, by showing that image quality ratings are affected by sound quality. In this experiment, we therefore wanted to explore whether such cross-modal transfer would emerge between interaction method and image quality.

Finally, in Chapter 2 we learned that interacting with stereoscopic content by means of rotating the volume, makes the spatial relationships of complex and ambiguous wireframe structures easier to understand. Both object motion and movement parallax increased the accuracy while performing a task, and decreased perceived workload. Combining stereo and motion decreased completion times, without affecting accuracy. Results of Experiments 2 and 3 (Chapter 2) showed that stereo vs. non-stereo manipulations did not affect workload. A reason argued in the discussion of Chapter 2, was that for these two experiments the level of disparity level was either too low or too high. In the current experiment, we therefore employed a disparity level between those used in Experiments 2 and 3

In sum, the current study adds to our understanding of 3D interaction in several ways: (1) it investigates effects of stereoscopic presentation on user experience; (2) it investigates gesture vs. mouse-based interaction in a performance-driven context, assessing effects on workload and allowing us to explore which attributes of user experience impact user preference; (3) additional measures – affect and image quality – provide us with an even broader perspective on user experience of 3D interaction; (4) lastly, the current study employed a disparity level between values in earlier investigation, aiming to optimize the stereo effect of the visualization and via this route further establish the hypothesized reduction of workload.

4.2 Experiment 7: Comparing gesture and mouse-based interaction in a performance-oriented context

4.2.1 Method

Design

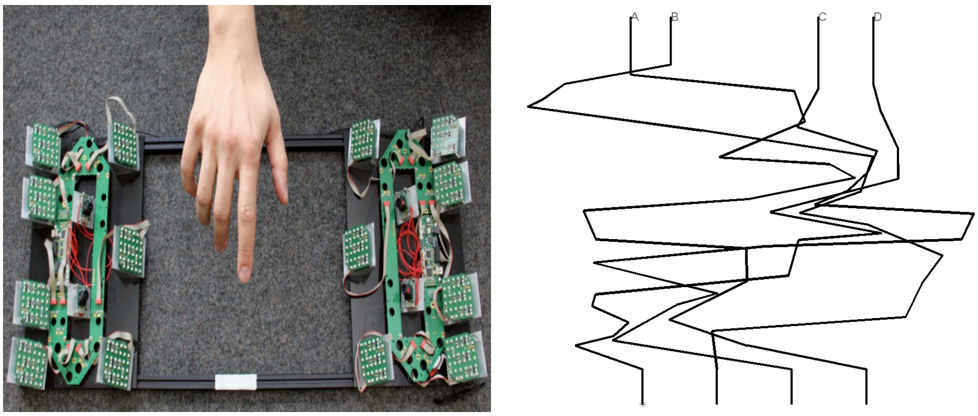
The study followed a 2x2 repeated-measures design, with Visualization method (mono vs. stereo) and Interaction method (mouse vs. gestures) as independent factors. The dependent variables were completion time, accuracy, workload, discomfort, user experience, and image quality. Each condition consisted of 12 unique tasks (6 easy, 6 difficult) randomly assigned to the four conditions.

Participants

Twenty-nine participants, (26 males and 3 females), between 15 and 37 years of age, all with normal or corrected to normal vision, took part. All participants had a stereo acuity better than 40 seconds of arc, tested with the Randot® stereotest. Participants were students from Technical University of Berlin and Fraunhofer HHI.

Stimuli

In the current experiment, users performed a path-tracing task (see Figure 27b) as described in Chapter 2.



(a) (b)

Figure 27. *The dual-view gesture tracker hardware (a), and an impression the task used in the experiment (b)*

In this experiment, the same sets of lines were used as those in Experiment 3. As a reminder, difficulty level was a function of the number of line segments in each stimulus with easy stimuli containing 20 segments and difficult stimuli 24 segments. The 48 tasks were randomly distributed over the four conditions, each containing six easy and six difficult tasks. The maximum disparity used in this experiment was 20 min of arc. The object could be rotated using either the mouse or using gestures.

Setting and Apparatus

The experiment was carried out in the Media-lab at Fraunhofer HHI in Berlin. The stimuli were displayed on a Heinrich Hertz Free2C autostereoscopic 3D Display, as described in the method section of Experiment 2. The gestures were detected using dual-view gesture detection hardware (see Figure 27a). This set-up consists of two cameras and infrared light arrangements capturing the hand(s) from two different viewpoints. The direction of the emitted infrared light is adjustable to provide uniform illumination within the interaction space. The distance between the tracker modules was set to a value of 900 mm. A large field of view was realized by the use of lenses with a focal length of 2.4 mm. The camera base in each tracker module was adjusted to a value of 190 mm. HHI developed software analyzing the two viewpoints, such that occluded parts in one view can be detected in the other view (and vice versa), thus enabling the system to detect all relevant gestures of the hand. Compared to the gesture tracker used in Experiment 3, the new setup increased the tracking accuracy (Hopf, Neumann & Przewozny, 2011), and the gesture used to rotate the volume was modified such that it better corresponds to the gesture found in Experiment 4. Similar as in Experiment 5, participants used two spread fingers, however the position of the hand was not directed towards the floor (i.e., a horizontal alignment) but the user's hand was vertically aligned. Again a movement to the left and right rotated the volume around the y-axis. Both in the mouse and gesture condition the distance from the display was approximately 75 cm. Standard office lighting conditions were used during the experiment.

Measures

The dependent variables used in the current experiment are a combination of the performance-related measures used in Chapter 2 and the user experience measures used in Chapter 3. The performance measures were accuracy, completion times and workload (see for more details Chapter 2, section 2.2.4). User experience was measured using self-reports. Similar to Chapter 3, physical fatigue was measured by assessing postural discomfort adapting the technique used by Corlett and Bishop (1976), asking participants' perceived fatigue using the Borg scale (Borg, 1982). Hedonic and pragmatic qualities were assessed with the Attracdiff questionnaire (Hassenzahl, 2004) as used in Chapter 3. The subscales of the Attracdiff questionnaire were internally consistent with $\alpha = .75$ for pragmatic quality (PQ), $\alpha = .84$ for hedonic quality-stimulation (HQS), $\alpha = .85$ for hedonic quality-identification (HQI) and $\alpha = 60$ for attractiveness (ATT). In addition, we measured users' mood using the positive and negative affect scale (PANAS; Watson & Clark, 1994) consisting of two subscales. The positive affect subscale consists of 10 items (interested, excited, strong, enthusiastic, proud, alert, inspired, determined, attentive, active) and was internally consistent with $\alpha = .84$. The negative affect subscale also consists of 10 items (distressed, upset, guilty, scared, hostile, irritable, ashamed, nervous, jittery, afraid), and was internally consistent with $\alpha = .83$. The response scale ranged from (1) very slightly or not at all to (5) extremely. We further added two items concerning the fluency and naturalness of the interaction, measured on a 9-point scale ranging from (1) not at all to (9) extremely. In addition, we asked users to reflect on the perceived image quality, perceived naturalness, and perceived brightness of the images on a 9-point scale from (1) bad to (9) excellent and perceived depth from none (1) to (9) excellent.

Procedure

Upon arrival at the HHI Media-lab, participants were tested for their stereo acuity using the Randot® stereotest. When participants successfully completed the stereo acuity test, they were seated behind the computer screen and received instructions regarding the experimental procedure and the questionnaires used during the experiment. Participants were instructed to perform the task as rapidly and accurately as possible. Before the start of the experiment, participants practiced the gesture and mouse conditions. The experiment consisted of 4 blocks each with 12 tasks (e.g., 6 easy and 6 difficult), and after each block participants filled in the questionnaires described above. The experiment took approximately 90 minutes and users were compensated with 15 Euros for their participation.

Statistical Analysis

The results were analyzed using a repeated-measures ANOVA analyzing the effects of Visualization method (mono vs. stereo), Interaction method (mouse vs. gestures), and their interaction. For completion times and accuracy the effect of Difficulty (easy vs. difficult) was also analyzed, as well as both two-way and three-way interactions between Difficulty, Visualization method and Interaction method. For accuracy, we first calculated the percentage of correct responses for each of the 8 conditions (i.e., Visualization method, Interaction method, Difficulty). In line with Chapter 2, within each participant we regarded completion times exceeding ± 3 SD as outliers and replaced these values with a completion time corresponding to the mean ± 3 SD (2.2% of the data). In the current experiment, effect sizes were again reported in terms of partial η^2 and Cohen's *d*, as discussed in section 2.5.1

4.2.2 Results

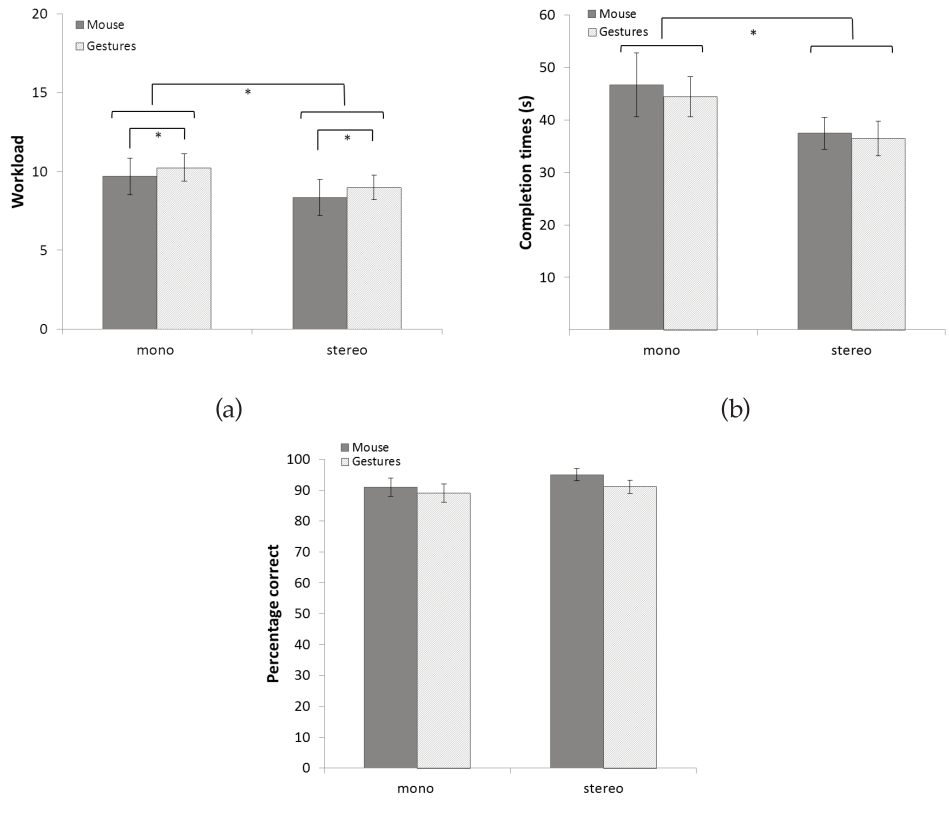
In this section the effects of Interaction method, Visualization method and Difficulty (the latter only for performance measures) will be discussed. First, we will analyze the results for performance indicators - accuracy, completion times and workload - followed by user experience indicators - hedonic and pragmatic quality, positive and negative affect. Finally, we will discuss the results on image quality and user preferences.

Performance measures

Accuracy. The repeated-measures ANOVA with accuracy as dependent variable showed a significant main effect of Difficulty [$F(1,28) = 16.51$; $p < .001$, partial $\eta2 = .37$], indicating higher accuracy for the easy task ($M = .95$, $SE = .01$) compared to the difficult task ($M = .89$, $SE = .02$; $d = .72$). As can be seen in Figure 28 the results did not reveal a main effect of Visualization method [$F(1,28) = 2.79$; $p = .11$, partial $\eta2 = .09$], Interaction method [$F(1,28) = 2.38$; $p = .13$, partial $\eta2 = .08$], or any two- or three -way interactions between Difficulty, Visualization method and Interaction method (all F < 1, *ns*).

Completion times. The repeated-measures ANOVA with completion times as a dependent variable indicated a main effect of Difficulty [$F(1,28) = 47.52$; $p < .001$, partial $\eta2 = .63$], showing longer completion times for the difficult tasks ($M = 49.40$, $SE = 4.62$) than for the easy tasks ($M = 33.16$, $SE = 2.88$; $d = .80$). Figure 28a shows lower completion times when participants performed the tasks in stereo. This was confirmed by the repeated measure ANOVA [$F(1,28) = 6.46$; $p < .05$, partial $\eta2 = .19$]12, indicating that in stereoscopic conditions ($M = 37.0$, $SE = 3.48$) participants were faster (i.e., had lower completion times), as compared to monosopic visualizations ($M = 45.6$, $SE = 4.52$; $d = .40$). Results did not reveal an interaction between Difficulty and Visualization [$F(1,28) = 3.22$; $p = .08$, partial $\eta2 = .10$]. In addition, no significant difference in completion times was found between mouse or gesture-based interaction or any interactions between Interaction method, Visualization method and Difficulty (all F < 1, *ns*).

[12] *Analyzing this with the Log10 transformed completion times showed a stronger effect for Visualization method ($F(1,28) = 13.52$; $p < 0{,}001$; $d = .43$, partial $\eta2 = .33$)*



(a) (b) (c)

Figure 28. *Results of the three performance measures accuracy (a), completion times (b) and workload (c) as a function of Interaction method and Visualization method. The error bars depict 95% confidence intervals.*

Workload. As shown in Figure 28, workload was lower for conditions with stereo and when participants interacted with the mouse. A repeated-measures ANOVA indeed revealed a significant main effect of Visualization method [$F(1,28) = 7.08$; $p < .05$, partial $\eta 2 = .20$] and of Interaction method on workload [$F(1,28) = 4.43$; $p < .05$, partial $\eta 2 = .14$]. Workload was rated lower for stereo ($M = 8.99$, $SE = .50$) than for mono visualizations ($M = 10.25$, $SE = .49$; $d = .47$). In addition, mouse based interaction ($M = 9.02$, $SE = .52$) decreased perceived workload compared to gesture-based interaction ($M = 10.22$, $SE = .52$; $d = .43$). The results did not reveal an interaction between Interaction method and Visualization method ($F < 1$, *ns*).

Chapter 4

User experience

Pragmatic and hedonic quality. The results on pragmatic and hedonic quality are presented in Figure 29. The repeated-measures ANOVA with pragmatic quality as dependent variable showed a significant main effect of Visualization method [$F(1,28) = 22.76$; $p < .001$, partial η2 = .45], revealing a higher score for stereo ($M = 4.74$, $SE = .13$) than for mono visualizations ($M = 4.24$, $SE = .14$; $d = .69$). Interaction method did not reveal a significant main effect between responses for the mouse and gestures (F < 1, *ns*), but the interaction between Visualization method and Interaction method was significant [$F(1,28) = 4.34$; $p < .05$, partial η2 = .13]. As shown in Figure 29, this interaction suggests that the effect of Visualization method on pragmatic quality was largest for mouse-based interaction and less pronounced for gestures. For hedonic quality-stimulation, the results revealed significant main effects of Visualization and Interaction method. The main effect of Visualization method [$F(1,28) = 4.59$; $p < .05$, partial η2 = .12] indicated that stereo ($M = 4.72$, $SE = .15$) rendered slightly higher scores for hedonic quality-stimulation than the mono visualizations ($M = 4.53$, $SE = .16$; $d = .23$).

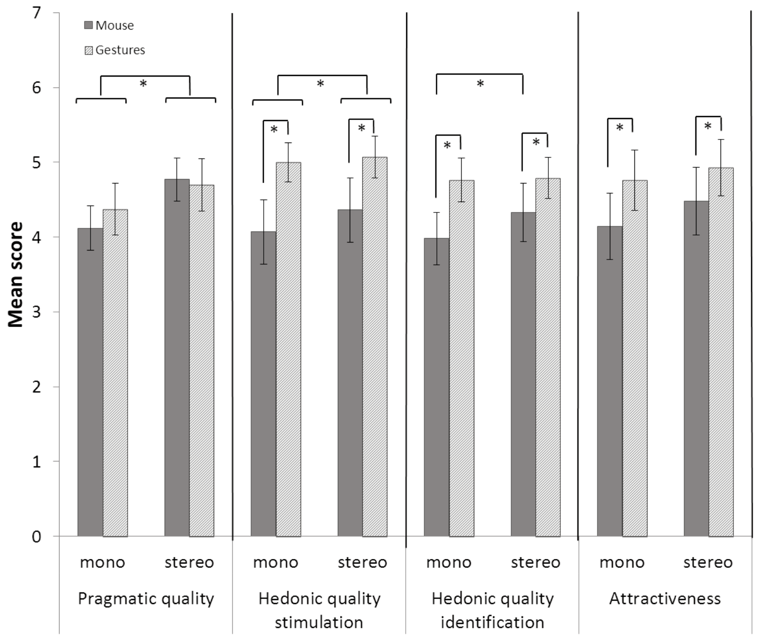
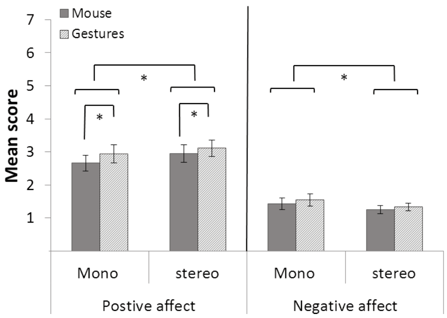


Figure 29. *Mean scores in terms of pragmatic, hedonic quality (stimulation and identification) and attractiveness as a function of Interaction method and Visualization method. The error bars depict 95% confidence intervals.*

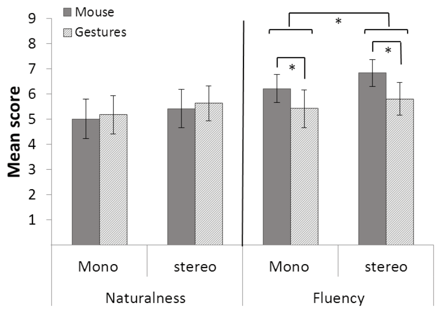
The main effect of Interaction method [$F(1,28) = 20.34$; $p < .001$, partial $\eta 2 = .42$] was larger, showing higher scores in terms of hedonic quality-stimulation for gestures ($M = 5.03$, $SE = .13$) than for the mouse ($M = 4.22$, $SE = .21$; $d = .88$; see Figure 29). The interaction between Visualization method and Interaction method was not significant [$F(1,28) = 1.99$; $p = .17$, partial $\eta 2 = .07$]. For hedonic quality-identification the results showed main effects of Visualization method and Interaction method, as well as an interaction between Interaction method and Visualization method. The main effect of Visualization [$F(1,28) = 6.28$; $p < .05$, partial $\eta 2 = .18$] showed that stereo visualizations ($M = 4.56$, $SE = .15$) result in slightly higher scores than mono visualizations ($M = 4.37$, $SE = .12$; $d = .26$). Again, the main effect of Interaction method [$F(1,28) = 9.80$; $p < .001$, partial $\eta 2 = .26$] was larger, rendering higher scores for gestures ($M = 4.78$, $SE = .14$) than for the mouse ($M = 4.16$, $SE = .18$; $d = .72$). But the interaction between Visualization and Interaction method [$F(1,28) = 5.24$; $p < .05$, partial $\eta 2 = .16$] indicated that the effect of stereo visualizations on hedonic quality-identification was more pronounced for mouse-based interaction than for gesture-based interaction (see Figure 29). In terms of attractiveness the results revealed a main effect of Interaction method and a non-significant trend for Visualization method. The main effect of Interaction method [$F(1,28) = 5.80$; $p < .05$, partial $\eta 2 = .17$] showed that gestures ($M = 4.85$, $SE = .18$) were experienced as more attractive than the mouse ($M = 4.31$, $SE = .21$; $d = .51$). In addition, Visualization method rendered a non-significant trend [$F(1,28) = 3.86$; $p = .06$, partial $\eta 2 = .12$], showing that stereo conditions ($M = 4.71$, $SE = .17$) were estimated slightly more attractive than mono conditions ($M = 4.45$, $SE = .18$; $d = .28$). The interaction between Visualization method and Interaction method on attractiveness was not significant ($F < 1$, *ns*).

Positive and negative affect. In terms of positive affect, the results showed a significant main effect of Visualization method and Interaction method (see Figure 30). The main effect of Visualization method [$F(1,28) = 7.84$; $p < .01$, partial $\eta 2 = .22$] revealed higher levels of positive affect for stereo ($M = 3.03$, $SE = .13$) than for mono visualizations ($M = 2.80$, $SE = .12$; $d = .34$). In addition, the main effect of interaction method [$F(1,28) = 13.39$; $p < .01$, partial $\eta 2 = .33$] showed that gesture-based interaction ($M = 3.02$, $SE = .13$) induced higher levels of positive affect than did mouse-based interaction ($M = 2.80$, $SE = .12$; $d = .33$). The interaction between Visualization and Interaction method on positive affect was not significant [$F(1,28) = 1.11$; $p = .30$, partial $\eta 2 = .04$]. As shown in Figure 30a, stereo visualizations ($M = 1.29$, $SE = .06$) induced significantly lower levels of negative affect than did mono visualizations ($M = 1.49$, $SE = .08$; $d = .53$) with

[F(1,28) = 15.71; p<0.001, partial η2 = .36]13. No main effect of Interaction method [F(1,28) = 2.18; p<0.15, partial η2 = .07] or interaction between Visualization and Interaction method (F < 1, *ns*) was found for negative affect.



(a)



(b)

Figure 30. *Mean scores in terms of positive and negative affect (a) and fluency and naturalness of the interaction (b), as a function of Interaction method and Visualization method. The error bars depict 95% confidence intervals*

Naturalness and fluency of the interaction. With two single self-report items, we asked participants to reflect on the naturalness and fluency of the interaction. For naturalness the results did not reveal a significant main effect of Interaction method (F < 1), or Visualization method [F(1,28) = 3.15; p = .09, partial η2 = .10], nor an interaction effect between these factors (F < 1, *ns*). As shown in Figure 30, the interaction was experienced as more fluent in conditions using stereo visualizations (M = 6.31, SE = .23) than in mono visualizations (M = 5.81, SE = .25; d = .39), with [F(1,28) = 5.08 (p < .05, partial η = .15]. In addition, the main effect of Interaction method [F(1,28) = 5.38; p < .05, partial η2 = .14] showed that gestures (M = 5.60, SE = .33) were evaluated as less fluent than the mouse (M = 6.52, SE = .24; d = .60). The interaction between Visualization and Interaction method was not significant (F < 1, *ns*).

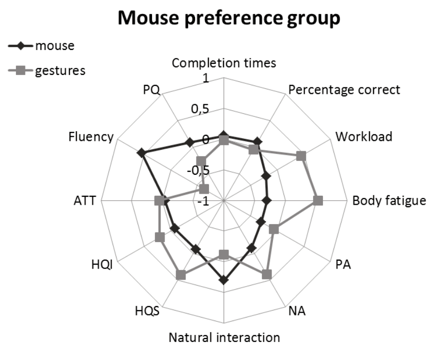
Body fatigue. Gesture-based interaction (M = .80, SE = .15) resulted in higher levels of body fatigue than did mouse-based interaction (M = .44, SE = .08; d = .58), with [F(1,28) = 9.69 p < .001, partial η2 = .26]. No significant main effect of Visualization method [F(1,28) = 2.07; p < .16, partial η = .07] or interaction between Visualization method and

[13] *The data for negative affect were positively skewed (skewness = 1.46). Analyzes of the Log10 transformed NA-scores showed a stronger effect for Visualization method (F(1,28)=16.68; p < .001;partial η2 = .37)*

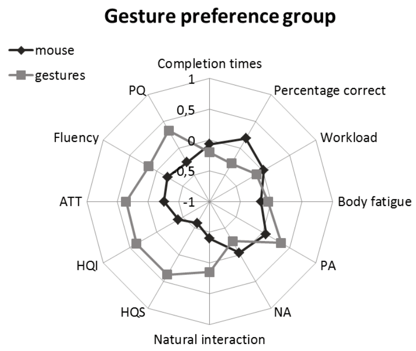
Interaction method (F < 1, *ns*) emerged. For both gesture ($M = 2.4$, $SE = .30$) and mouse ($M = 2.50$, $SE = .26$) weak to moderate levels of fatigue were experienced for the eyes. For mouse-based interaction only very minor complaints were mentioned in the right wrist ($M = .91$, $SE = .21$) and right hand ($M = 1.03$, $SE = .18$). For gestures users mentioned some fatigue in their right hand ($M = 2.07$, $SE = .31$), right wrist ($M = 1.0$, $SE = .24$), right forearm ($M = 1.16$, $SE = .27$), right upper arm ($M = 1.57$, $SE = 1.98$) and right shoulder ($M = 1.97$, $SE = .31$), although these levels represent a low level of fatigue, as body fatigue was measured on a scale ranging from 0 to 10.

Image quality. To understand whether Visualization method and Interaction method influenced users' perception of the images, we asked users to reflect on image quality, brightness, naturalness and depth perceived in the images. Results showed a main effect of Visualization method on naturalness [$F(1,28) = 6.30$; $p < .05$, partial η2 = .18]; images were perceived as more natural when displayed in stereo ($M = 5.67$, $SE = .32$) compared to the mono visualizations ($M = 5.03$, $SE = .32$; $d = .32$). Results did not reveal a main effect of Interaction method or an interaction between Visualization method and Interaction method on naturalness (F < 1, *ns*). Visualization method had a large main effect in terms of perceived depth [F(1,28) = 16.77, $p < .001$, partial η2 = .38], showing higher levels of perceived depth in the stereo condition ($M = 6.40$, $SE = .25$) compared to the mono condition ($M = 4.93$, $SE = .36$; $d = .89$) Again, neither Interaction method nor the interaction between Visualization method and Interaction method was significant (both F < 1, *ns*). In terms of image quality, no main effect of Visualization method [$F(1,28) = 1.34$; $p = .26$, partial $\eta^2 = .05$] or Interaction method, nor an interaction between Interaction method and Visualization emerged (all F < 1, *ns*). The findings for brightness were similar to those reported for image quality, revealing no main effect of Visualization method [$F(1,28) = 3.15$; $p = .09$, partial $\eta^2 = .10$] or Interaction method [$F(1,28) = 3.15$; $p = .09$, partial $\eta^2 = .10$], nor an interaction between Interaction method and Visualization method [$F(1,28) = 3.15$; $p = .09$, partial η2 = .10].

User preference. After users had completed all the tasks, we asked them which visualization method and interaction method they preferred and subsequently explored whether their scores were different depending on their final preference. With regard to the interaction method, 39 percent of the users indicated that they preferred gestures versus. 61 percent of the users who preferred the mouse. With regard to visualization method the preference was more skewed as only 7 percent of the users preferred the 2D visualization versus 93 percent who preferred the 3D visualizations.



(a)



(b)

Figure 31. *Graphical representation of the experience of users who preferred the mouse (a) and gestures (b). This figure shows that participants who preferred the mouse, perceived interacting with the mouse as more fluent and pragmatic than gestures. In addition participants in this group also perceived more workload and body discomfort when interacting with gestures. Incontrast, the group that preferred gesture-based interaction perceived gestures as more hedonic, attractive, fluent and pragmatic. In addition, perceived workload and body discomfort were similar between the mouse and gestures. The values represent z-scores for each measure.*

Since the group that preferred 2D visualizations was too small we only explored the relation between preference and users experience scores for interaction method. In Figure 31, the standardized z-scores for both mouse (black lines) and gesture–based interaction (grey lines) are presented for participants who preferred the mouse (Figure 31a), and participants who preferred gestures (Figure 31b). Overall Figure 31 showed that for both groups, gestures elicited higher scores in terms of hedonic quality, although the difference was larger for the group that actual preferred the gesture-based interaction. As shown in Figure 31a, participants who preferred the mouse experienced the mouse as more fluent and pragmatic than gestures. However, gestures were associated with more discomfort and higher levels of workload and negative affect compared to the mouse. In contrast, the group who preferred gestures showed a different pattern. Figure 31b showed that not only the scores for hedonic qualities were higher for gesture-based interaction, gestures were also experienced as more fluent and pragmatic. In addition, the group preferring gestures did not experienced more workload or discomfort when interacting with gestures. Since for both groups the actual task performance (completion times and percentage correct) were similar, results suggest that the actual preference of users depends more on their subjective experience of the interaction technology, and is therefore a valuable measure to take into account when studying interaction technologies.

4.3 Discussion

Developments in display technology and interaction technology have changed the way we interact with computers in both leisure as well as professional applications. In Chapter 2 we learned that when users were able to interact with the content by means of rotating the image using mouse or head movements, task performance increased and perceived workload decreased compared to static visualizations. Stereoscopic presentation of spatial structures reduced completion times. In addition, in Experiment 1 also workload was slightly reduced for stereo visualization, however this result was not replicated in Experiment 2 and 3, and therefore additional research is needed. Natural and embodied interaction with stereoscopic content requires new and innovative methods of interaction, such that users can easily manipulate the content in three dimensions. In Chapter 3, we learned that embodied interaction (i.e., gestures and Wii) increased user experience in terms of hedonic quality and fun. In Chapter 3, however, the focus was less on the actual performance of a task, but instead on how users experienced the interaction technology. In the current chapter methodologies from Chapters 2 and 3 were combined, while optimizing both the gesture-based interaction (more accurate detection) and the stereoscopic depth (using disparity levels between those values used in Experiments 2 and 3) compared to the studies reported in Chapters 2 and 3. In Experiment 7, we were interested in the added value of stereoscopic displays and gesture-based interaction for task performance, user experience, and image quality. Furthermore, we wanted to gain more insight into which factors influence the actual preference of users for mouse or gesture-based interaction. Therefore, in the current experiment we manipulated visualization method (i.e., stereo vs. mono) and interaction method (i.e., gesture vs. mouse-based), and measured task performance (i.e., completion times, accuracy, and workload) and user experience (i.e., hedonic and pragmatic qualities, positive and negative affect, and naturalness and fluency of the interaction). In addition, we asked participants to reflect on the images in terms of naturalness, image quality, depth and brightness. At the end of the experiment, we asked which interaction technology and visualization method was preferred by the user.

Figure 32 summarizes the results of Experiment 2, 3 and 7, presenting the results of motion either combined with stereo or without stereo. In all three experiments stereo decreased completion times compared to mono visualizations, leaving accuracy unaffected. The effect stereo that had on completion times was consistent across Experiment 2, 3 and 7, showing effect sizes around d = .40. Although participants completed the task faster in all three experiments, users' subjective experience in terms of perceived workload was not consistent across the three experiments. In Experiment 2 and 3 stereo did not reduce perceived workload, whereas it did in Experiment 1 and 7. For both studies, small to medium effect sizes were found (d = .37 Experiment 1 and d = .47 in Experiment 7). Since these four experiments showed mixed results, no firm conclusions can be drawn. However from both Experiment 1 and the combination of Experiments 2, 3 and 7, we suspect that the disparity level is an important factor determining workload reductions. Yet, more research is needed to better qualify these effects by testing a wider variety of tasks combined with different disparity levels. In line with completion times and workload, results in Experiment 7 also showed that pragmatic quality received higher scores for stereo visualizations compared to mono visualizations.

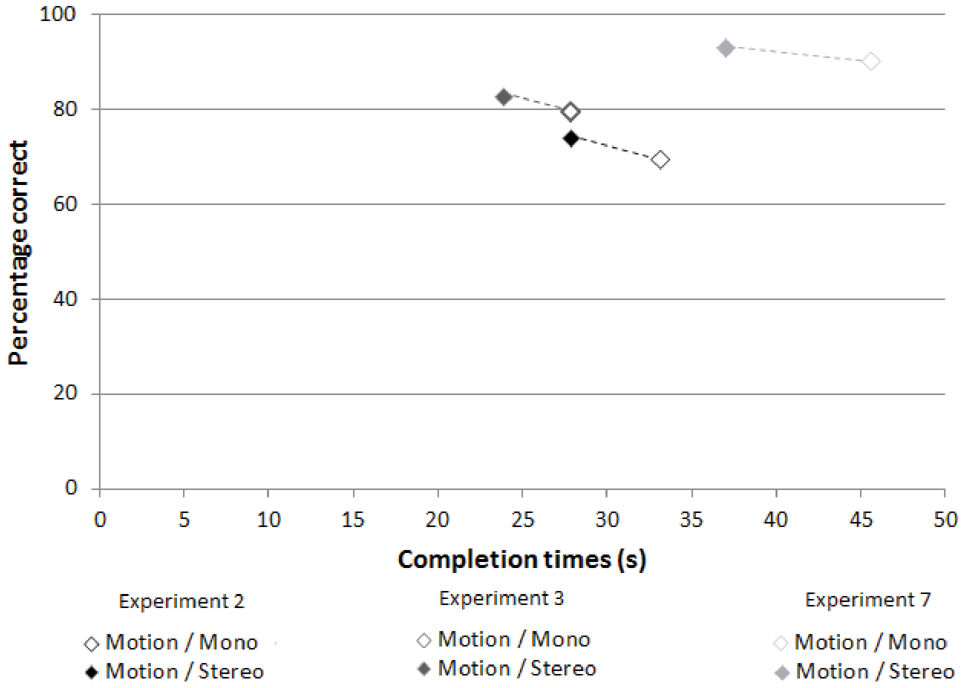


Figure 32. *Results of Experiment 2, 3 and 7, presenting the effects of Visualization method combined with motion on accuracy (y-axis) and completion times (x-axis). The symbols and grayscales in this figure present the three experiments (Experiment 2: black, Experiment 3: grey, Experiment 7: light grey). In addition, the open symbols represent mono visualizations and the filled symbols stereo visualizations. The dotted lines between the various symbols are differences between the mono and stereo visualizations.*

In addition to performance oriented measures, we also measured user experience. Interestingly, stereo visualizations did not only increase task performance but also resulted in higher levels of user experience, independent of the interaction method used. More specifically users reported slightly higher levels of hedonic quality-identification (d = .23) and hedonic quality-stimulation (d = .26), and higher levels of positive (d = .34) and lower levels of negative affect (d = .53). Previous literature suggests that positive feelings of users contribute to task performance (Isen, 2001). In line with this thought, the positive user experience found in the current study may have affected participants' task performance. From results of the current experiment we can however not determine to what extent the experience contributed to the performance of the task. Moreover, as we will discuss in the next paragraph, for gesture-based interaction no such performance benefit was found, although gestures were also perceived more positively than mouse-based interaction.

Both mouse and gesture-based interaction revealed similar results in terms of accuracy, completion times and pragmatic quality. Workload, however, decreased when interacting with the mouse (d = .43). From the items in the workload questionnaire, both physical fatigue and frustration were the most important factors increasing participants' perception of workload while gesturing. This is in line with findings in terms of body discomfort, showing that participants experienced more physical fatigue in gesture-based interaction than in mouse-based interaction (d = .58). Moreover, users perceived gesture-based interaction as less fluent than mouse-based interaction (d = .60). Potentially, this may still be attributed to the current stage of development of this particular tracker system, since the gesture tracker was not yet as accurate as mouse-based interaction, and the tracker sometimes responded to movements that were not intended as interactions. When in the near future detection accuracy and speed are improved, perceived workload might decrease to a level where it becomes comparable or even lower than mouse-based interaction (Goldin-Meadow et al., 2001; Cook et al., 2011). In line with findings reported in Chapter 3, gestures revealed higher scores in terms of hedonic quality-stimulation than the mouse, showing that gestures were experienced as innovative and challenging. Hassenzahl (2004) suggested that these experiences contribute to participants' impressions of the technology by extending their skills and knowledge of how we can interact with computer displays. In contrast to Experiment 5 where we also explored the effect of interaction method on user experience, Experiment 7 focused more on users' tasks performance. However as discussed above, pragmatic quality was similar between mouse and gesture-based interaction. Although the effect size of interaction method on hedonic quality-stimulation was smaller in Experiment

7 (d = .88) compared to Experiment 5 (d = 3.72), gesture interaction had a large effect on hedonic quality stimulation in both experiments. A similar trend was observed in terms of hedonic quality-identification, for which the effect size of interaction method was slightly larger in Experiment 5 (d = .95) compared to Experiment 7 (d = .72). Focusing on users' task performance might nuance their evaluations in terms of hedonic aspects, rendering somewhat more subtle differences than in Experiment 5. Nevertheless, the effect sizes found in both studies revealed medium to large effects. The higher level of hedonic quality-identification suggests that users can also express themselves through the gesture-based interaction, by communicating personal values such as the perceived connection with the content. In addition, users had slightly more positive feelings when using gestures than when interacting with the mouse (d = .33), which is line with James (1884), who argued that motion gives rise to a more visceral experience affecting user's emotions. In the introduction we hypothesized that a positive affect, induced by gesture-based interaction, could increase cognitive processes and creativity (Ashby et al., 1999; Isen, 2001) and therefore increase task performance. However in these earlier studies such effects were mainly apparent for tasks that are important for the user, which is not necessarily true for the task used in Experiment 7.

At the end of the experiment we asked participants which of the interaction methods and visualization methods they preferred. In terms of visualization method, the results showed a clear preference for stereo visualizations (93%). However in terms of Interaction method the preference was more mixed; 39% of the users preferred gesture interaction and 61% preferred the mouse. An exploration of the user experience indicators concerning mouse and gesture-based interaction for each group (i.e., the group of users who preferred the mouse and the group of users who preferred gestures), suggested that participants experienced the technology different. These differences were shown for various subjective indicators, such as hedonic quality and fluency of the interaction, and workload. Participants who preferred the mouse experienced the mouse as more fluent and more pragmatic. In addition these participants experienced more body discomfort and workload when interacting with gestures. On the other hand, participants who preferred gestures experienced gestures as more hedonic, attractive, fluent and pragmatic, and did not experience more discomfort or workload when interacting with gestures compared to the mouse. Interestingly, the objective performance (completion times and percentage correct) revealed similar results for both interaction methods among both groups. This suggests that the final preference is more strongly affected by subjective indicators such as workload, hedonic quality and fluency. This emphasizes the idea that for a full understanding how individuals experience and evaluate interaction technology, a broader perspective of user experience is required. Since not all users preferred gesture-based interaction, results also showed that there are individual differences and, as stated earlier, a one-size-fits all solution probably not exist.

In addition to performance measures and user experience, we also evaluated the perceived quality of the images in terms of naturalness, image quality, brightness and perceived depth. Results showed that image quality was not affected by the visualization method, although perceived naturalness and depth revealed higher scores for stereo visualizations than mono visualizations. This is in line with previous studies evaluating image quality of stereoscopic displays (Lambooij et al., 2010). In terms of interaction method, results did not reveal differences for image quality, naturalness, depth and brightness between mouse and gesture-based interaction. This showed that there was no cross-modal transfer between interaction quality and image quality, as Beerends and De Caluwe (1999) reported between image and sound quality. However, in that particular study they systematically introduced degradations in both audio and video quality, whereas in the current study we only had one manipulation (gestures or mouse). Although our results did not reveal a difference in image quality metrics, this did not necessarily suggest that there should not be a cross-modal transfer between interaction quality and perception of images.

4.3.1 Practical Implications

In line with Chapter 3, hedonic aspects were more favorable for embodied interaction than non-embodied interaction. Results further showed that to have a complete understanding of advantages and disadvantages of the interaction, both task performance and user experience should be taken into account. The two distinct groups (users preferring gestures or the mouse) taught us that for a better understanding of users' preferences, experience should be studied beyond traditional usability measures. The potential theoretical advantages of gesture-based interaction (e.g., embodied interaction, transparency, potentially reducing cognitive load, support learning), combined with enhanced user-experience found in Chapter 3 and the current chapter, showed that gestures can have value for both performance and entertainment settings. However, before gestures are fully accepted as additional interaction method, there are still several challenges concerning gesture-based interaction, including the issue of increased bodily fatigue when using gestures, as well as the issue of how to determine the start and end of a gesture and how to detect a potentially large number of gestures accurately.

In line with Chapter 2, results showed that stereoscopic displays decreased completion times as compared to monoscopic visualizations. Moreover, stereoscopic visualizations were perceived as more pragmatic and participants experienced less workload. In addition to increased task performance, user experience indicators such as hedonic quality and positive affect increased when visualizing the task in stereo. These findings, combined with increased naturalness and viewing experience found in previous literature (Lambooij et al., 2010), illustrate the broad range of applications for which stereoscopic displays can be used. However, we should note that our current findings are based on one particular task that did not have many pictorial depth cues, which might benefit more from stereo than environments or tasks that contain more depth cues. In future research, it would be interesting to study the added value of stereo in more enriched 3D environments for a wider variety of tasks, to better understand the contribution of stereo in terms of task performance, perceived workload and user-experience.

CHAPTER 5

The effects of interaction gain on distance perception

"Perception is not something that happens to us, or in us. It is something we do."
Noë (2004, p. 1)

5.1 Introduction

The above statement by Noë (2004) nicely describes the coupling between action and perception. Traditionally, perception and action were treated as independent processes, however various authors challenged this view (Gibson, 1979; Hommel, Müsseler, Aschersleben & Prinz, 2001 Zwickel & Prinz, 2012). When interacting with objects and environments around us, we are generally not consciously aware of the relation between our own movements in an environment and perceptual impressions of that environment. In our real, physical world, body movements (e.g., walking, grasping), and corresponding perceptual changes of the environment have a stable, invariant relationship. However, in virtual environments, our perceptual experience is not necessarily coupled to our movements in the same way as in daily life. In the previous chapters, we have explored the effects of various interaction methods on user experience, and we have shown that more embodied interaction increased experiences of enjoyment and hedonic quality. However, as already mentioned in Chapter 1, embodied interaction might also have repercussions on our perceptual experience. In the current chapter, we will make a first attempt studying the effect of embodied interaction on the perception of distances in a virtual environment. Before describing our research question in more detail, we will first discuss background literature on perception-action coupling in general.

5.1.1 Perception action coupling

Traditionally, visual perception has been studied in terms of a passive observer, whose brain interprets the light falling on the retina. James Gibson (1979) was one of the first challenging this notion by arguing that in order to perceive the world, one must view it from the perspective of an active observer. Gibson argued that objects in the environment are observed in terms of action possibilities, which he called affordances. Objective characteristics of an object are always present (e.g., the hardness, form, size and heaviness of a rock), however the perception of that rock can be different depending on persons' current state. When running in the woods, a rock can be used to rest on (when the observer is tired), or it might be seen as an obstacle (when the observer wants to continue running). In line with this view, O'Regan and Noë (2001) emphasized the importance of the concept of sensorimotor contingencies, describing the relation between our actions (e.g., head rotation, grasping an object) and corresponding changes in retinal images (e.g., changing perspective, increasing object size). In this view, our perception of objects is not only affected by changes in retinal images caused by movements, but the movements themselves are also part of the perceptual experience. Or as O'Regan and

Noë (2001, p. 1019) put it: “Whereas Gibson stresses the use of sensorimotor invariants as sources of information, we are stressing the idea that sensorimotor invariants are part of what constitute sensations and perceptual content.” Both theories from Gibson (1979) and Noë (2004) are controversial (Clark, 2008), since there are different interpretations to what extent the human body affects perception and cognition, as discussed in Chapter 1 (Gallagher, 2011).

5.1.2 Embodied perception

Nowadays the term embodied perception is frequently used to refer to the role the human body plays during the perception process. Perception is not merely a process of analyzing incoming data, but that it is influenced by behavioral intentions, physical state, and emotions of the perceiver. As a theory, embodied perception stresses the importance of relating perception to the individual’s opportunities, and costs of acting in the environment. This perception-action coupling strongly resonates with the theory of ecological perception by Gibson (1979), discussed in the previous section. Various studies exploring the role of our body in perception have been performed by Proffitt, Witt and colleages (Witt, Linkenauger, Bakdash & Proffitt, 2008; Witt & Proffitt, 2005; Proffitt, 2006; Proffitt, Stefanucci, Banton & Epstein, 2003). In these studies, results have shown that aspects such as participants’ skills and perceived effort can change participants’ perceptions. For instance, participants who were better at softball or golf estimated a ball as larger (Witt, Linkenauger, Bakdash & Proffitt, 2008; Witt & Proffitt, 2005), and participants who were wearing a heavy backpack estimated hills to be steeper, and distances to be further away than participants who did not carry such a heavy load (Proffitt, 2006, Proffitt et al., 2003). However, more research is needed to confirm these findings in different settings/laboratories, since replications in other laboratories have not consistently revealed significant effects of required effort on distance or slope estimation (Durgin, Baird, Greenburg, Russell, Shaughnessy & Waymouth, 2009; Hutchison & Loomis, 2006).

Recently, Zwickel and Prinz (2012) reviewed a large number of studies related to action-perception coupling, describing different theories explaining the coupling between action and perception. One of the theories is based on affordances, as discussed above. Another approach explaining action-perception couplings is the Theory of Event Coding (TEC) (Hommel et al., 2001), assuming that both action and perception are coded within the same processing stages. An alternative view is based on attention, assuming that the planning of an action changes the attention of a person and therefore influences perception (Schneider & Deubel, 2002). As discussed by Zwickel and Prinz (2012), there are also theories that assume that goals only play a minor role and that perception is mainly motor based. These different views on action and perception illustrate that there is still no general accepted view on how findings on action-perception couplings can be explained. Studies reviewed by Zwickel and Prinz (2012) showed that there is a large body of evidence that action and perception interfere upon each other. Interestingly, some studies showed that action enhances perception (i.e., an assimilation effect) while other studies showed that action attenuates perception (i.e., a contrast effect). An example of a contrast effect is shown by Hamilton, Wolpert and Frith (2004), where participants judged the weight lifted by actors as heavier when lifting a light weight themselves at the same time. On the other hand, Wohlschläger (2000) showed evidence for an assimilation effect, where the perceived direction of rotation of ambiguous dots was influenced by the turning direction of a knob held by the participants. Zwickel and Prinz (2012) concluded that action either attenuates or enhances perception depending on various factors, such as: perceptual ambiguity of the stimuli; if action-perception is functional related or unrelated; if action and perception share overlapping features; and if stimuli are presented concurrently (for an elaborate discussion see Zwickel and Prinz (2012)). A recent study by Zwickel, Grosjean and Prinz (2010), studied whether proprioceptive information, or the planning of a movement, explains the action-perception coupling. In this experiment, participants had to detect the deviation of a vertical moving point, while simultaneously moving their hands to the left or the right. To isolate proprioceptive information, the hands were transported by a motor. In these conditions, no effect of action on perception was observed. On the other hand, using a fixed pen (i.e., participants were able to plan a movement, but could not move their hand), resulted in an assimilation effect, showing a faster detection of stimulus motion in the direction of the intended hand movement. This result suggests that the intention of a movement plays a more important role than pure proprioceptive information. These studies clearly show the complexity of the interference process between action and perception, and the role our body play in these processes.

Another relevant line of research is related to tool use and distance perception. In Chapter 1, we already mentioned the flexibility of our brain to incorporate tools in our body schema (Clark, 2003; Haans & IJsselsteijn, 2012). Berti and Frassinetti (2000) provided indications that using a tool can extend persons' peripersonal (within arm reach) space to that of extrapersonal (beyond arm reach) space. In this experiment, they studied a patient with damage to the right hemisphere, having a left-sided neglect in the near space but not in the far space. A patient with a neglect ignores stimuli from either the left or the right side, and consequently, when asked to divide a line in half (i.e., a so-called line bisection task), the estimation of the midpoint of the line is either shifted to the right or the left. In the case of Berti and Frassinetti's (2000) patient, the patient's neglect would extend to the left side, leading to a rightward displacement error in the line bisection task. When the patient used a light pen, the neglect appeared in the near field but not in the far field. However, when the patient used a stick, the neglect appeared in both the near and far field conditions, showing that the near field was extended towards the far field. Recent studies also showed that when participants were holding a tool, objects appeared closer than when they were not (Witt, Proffitt & Epstein, 2005; Osiurak, Morgado & Palluel-Germain, 2012). However, this effect appeared only when they intended to use the tool (Witt et al. 2005) and/or when the stick had a sufficient length (Osiurak et al., 2012).

This discussion showed that our perception of the environment can be influenced by factors such as physical state, perceived skills and tool use. In the current chapter, we will apply the concept of embodied perception using virtual environments, focusing on the relation between our physical movements and perceptual changes of the environment. We will ask users to estimate the distance between two objects while interacting with these objects in the environment. Estimating distances is a basic activity used for many daily tasks such as reaching out to objects in front of us, or interpreting the size of a room. However, research has shown that we are not always accurate at estimating distances, and that, in particular, distances in virtual environments are frequently significantly underestimated.

5.1.3 Distance perception

Various depth cues (pictorial, motion, and binocular depth cues) help us when estimating the distance between ourselves and objects (egocentric distance estimates), or between two objects (exocentric distance estimates). In Chapter 2, we learned that these depth cues enhance tasks such as aligning objects or identifying spatial relationships. Also for estimating distances, these depth cues are important. Various authors have studied the relative contribution of different depth cues on the perception of distance (e.g., see Sedgwick, 1986; Cutting & Vishton, 1995; Hershenson, 1999 for overviews of this work). Such studies have employed a range of methods to assess distance estimates, including verbal reports, blind walking, and even throwing balls. Notably, these different assessment strategies may result in different findings. For example, when participants were asked to walk blindfolded to a point in the environment, they were more accurate than when they were asked to verbally estimate that distance (Cutting & Vishton, 1995). A more elaborate discussion of these methods can be found in Loomis and Philbeck (2008).

In real life, egocentric distances are slightly underestimated, whereas for exocentric distances, estimates are more accurate (Cutting & Vishton, 1995). Interestingly, in virtual environments results seems to be exaggerated, showing underestimations up to 50 percent for egocentric distance estimates (for both verbal and blind walking estimation; Witmer & Kline, 1998; Loomis & Knapp 2003; Thompson, Willemsen, Gooch, Creem-Regehr, Loomis & Beall, 2004), and overestimations (Wartenberg & Wiborg, 2003; Waller, 1999), or near veridical estimates (Richardson & Waller, 2007) for exocentric distances. The fact that egocentric distances are underestimated has challenged researchers to explore factors that could be contributing to such an estimation error, including the limited field of view when wearing HMDs, inaccurate stereo visualizations, limited cue availability, limited resolution and quality of the images, errors in accommodation, and weight of the helmet. However, none of these factors fully explained the underestimation of distances in VR (Waller, 1999; Creem-Regehr, Willemsen, Gooch & Thompson, 2003; Thompson et al. 2004; Willemsen, Gooch, Thompson & Creem-Rehehr, 2008; Willemsen, Colton, Creem-Regehr & Thompson, 2009). Providing participants with feedback on how accurately they estimated distances (Waller, 1999; Richardson & Waller, 2007), or showing them their virtual self-representation or avatar (Mohler, Creem-Regehr, Thompson & Bülthoff, 2010), increased the accuracy of distance estimates. In addition, Richardson and Waller (2007) showed that distance estimates in virtual environments became more accurate when participants were allowed to interact, by means of walking through the environment, prior to their distance estimation. Both studies suggest that the human body plays an important role when estimating distances in the (virtual) environment

5.1.4 Gain

Virtual environments have the ability to simulate real world settings or to present an imaginary world. Importantly, the laws of physics and the invariant action-perception relations discussed earlier do not necessarily hold for virtual environments. For example, in VR environments, a participant can walk through solid objects, change laws of gravity, have an entirely re-arranged virtual body. One of the most basic factors related to how we interact with the environment is the gain of the interaction. The gain, a ratio of output to input, describes the relation between our movements and our perception of those movements. Whereas in daily life the gain always has a ratio of one, in virtual environments this can be any number. When implementing a gain that is larger than one, movements in the virtual environment are larger than the movement of our physical hands. In fact, previous studies have shown that in many applications a gain larger than one is used, since this increases pointing efficiency, and decreases body discomfort (Johnsgard, 1994; Casiez, Vogel, Balakrishnan & Cockburn, 2008), although it has a lower fidelity, since in daily life the gain is always one. Various sources of information can be used to sense the movements of our limbs. First, proprioceptive senses allow us to internally determine our movements (distance and speed) and the position of our hands and arms in the environment (for a more detailed discussion see Proske, 2006). Another mechanism is efferent copy, where a copy of our outgoing motor command is sent to the brain and used to predict visual changes caused by our movements (Miall & Wolpert, 1996). In daily life we use both sources of information unconsciously, and we have learned to trust this information for many daily activities concerning both action and perception.

5.1.5 Rationale for the studies

In the introduction, we discussed the role of the human body in perceiving the environment. In daily life, when we interact with objects around us, the relation between our own movements and our perception of those movements (i.e., gain) is always one. However, in computer-mediated interaction this gain is flexible, and may differ between applications. This allows us to investigate the role of embodied cues in computer-mediated interaction, since we can vary the physical movements without changing the visual displacements. In the remainder of the current chapter, we will investigate the role of embodiment in perception by exploring whether the size of our hand movements influences our perception of distances. We hypothesize that in virtual space, a mismatch between our physical movement and the projection of that movement, may affect our perception of 2D en 3D space. Zwickel and Prinz (2012) argued that assimilation effects occur for functionally related tasks and ambiguous stimuli. Since in both experiments, stimuli were presented somewhat ambiguous (e.g., only one block presented each time), and the tasks were functionally related (traveling and estimating distances), we expect that larger gains – and consequently smaller body movements – will result in lower estimates of the same distance. We will study this in a desktop environment (Experiment 8), and in more immersive 3D environments using a head mounted display (Experiment 9). Previous literature on distance estimations in virtual environments (either on display screens or via an HMD) has shown that performance was not as accurate as estimations in natural environments. However, accuracy increased when participants were allowed to interact with the environment (Richardson & Waller 2007). Whether gain played a role in this phenomenon has not been investigated. However, gain is a fundamental transformation of the action-perception relation, and therefore an interesting factor for studying the differences in real world and virtual world distance estimations found in previous literature. For this reason, we will study whether gain manipulations influences the accuracy of distance estimations.

In Experiment 8, participants interacted with a LCD display using a mouse. We manipulated gain level by introducing two gain levels (1 and 5), where a gain of 1 denotes a direct one-to-one mapping of controller movements and cursor movement, and a gain of 5 a mapping where the cursor movement is 5 times larger than the controller movement. Participants estimated distances between two rectangles along the horizontal and vertical axes. From a pilot study, we learned that when presenting the two rectangles simultaneously, participants estimated the distances first, before actually moving between the two rectangles. Therefore, we changed the task preventing participants from estimating the distance before actually interacting with the content. Gain was manipulated within participants, however half of the participants started the first block with a gain of 1, and the other half with a gain of 5. This allowed us to test whether gain affected participants' distance estimates within a session, or whether the gain participants started with, was used as a reference frame within a session.

In Experiment 9, the aim was to extend findings of Experiment 8 using a 3D environment wearing a Head Mounted Display (HMD). Again, we investigated whether the size of users' hand movements changed the perceived distance between two objects. In addition, we explored which of the two gain levels (i.e., high vs. low) resulted in the most accurate estimations. Third, we investigated whether the use of an interaction device impacted these effects, by comparing conditions in which participants were holding a trackable tool to conditions in which they were wearing a trackable glove. Our hypothesis was that participants would rely more on hand and arm movements using more direct interaction (like a glove), than during less direct styles of interaction (i.e., a tool). In addition, previous studies showed that a stick extended participants' peripersonal space, and decreased participants' distance estimates. Fourth, we wanted to explore whether reaching distance moderates participants' distance estimation. For this reason, we introduced a range of reaching distances for horizontal distance judgments. As dependent variables, we employed verbal distance estimates, and asked participants to reflect on hedonic and pragmatic quality, embodiment, and body fatigue. These insights could inform the design and implementation of interaction technologies in VR.

5.2 Experiment 8: effects of mouse gain on distance estimation along X and Y axes

In the first experiment of this chapter, we manipulated mouse gain to vary the size of the hand movements required to select a distant object. We were interested if the distance travelled with the hand influences the perception of – visually equal – distances between two objects presented on the screen.

5.2.1 Method

Participants

Forty-six participants (30 males and 16 females) between 19 and 27 years old, all with normal or corrected to normal vision took part in this study. Participants were recruited from a database containing both students as well as individuals unrelated to the university.

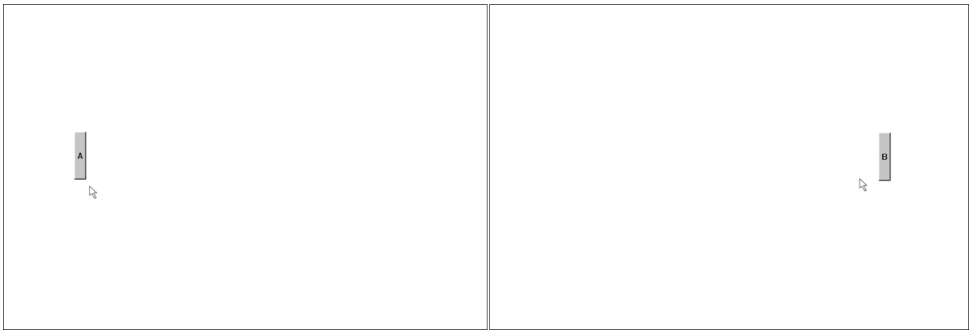
Design

The study followed a 2x2x10 repeated-measures design, with Gain (1 vs. 5)[14], Direction (horizontal (x-axis) vs. vertical (y-axis)) and Distance (10 distances) as within-subjects factors. Due to the wide screen display set-up, the distances chosen for estimates on the horizontal axis (i.e., '2', '4.5', '7', '9.4', '11.9', '14.4', '16.9', '19.3', '21.8', and '24.3' cm) were slightly different from those selected for the vertical axis (i.e., '2', '4.5', '5.7', '7', '8.2', '9.5', '10.7', '11.9', '15.5', and '16.9' cm). In each condition participants made three distance estimates, resulting in a total of 120 distance estimates. The dependent variable was accuracy of the estimates, calculated as the percentage under/over estimation.

Setting and Apparatus

The experiment was carried out in the 3D/e lab of the TU/e, where we created two identical set-ups. The task was presented on a Dell S2309W 23 inch wide screen monitor with a resolution of 1920x1080 pixels. Participants interacted with the computer using a mouse. The gain was manipulated by running an AutoHotKey script within Authorware.

[14] *After the experiment, we noticed that windows has an option 'enhance pointer precision' that internally adapt the gain based on the speed of the physical mouse movement. A faster movement of the mouse will decrease the gain, whereas a slower movement of the mouse will increase the gain. Therefore the gain-settings originally chosen in this experiment are different than the 1 and 5 originally planned. Results revealed that the speed of the physical mouse was faster in the gain of 1 condition and slower in the gain of 5 condition. This means that the gain experienced by the users will somewhat smaller than 5 in the conditions with a gain of 5, and somewhat larger than 1 in the conditions with a gain of 1.*



(a) (b)

Figure 33. *Two screenshots of the task. When participants started the task, only one rectangle was visible (a). When participants clicked on A, and moved 1.33 cm to the right or above, the second rectangle became visible and the first disappeared (b).*

Stimuli and task

In the current experiment, participants estimated distances between two rectangles displayed on a computer screen separated horizontally or vertically (see Figure 33). The rectangles were .50 cm wide and 2.1 cm high, and positioned along the horizontal or vertical axis. Ten different distances were used for both the x and y-axes. Each first rectangle appeared at a different position on the vertical and horizontal axes to avoid that participants' recognize previously estimated distances. The task used in this experiment was based on a pilot study. In this pilot study, the task was to estimate the distance between two rectangles which were both shown on the screen, using two gain levels. Half of the participants interacted with the task by clicking on the two squares, and the other half dragged one square to the other. Results showed no difference between the dragging and clicking, nor any difference between the two gain levels. At the end of this pilot study, we asked participants what their strategy was while estimating distances. Participants mentioned that they first estimated the distances between the two squares, before actually moving the mouse from the first to the second square. Based on this finding we changed the task in a simple yet crucial way, which prevented participants from performing distance estimates before hand movements were performed. As shown in Figure 33, participants only saw one rectangle at the time. Once participants clicked on the first rectangle and moved their mouse 1.33 cm to the right (during horizontal distance estimation), or upwards (during vertical distances estimation), the first rectangle disappeared and a second rectangle appeared on the screen. After participants had clicked on the second rectangle, they entered the perceived distance using a keyboard. The experiment was programmed in Authorware.

Measures

The dependent variable in the experiment was the distance estimate, computed as a percentage according to the function proposed by (Waller, 1999)

$$\%\ estimation = \frac{Estimated\ Distance}{Actual\ Distance} \times 100 \quad \textit{(100\% = veridical)}$$

Procedure

Upon arrival at the 3D/e lab, participants were seated behind the computer and received instructions regarding the experimental procedure. Before the start of the experiment, participants practiced three distance estimates in the condition they started with. The experiment consisted of two blocks (i.e., horizontal and vertical distance estimations) each consisting of 60 tasks (30 with a low gain and 30 with a high gain). All participants started with estimations in the horizontal direction, and estimated the distances for both low and high mouse gain. After this block, participants started with distance estimations in the vertical direction. Half of the participants started each block with a gain of 1, and the other half with a gain of 5. The experiment took approximately 20 minutes and participants were compensated with 5 euro for their participation.

Statistical Analysis

Distance estimates were checked for typing errors (i.e., values above 100 and values with only a 0 (3 out of 3120 estimates were removed)). In addition, within each distance we regarded distance estimates exceeding ± 3 SD as outliers and replaced these values with a distance estimate corresponding with M ± 3 SD (.5% of the data). The distance judgments for each trial were analyzed using a repeated-measures ANOVA with gain and distance as within-subjects factor. Horizontal and vertical distance estimates were analyzed separately, since the distances were slightly different between the two directions. For Distance, and the interaction between Distance, Gain and Direction the test of Sphericity was violated, and therefore we reported these results with the Greenhouse-Geisser correction. In the current experiment, effect sizes are again reported in terms of partial η2 and Cohen's d, as discussed in paragraph 2.5.1. However, for between subjects comparisons reported in this chapter we will apply Cohen's ds, to compare two groups of independent observations. Cohen's d_s is calculated by: $d_s = \frac{M1-M2}{SD_{pooled}}$ where $SD_{pooled} = \sqrt{\frac{(n_1-1)\cdot SD_{1\,1}^2+(n_2-1)\cdot SD_1^2}{n_1-n_2-2}}$ (Cohen, 1988). For within-subject comparisons, the formula described in paragraph 2.5.1 is used.

Chapter 5

5.2.2 Results

As shown in Figure 34 the results showed that participants' estimates were, on average, higher than 100%, revealing an overestimation in their distance estimates. A one-sample t-test confirmed this for both the horizontal (M = 116.40, SE = 1.10; t(919) = 14.88, p < .001; d = .49) and vertical axis (M = 134.0, SE = 1.30; t(919) = 26.78, p < .001; d = .88). A repeated-measures ANOVA with percent estimation as dependent variable showed a significant main effect of Distance, for both horizontal [F(1.9, 83.4) = 24.6; p< .001, partial η2 = .35] and vertical distance estimates [F(2.4, 109) = 25.20; p < .001, partial η2 = .36]. Post hoc tests revealed that shorter distances were overestimated more than longer distances. Results did not reveal a significant main effect of Gain in the horizontal [F(1, 45) = 2.53; p = .12, d =.11 partial η2 =.05] and vertical direction (F<1, ns). In addition, results did not show an interaction between gain and distance for both horizontal [F(5, 226) = 1.25; p = .28, partial η2 = .03] and vertical estimates [F(5, 243) = 1.46; p = .20, partial η2 = .03]. For each direction (horizontal and vertical), half of the participants started with a gain of 1, and the other half with a gain of 5. After completing the 30 distance estimations with a particular gain level (session 1), the same tasks were performed with the other gain manipulation (session 2).

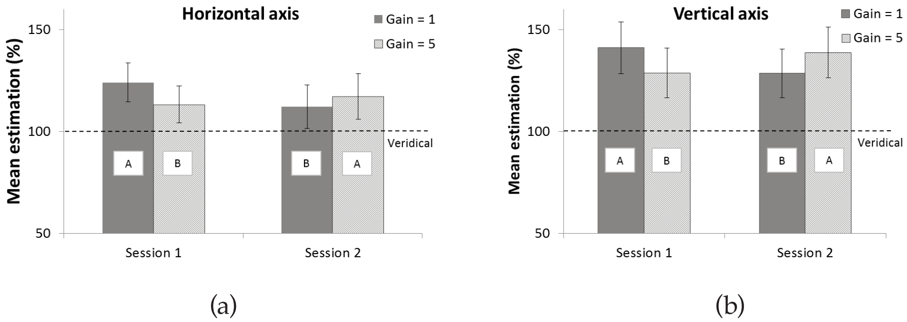


Figure 34. *The effect of gain on horizontal (a) and vertical (b) distance estimates. Session 1 is the first distance estimation at the start of both horizontal and vertical distance estimations. In the second session, the group that started with a gain of 1 (indicated with the letter 'A') now interacted with a gain of 5 and the group that started with a gain of 5 (indicated with the letter 'B') changed to a gain of 1. The scores are presented in percentage under/overestimations; scores over 100 percent represent overestimations and scores under 100 underestimations. The error bars depict 95% confidence intervals. This figure shows an overall overestimation of both horizontal and vertical distance estimations. In addition, in session 1 a non-significant trend was observed towards larger distance estimations when interacting with a gain of 1 compared to a gain of 5. However results in session 2 suggest that participants calibrate their estimates in the first session, and used this during session 2, regardless of the gain level participants interacted with.*

Figure 34 presents the results for horizontal and vertical distance estimates, using separated plots for session 1 (gain level participants started with) and session 2 (gain level participants used second). In line with informal comments of participants, results in Figure 34 suggest that for each direction, the distances estimated in the first session were used to calibrate their estimates throughout the experiment. In other words, results of session 2 seem to be biased by the gain settings participants were confronted with during the first session. Therefore, we separately analyzed the data of the first session using a repeated-measures ANOVA with distance estimation as dependent variable, Distance as within-subjects variable and Gain as between-subjects variable. Both horizontal and vertical distance estimates were analyzed separately. In line with the previous analyses, smaller distances were overestimated more than larger distances for both horizontal [$F(1.9, 84.8) = 18.90$; $p < .001$, partial $\eta2 = 0.30$] and vertical distance estimations [$F(3.4, 147.9) = 20.10$; $p < .001$, partial $\eta2 = .31$]. Although Figure 34 suggests that distance estimates were larger using a gain of 1 than when using a gain of 5, this result was not significant for estimations along the horizontal [$F(1, 44) = 2.60$; $p = .11$, partial $\eta2 = .06$] or vertical axis [$F(1, 44) = 1.91$; $p = .17$, partial $\eta2 = .04$]. Cohen's d effect size nevertheless showed a moderate effect for estimations along the horizontal axis ($d_s = .49$) and small to moderate effect for the vertical axis ($d_s = .42$). No interaction was found between Gain and Distance ($F < 1.1$, ns).

5.2.3 Summary of results

The aim of the current experiment was to investigate whether the size of our hand movements influences the perceived distance between objects. Our hypothesis was that when participants' hand movements were larger (i.e., using a gain of 1), the perceived distance between the two objects would be perceived as larger. Although Figure 34 suggested a trend, in the expected direction, results did not show significant differences between the participants using a higher gain level (i.e., smaller hand movements) vs. a lower gain level (i.e., larger hand movements). While no significant differences emerged, effect sizes for estimations on the horizontal direction ($d_s = .49$) and vertical direction ($d_s = .42$) were small to medium. However, this effect only emerged when comparing the estimates in the first session of both horizontal and vertical direction. This suggests that the gain level participants started with was used to calibrate their estimates used throughout the session, even when the gain was altered during the experiment. Results further showed an overall overestimation of distances presented on the display (effect sizes of $d = .49$ for horizontal judgements and $d = .88$ for vertical judgements). This is in line with previous studies performed by Roscoe (1984) and Waller (1999). Extending these findings beyond display effects, Künnapas (1955) showed that when a frame was presented around a line, estimates of the size of the line were larger when this frame was larger. This might also play a role in judgments on computer displays, since displays also have a frame around the screen.

Various factors can explain why differences between larger hand movements on participants distance estimates were not statistically significant. First of all, the total number of participants (n = 46) may have been too low for a between-group comparison. A power analysis showed that with this number of participants only large effect sizes can be detected (with a power of .80 and α = .05). In addition, as discussed in the method section, Windows internally adjusted the gain settings depending on the speed of the mouse. Therefore, the gain settings experienced by the users were actually lower than five for a gain of 5, and higher than one for a gain of 1, rendering a more subtle manipulation than intended. In Experiment 9, we aim to replicate this experiment in a more immersive 3D environment and with interaction methods that are more embodied. In addition, the new setup allowed us to also include the z-axis in our investigation. Compared to Experiment 8, we will use a larger number of participants, and the gain level will be controlled to ensure gain levels of 1 and 5. Furthermore, in Experiment 9 we will manipulate the gain between instead of within participants, since results in the current experiment suggested that the gain level participants started with is used as reference for their subsequent estimations. An additional advantage is that participants are blind to our experimental manipulation. Using a 3D environment will also increase the practical relevance, since an increasing number of applications will use 3D environments for training and business purposes.

5.3 Experiment 9: Distance perception in 3D space

A recent article in the De Volkskrant discussed the future role of 3D environments for visualizing the interior of a house to new potential buyers (Ammelrooy, 2012). For potential buyers an accurate and realistic perception of the size of the rooms is important. Previous studies showed that participants typically underestimate distances in virtual environments (Witmer & Kline 1998; Loomis & Knapp 2003; Thompson et al., 2004). On the other hand, when participants are able to walk in the virtual environment, estimates became more accurate (Richardson & Waller, 2007). In Experiment 8, results showed a trend that the gain that is introduced during such interactions also impacts the perception of distance, and thus potentially biases the experience of the environment. In the current study we aimed to replicate and extend findings of Experiment 8, this time using a 3D virtual environment. In contrast to Experiment 8, participants not only estimated exocentric distances, but also estimated egocentric distances, and sizes of the objects in the environment. This allows us to compare findings from the current experiment with findings from Experiment 8 (exocentric distance estimation), and previous literature on distance estimation in VR (egocentric distance estimation).

As we discussed earlier, users may interact with virtual content using a tool (i.e., a device that is tracked by the apparatus), or without such a device (when the apparatus is able to track gestures without the need for holding a device). In the current study, we are also interested in learning whether holding a tool, would render results different from deviceless interaction. Interacting with a tool might feel less direct, perhaps making one depend less on proprioceptive cues, than when interacting directly with one's hands. Moreover, previous studies showed that holding and using a stick, changed people's perception, making objects appear closer with, than without using a tool (Witt et al., 2005; Osiurak et al., 2012). Nevertheless, the fact that participants have to reach less when holding a stick, might explain why distances were perceived as closer than when pointing to an object without a stick. Therefore, in addition to the above manipulations (i.e., gain and interaction method), we also investigated the effects of reaching distance on persons' distance estimations.

5.3.1 Method

Participants

Seventy-seven participants (55 males and 22 females), between 18 and 34 years of age, all with normal or corrected-to-normal vision, took part. Participants were recruited from a database containing both students as well as individuals unrelated to the TU/e. All participants had a stereo acuity better than 40 second of arc, tested with the Randot® stereotest.

Design

In the current experiment, participants estimated distances between two objects (exocentric), between themself and an object (egocentric), and judged the sizes of objects in a virtual environment. The effects of gain and interaction method on these estimates were investigated using a repeated-measures design with two gain levels (1 vs. 5) and two interaction methods (glove vs. device) manipulated between participants and distances manipulated within participants.

For the exocentric distance estimates, 10 distances (5, 10, 15, 20, 25, 30, 35, 40, 45, and 50 cm) were provided along two directions (z-axis vs. x-axis). Participants were equally distributed over the four conditions (i.e., glove-low gain, glove-high gain, device-low gain and device-high gain). The exocentric distance estimation task consisted of two blocks, i.e. estimation along the x-axis and along the z-axis). Within each block participants performed 30 distance estimations (i.e., each distance was estimated three times), resulting in a total of 60 distance estimates, The different distances were randomly shown to the participant. In addition, for distance estimations on the x-axis, three distinct positions from the participants (along the z-axis) were used (50, 65, and 80 cm) to investigate the effect of reaching on perceived distances.

For egocentric distance estimates, 5 distances along the z-axis (40, 50, 60, 70, and 80 cm) were provided. This egocentric estimation task consisted of one block in which participants estimated the five distances twice[15], resulting in a total of 10 egocentric distance judgments. After participants completed these ten estimates, we asked them to estimate the size of the red ball, the Rubik's cube and the length and width of the table.

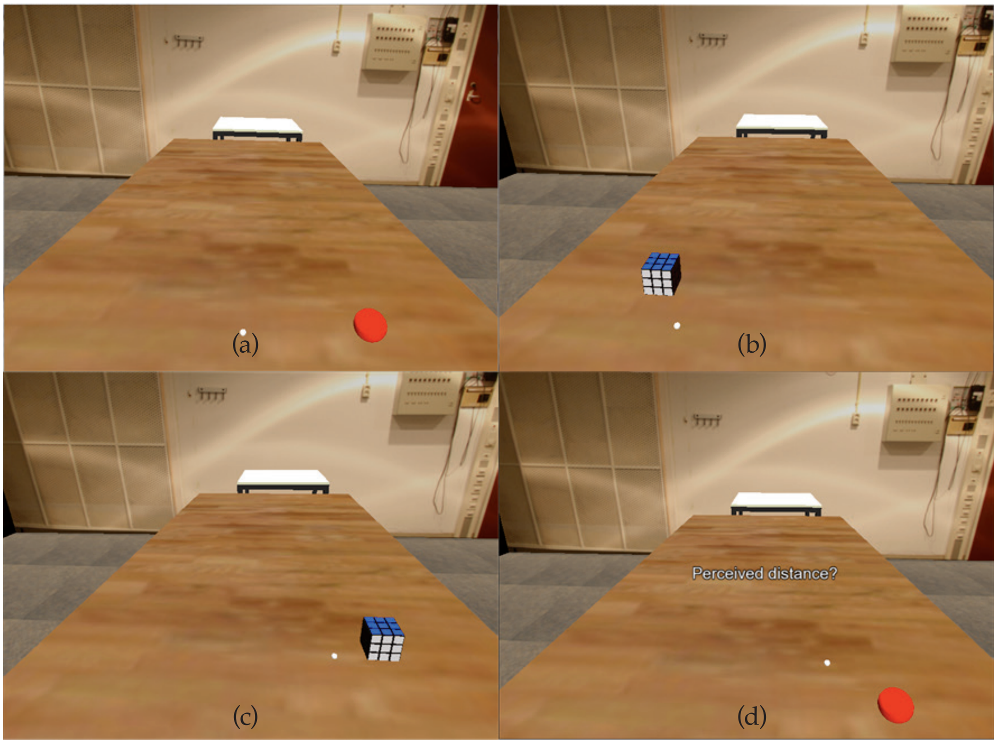


Figure 35. *Screenshots of the experimental task during exocentric distance estimation along the x-axis. (a: upper left panel) Before each trial participants were asked to move the white pointer into the red ball. (b: upper right panel) When participants moved the pointer into the red ball a Rubik's cube appeared on the table. (c: lower left panel) When participants selected the cube it disappeared, and a second cube appeared on the table when participants moved the pointer cm to the right. (d: lower right panel) Participants continued moving to that cube and after selecting this cube they were asked to estimate the distance perceived between the two cubes*

[15] *Note that the first 8 participants estimated the distances only once.*

Stimuli and task

In the current experiment, participants were placed in a fully immersive virtual environment, representing the lab environment. The environment consisted of an exact copy of the Lab with a table in the center of the room (see Figure 35). The table had a size of 90 x 199 cm. On the table we placed a Rubik's cube (5.7 x 5.7 x 5.7 cm) used for the distance estimation tasks. Participant could interact with the environment by steering a white ball (with a diameter of .50 cm) used as pointer. This pointer followed the movements of the hand or the device, depending on the experimental condition (see Figure 36). Participants that used a glove could make a selection by tapping the thumb and pointing finger, whereas participants that used a device (see setting and apparatus paragraph for more detail) pressed a button with their thumb.

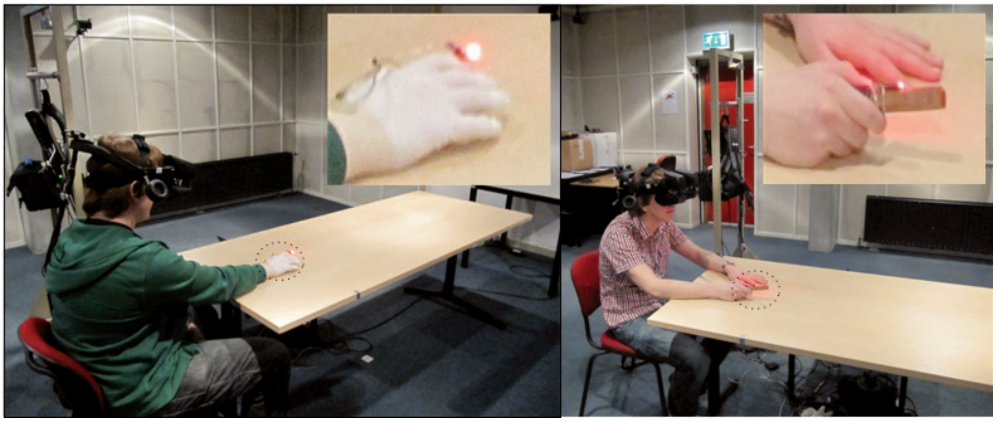
The task for judging exocentric distances is presented in Figure 35. This task consisted of several trials during which participants estimated the distance between two cubes. Before each trial participants moved the pointer into a red ball (diameter = 2.5 cm), which is used as starting point before interacting with the cubes (see Figure 35a). The red ball was located on the table, 40 cm in front of the participants, and 20 cm right from the center of the table. When the pointer was moved into the red ball, a Rubik's cube appeared on the table (see Figure 35b). Participants moved the pointer into the cube and were alerted with a short 'beep', and the color of the pointer changed from white to green. Participants were asked to remember the location of the first cube, and select this cube by pressing a button (for the tool-based interaction), or tapping the thumb and pointing finger (for the glove-based interaction), to start the estimation task. After they selected the cube and moved 3 cm to the right (for horizontal estimates), or 3 cm forward (for estimates in depth), a second cube appeared on the table (see Figure 35c). Participants continued moving towards the second cube, and after selecting this cube they made a verbal estimate of the perceived distance between the first and second Rubik's cube (see Figure 35d).

For the egocentric distance estimation task the same virtual environment was used, however, in this task only one Rubik's cube appeared on the table at different distances. Again, participants started at the location of the red ball, and moved the pointer to the cube in front of them. When the cube was selected they were again asked to make a verbal estimate of the distance between themselves and the cube.

When participants completed the egocentric distance judgments, we presented the objects for the last time, and asked participants to judge the size of the red ball, the size of the Rubik's cube and the length and width of the table that they were seated at respectively.

Setting and Apparatus

The experiment was carried out in the VR/e lab of the TU/e. This room was simulated in a virtual environment programmed in WorldViz Virtual Reality toolkit Vizard, and presented on a Head Mounted Display (HMD). In this experiment, we used a NVIS nVisor SX111 HMD, with 102H x 64V degrees FOV (111 degrees diagonal), and a resolution of 1280x1024. The stereoscopic view on the HMD was created by presenting a different view for each eye, with an overlap of 50 degrees (66%). The head position was measured using a 3 DOF wireless InterSence IntertiaCube3 position tracker. Both the position of the glove and the device were tracked using the PhaseSpace impulse position tracker, tracking a LED marker mounted on both the glove and device (see Figure 36). To steer the cursor, a white flexible glove was used with a LED marker on the pointing finger. In addition, both the thumb and pointer finger were mounted with a wire, that gave a signal when both fingers are tapped together, used for the selection of objects. The device consisted of a wooden stick mounted with a press button and a LED marker. Users pressed the button to select objects, and could move the stick to steer the cursor.



(a) (b)

Figure 36. *Screenshots of the experimental setting, showing two participants wearing a head mounted display and interacting in the VE using a glove (a) or holding a device (b)*

Procedure

Upon arrival at VR/e lab, participants were tested for their stereo acuity using the Randot® stereotest. After the stereo acuity test, participants read and signed a consent form, and were made aware that verbal responses were being recorded. In addition, we asked participants to indicate if they experienced dizziness or nausea, so that we could stop the experiment. When participants were seated behind the table and the HMD was placed on their heads, participants received instructions concerning the task. Participants were assigned to one of the four conditions discussed above. Half of the participants started with the exocentric distance estimation task along the x-axis and half of the participants started with estimations along the z- axis. Participants first completed 30 distance estimates in one direction (x or z), followed by 30 distance estimates along the other axis. When participants finished the exocentric distance estimation task, we took off the HMD and asked them to fill in a questionnaire concerning hedonic and pragmatic qualities, and questions regarding embodiment and fatigue. After participants had completed the questionnaire, the HMD was again placed on their head, and they continued with the egocentric distance estimation task. Participants used the same gain and interaction method as during their exocentric distance estimation task. When they had completed the 10 distance estimates, we also asked them to judge the size of the red ball, the size of the Rubik's cube and the length and width of the table. At the end of the experiment, participants were thanked for their participation. The experiment took approximately 30 minutes and participants received a compensation of 5 euro for their time.

Measures

Distances were verbally estimated by participants in cm, and later converted to percentage scores representing under/overestimation of distances (see measure section in paragraph 5.2.1). In addition, we included the AttrakDiff questionnaire (for an elaborate description of the items and the scales we refer to Chapter 3). The items for pragmatic quality (PQ) were internally consistent with Cronbach's $\alpha = .64$, hedonic quality-stimulation (HQS) with $\alpha = .79$, hedonic quality-identification (HQI) with $\alpha = .67$ and attractiveness (ATT) with $\alpha = .59$. To measure experienced embodiment and fatigue we included eleven items inspired by an existing presence questionnaire (Witmer & Singer, 1998) and previously used questions applied in Chapters 3 and 4. A Principle Axis Factoring analysis with Oblimin rotation revealed four factors based on Kaiser's criterion. However, our aim was to measure embodiment and fatigue and therefore we forced the analysis to extract two factors. Results showed that embodiment consisted of 6 items (Directness, Naturalness, Interaction naturalness, Overall estimation, Involvement) with an internal consistency of $\alpha = .72$. Fatigue consisted of one item (fatigue). The items difficulty, competence, body involvement, and focus did not belong to these two factors and were excluded from further analyses.

Statistical Analysis

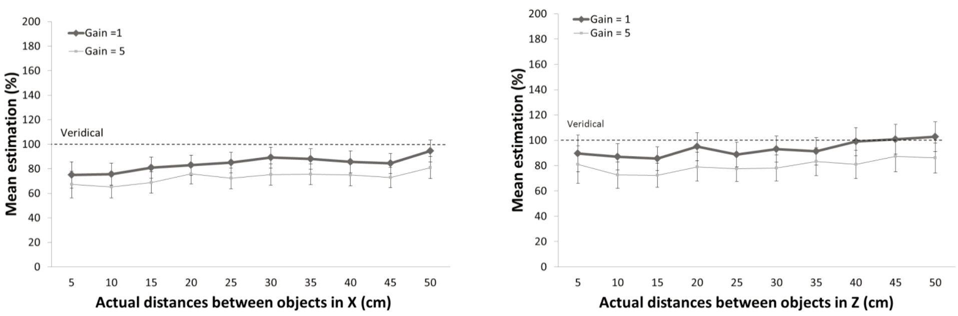
Two participants had complaints of nausea or headache while being exposed to the virtual environment, and hence their data were not included in the dataset. In addition, missing data were removed from the data set (one case) and within each distance we regarded distance estimates exceeding ± 3 SD as outliers and replaced these values with a distance estimate corresponding to M ± 3 SD (1.1% of the data for exocentric and .2% of the data for egocentric distance estimates). We averaged the three estimates per distance for each participant. For exocentric distance estimation these average scores were submitted to a 2 (Gain: 1 vs. 5) x 10 (Distance) x 2 (Direction: x vs. z) x 2 (Interaction method: glove vs. device) repeated-measures ANOVA, with Gain and Interaction method manipulated between participants, and Distance and Direction manipulated within participants. For egocentric distances a 2 x 6 x 2 repeated-measures ANOVA was used analyzing the effects of Gain (1 vs. 5), Distance (6 distances), and Interaction method (glove vs. handheld). Distance was manipulated within participants, and Gain and Interaction method were manipulated between participants. The effect of reaching distance on exocentric distance estimates were analyzed with a 3 x 2 x 6 x 2 repeated-measures ANOVA, analyzing the effects of Reaching distance (50, 65, and 80 cm), Gain (1 vs. 5), Distance (6 distances), and Interaction method (glove vs. handheld). The questionnaires and estimates for object sizes (i.e. red ball, table width and length, Rubik's cube) were analyzed using a Univariate ANOVA with Gain and Interaction method as independent variables. For the size estimation of the ball, the data of one person was missing, so this analysis was performed with 76 participants. For exocentric distances estimations the test of Sphericity was violated for Distance and Distance x Gain, and for egocentric distances the test of Sphericity was violated for Distance. For these effects, we applied Greenhouse-Geisser corrections. In the current experiment, effect sizes are again reported in terms of partial η2 and Cohen's d, as discussed in paragraph 5.2.1. We will use d_s when effect sizes are calculated for between-group comparisons and d for within group comparisons as explained in paragraph 2.4.1.

5.3.2 Results

First, we will discuss the results of the exocentric distance estimation task. Subsequently, we will discuss the results of egocentric distance estimation and size estimation of objects in the room. Finally, we will discuss the results of the questionnaires.

Exocentric distance estimation

The results showed that participants' estimates were on average below 100%, revealing underestimation in their distance estimates. A one-sample t-test confirmed that distances were underestimated in both the z ($M = 86.66$, $SE = 1.29$; $t(749) = -10.31$, $p < .001$, $d = .37$) and x direction ($M = 78.68$, $SE = 1.02$; $t(749) = -20.83$, $p < .001$, $d = .76$). The repeated-measures ANOVA with estimations on the exocentric task as the dependent variable showed a main effect of Gain [$F(1, 71) = 4.30$; $p < .05$, partial $\eta 2 = .06$], indicating larger estimates for a gain of 1 ($M = 88.75$, $SE = 4.19$) than for a gain of 5 ($M = 76.38$, $SE = 4.25$; $d_s = .48$). Results also revealed a main effect of Distance [$F(2.3, 165.9) = 9.02$; $p < .001$, partial $\eta 2 = .11$], indicating that smaller distances were underestimated proportionally more than larger distances (see Figure 37).



(a) (b)

Figure 37. *The effect of Gain on exocentric distance estimation along the x-axis (a) and z-axis (b). The scores are presented in percentage under/overestimations; scores over 100 percent represent overestimations and scores under 100 percent indicate underestimations. The error bars depict 95% confidence intervals. Results showed that distances in both X and Z directions were underestimated. In addition when participants interacted with a gain of 1 estimates were larger (less underestimated) than when using a gain of 5.*

Furthermore, results showed that estimates in the horizontal direction, i.e., along the x-axis, were underestimated more ($M = 78.58$, $SE = 2.74$), than estimates in depth, i.e., along the z-axis ($M = 86.55$, $SE = 3.57$; $d = .29$) with [$F(1, 71) = 12.77$ $p < .01$, partial $\eta 2 = .15$]. The interaction between Distance and Direction was also significant [$F(6.0, 427.7) = 4.64$; $p < .001$, partial $\eta 2 = .06$]. Post-hoc comparisons with Bonferroni correction, indicated that estimates for distances of 15 cm and 30 cm did not significantly differ between the x-axis and the z-axis, whereas the other distance estimates in the x-direction were smaller than those in the z-direction. Using a glove or a device did not have a significant main effect on the distance estimates ($F < 1$, ns), suggesting that the interaction method did not affect participants' distance perception. In addition, the analyses showed no significant interaction effects between Distance and Gain, Direction and Interaction method, Direction and Gain, or Interaction method and Gain (all $F < 1$, ns). The results further showed no three-way or four-way interactions between Distance, Direction, Interaction method, and Gain (all $F < 1$, ns). Estimates along the x-axis were performed at three different distances (e.g., 50, 65, and 80 cm) from the observer. To test whether estimates between objects along the x-axis depended on the reaching distance, a repeated-measures ANOVA was performed with distance estimation as the dependent variable and Reaching, Gain, Tool, and Distance as independent variables. Adding reaching distance to the model did not affect the earlier results (i.e., again Gain and Distance were significant, whereas Interaction method did not affect participants' distance estimates). Therefore, we will report only the effects on Reaching distance and the interactions between Reaching and the other independent variables.

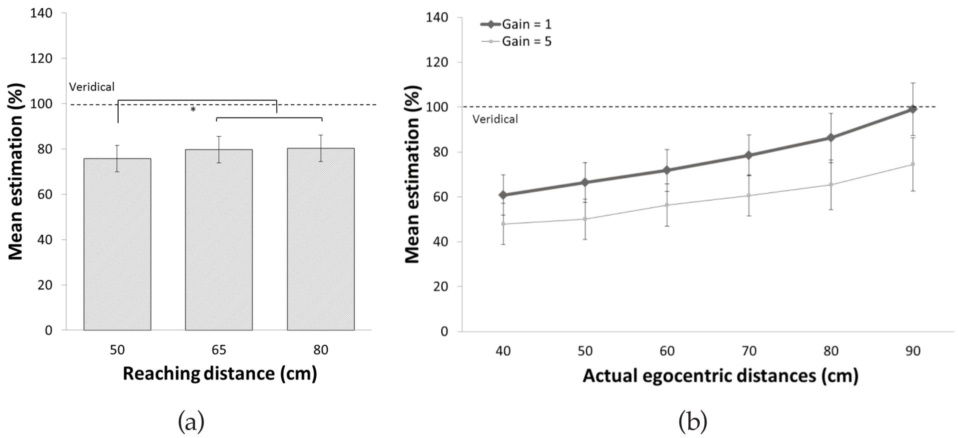


Figure 38. The effect of reaching distance on exocentric distance estimation (a), and the effect of Gain on egocentric distance estimation between the observer and the cube on the table (b). All scores are presented in percentage under/overestimations; scores over 100 percent represent overestimations and scores under 100 percent represent underestimations. The error bars depict 95% confidence interval. Results in (a) showed that when participants had to reach further (e.g., 65 and 80 cm) the estimated distance between the cubes on the x-axis was larger compared to a distance of 50 cm. Results in (b) showed that in both gain settings distances were underestimated, however egocentric distances were perceived larger when interacting with a lower gain.

As shown in Figure 38a, results showed a significant main effect of Reaching [$F(1, 142) = 13.01$; $p < .001$, partial $\eta 2 = .16$], indicating that estimated distances between the two cubes on the x-axis were estimated larger when participants had to reach farther (i.e., objects were located further from the participant). Post-hoc tests with Bonferroni correction showed that at reaching distances of 50 cm ($M = 75.73$, $SE = 2.96$), estimates were smaller than estimates at 80cm ($M = 80.28$, $SE = 2.96$; $d = .18$) and 65 cm ($M = 79.68$, $SE = 2.96$; $d = .15$) from the participants. The three-way interaction between Distance, Reaching and Interaction method [$F(10, 767) = 1.71$; $p = .07$, partial $\eta 2 = .02$] was not significant, nor were any other two, three, or four-way interactions between Reaching, Distance, Interaction method, and Gain (all $F<1$, ns).

Egocentric distance estimation

A one-sample t-test showed that the estimates of egocentric distances were also underestimated ($M = 67.67$, $SE = 1.58$; $t(455) = -20.47$, $p < .001$; $d = .96$). Comparable to the results on the exocentric estimation task, a repeated-measures ANOVA showed a main effect of Gain [$F(1, 71) = 7.36$; $p < .01$, partial $\eta 2 = .09$], revealing that a gain of 1 resulted in higher estimates ($M = 77.14$, $SE = 4.66$) than a gain of 5 ($M = 59.15$, $SE = 4.72$; $d_s = .62$; see Figure 38b).

Similar to what was reported in earlier paragraphs, smaller distances were underestimated proportionally more than larger distances [$F(2.2, 155.7) = 62.16$; $p < .001$, partial $\eta 2 = .47$]. The interaction between Distance and Gain was not significant [$F(2.2, 155.7) = 1.85$; $p = .1$, partial $\eta 2 = .03$]. Results revealed no significant main or interaction effects of Interaction method (all p's > .10).

Object sizes.

Figure 39 presents the results for the size estimation of the objects used in this experiment. A one sample t-test showed that only the red ball was significantly underestimated ($M = 84.63$, $SE = 4.29$; $t(73) = -3.59$, $p = .001$). Results of a Univariate ANOVA with Gain and Interaction method as fixed factors revealed that the diameter of the red ball was estimated smaller using a gain of 5 ($M = 73.18$, $SE = 5.96$) than using a gain of 1 ($M = 95.68$, $SE = 5.79$; $d_s = .62$) with [$F(1,70) = 7.34$; $p < .01$, partial $\eta 2 = .10$][16]. Table length was underestimated for a gain of 5 ($M = 93.99$, $SE = 5.61$), but overestimated using a gain of 1 ($M = 111.35$, $SE = 5.53$; $d_s = .51$) with [$F(1,71) = 4.85$; $p < .05$, partial $\eta 2 = .06$].

[16] *For this analysis the data of one participant was missing (in condition with a gain of 5 combined with the glove)*

Chapter 5

Although the width of the table was estimated near its veridical value for a gain of 1 (M = 102.37, SE = 5.09) and slightly underestimated using a gain of 5 (M = 91.00, SE = 5.16; d_s = .36), this difference was not significant [F(1,71) = 2.46; p < .12, partial η2 = .03]. The Rubik's cube was also underestimated using a gain of 5 (M = 82.79, SE = 5.39), however slightly overestimated using a gain of 1 (M = 107.11, SE = 5.31; d_s = .74) with [F(1,71) = 10.3; p < .01, partial η2 = .13]. Results did not reveal a significant main effect of Interaction method, nor an interaction between Gain and Interaction method (all F's < 1; n_s).

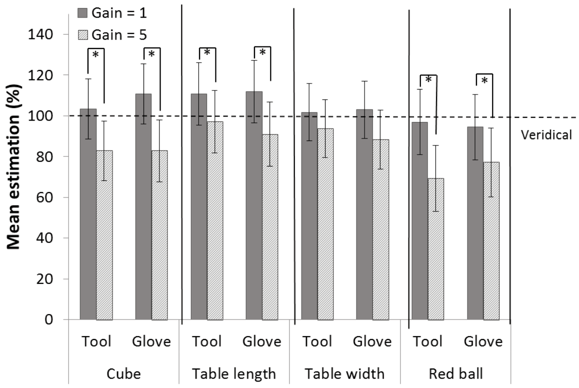


Figure 39. *Size estimations of the various objects (cube, table length, table width red ball) in the environment as a function of Gain and Interaction method. The scores are presented in percentage under/overestimations; scores over 100 percent represent overestimations and scores under 100 percent indicate underestimations. The error bars depict 95% confidence intervals. The comparisons indicated with a * are statistically significant (p < .05), Results showed that interacting with a lower gain increases sizes estimates of the cube, table length and the red ball.*

Subjective measures

The experience of participants interacting in the VE was measured in terms of embodiment, fatigue, pragmatic and hedonic quality and attractiveness (see Figure 40). A Univariate ANOVA revealed no significant main or interaction effects of Gain, or Interaction method on experienced embodiment (all F's < 1; ns). For fatigue a non-significant trend of Interaction method was found [F(1, 71) = 3.70; p = .06, partial η2 = .05], suggesting that the tool induced more fatigue (M = 4.10, SE = .25) than the glove (M = 3.41, SE = .25; d_s = .45). The results did not reveal a main effect of Gain [F(1, 71) = 1.33; p = .25, partial η2 = .02] nor an interaction between Gain and Interaction method [F(1, 71) = 1.13; p = .29, partial η2 = .02] on experienced fatigue. As shown in

Figure 40, neither Gain, nor Interaction method, had significant main or interaction effects on pragmatic quality (all F's < 1, ns). Similarly, Gain and Interaction method had no significant effects on hedonic quality-stimulation, hedonic quality-identification or attractiveness (all p's > .10).

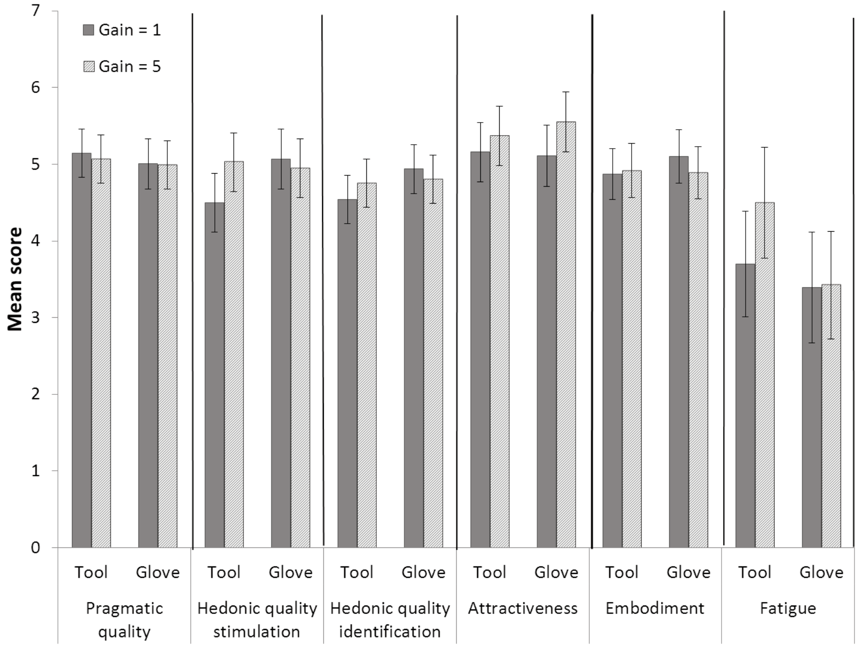


Figure 40. *Results from the AttrakDiff questionnaire (Pragmatic quality, Hedonic quality-stimulation, hedonic quality-identification and Attractivess), experienced embodiment and fatigue as a function of Gain and Interaction method. The error bars depict 95% confidence intervals. Results showed that gain levels and interaction methods revealed similar scores in terms of the various user experience factors.*

5.3.3 *Summary of results*

In the current experiment, the aim was to replicate and extend the findings reported in Experiment 8. The size of the hand movements was again varied by changing the gain of the interaction, while keeping the visual feedback to participants equal. The gain manipulation allowed us to test whether perception of distance is related to a person's movements – in other words, to what extent perception in mediated space is embodied. In line with our hypothesis, participants estimated distances as smaller when arm movements were smaller and vice versa. Results showed moderate effects for both exocentric (d_s = .48) and egocentric distance (d_s = .62) estimations, as well as for size estimations (between d_s = .36 to d_s = .74). Similar results were found for estimations of object sizes, revealing the largest difference between the two gain levels for the red ball (d_s = .62) and the Rubik's cube (d_s = .74). For both the table length (d_s = .51) and table width (d_s = .36; *ns*) the effects were much smaller. In addition, the results showed that the red ball was estimated significantly smaller than veridical, whereas, the table length, table width and Rubik's cube were, on average, estimated close to the original sizes.

Our second research question concerned the impact of the interaction method on distance perception. Results of the current study did not show an effect of interaction method, therefore the hypothesis that interaction method influences distance estimation cannot be accepted. In addition, we studied if reaching distance affects distance estimates, and found that reaching did indeed affected participants' distance estimation. When the two cubes were located 50 cm from the observer, the same distances were perceived as smaller than for objects placed at 65 cm or 80 cm from the observer, although the effect sizes were small (d = .15 and d = .18 respectively).

Lastly, we addressed user experience in terms of terms of hedonic and pragmatic quality, embodiment, and fatigue. Results revealed that neither gain, nor interaction method changed participants' responses in terms of hedonic and pragmatic quality. Also in terms of embodiment and fatigue, both gain conditions revealed similar experiences. Although perceived embodiment did not differ between glove and tool based interaction, holding a device did induce slightly more fatigue.

5.4 Discussion

In the current chapter, we investigated embodied perception of distances in mediated space. In line with Gibson (1979) and O'Regan and Noë (2001), we predicted that perception and action would go hand in hand, and that bodily action would play an important role when perceiving and interpreting the virtual environments. When interacting with objects in the real world, the relation between our own movements and the perception of those movements is generally one-to-one. When we move our hand 10 cm, our perception of the same movement is also 10 cm. We have therefore learned to integrate both visual and physical information when estimating the location and distance of objects. However, in mediated environments, the gain is under the user's or application designer's control. If the gain in an application differs from one, this might affect our perception of the environment, as physical feedback may bias the interpretation of the visual stimuli. In the current chapter, we carried out two experiments to test this hypothesis. In both experiments we manipulated gain, and administered a similar distance estimation task, first in a 2D screen environment (Experiment 8), and second in an immersive 3D virtual environment (Experiment 9). In the second experiment, we included not only exocentric, but also egocentric estimates and size estimates. We employed a gain of 1 (1 cm of movement equals 1 cm of visual displacement) and a gain of 5 (1 cm of movement equals 5 cm of visual displacement). Although in Experiment 8 results did not reveal a significant main effect of gain, effect sizes were similar to those found in Experiment 9. For exocentric distance estimation along the horizontal axis a medium effect was found in both Experiments ($d_s \approx .49$). For the egocentric distance estimation in Experiment 9, the effect was slightly larger ($d_s = .63$). In sum, results of both experiments indicated that when movements were larger (e.g., lower gain), distances were perceived as larger, which constitutes an assimilation effect as observed in various experiment reviewed by Zwickel and Prinz (2012). This suggests that people integrate visual and physical (motor control) when estimating sizes and distances within the environment, which is in line with previous thoughts on action-perception coupling (Gibson, 1979; O'Regan & Noë, 2001, Zwickel & Prinz, 2012). In the introduction three mechanisms (proprioception, efferent copy and effort) were mentioned that might be relevant for explaining the results in this chapter. In the current experiments we were not able to differentiate between these mechanisms. To better understand the underlying mechanisms, more research is required, where the experimental procedure described in this chapter can be used as point of departure. For example, to better understand to what extent effort contributes to estimations of distances, the weight of a controller,- and thus the effort required when interacting with this device -, can be manipulated. In addition, by only varying the weight of a controller, proprioceptive information remained constant and therefore more information is gained on the processes underlying embodied interaction.

In addition to exocentric and egocentric distances, we also asked participants to estimate sizes of objects in the room (e.g., table length, table width, red ball and Rubik's cube). Results showed that the effects of a gain manipulation were very similar to those found for exocentric and egocentric distance estimations: objects were perceived to be larger when the gain was lower and vice versa. However, in terms of under/over estimation the difference between the two gain levels was most pronounced for the red ball ($d_s = .62$) and the Rubik's cube ($d_s = .74$), while the effect was smaller for both the table length ($d_s = .51$) and table width ($d_s = .36$; ns). A potential explanation is that participants saw the physical table when entering the experimental room, making it likely that participants had a fair idea of its physical dimensions. Both the red ball and the Rubik's cube were not physically present in the real environment, rendering its size estimates more sensitive to experimental manipulation.

In the current chapter, our studies showed that computer displays resulted in an overestimation (d between .49 and .88), whereas HMDs resulted in an underestimation (d between .37 and .96). Previous studies by Roscoe (1984) and Waller (1999) also reported overestimation on computer displays, potentially caused by the border around the screen, which can act as a visual anchor. Also, studies by Witmer and Kline (1998), Loomis and Knapp (2003), and Thompson and colleagues (2004), corroborate our findings concerning underestimation of egocentric distances wearing a HMD. In contrast, earlier studies reporting exocentric distance estimation in VR revealed overestimation or near veridical performance (Waller, 1999; Wartenberg & Wiborg, 2003; Richardson & Waller, 2007), which is different from our findings in Experiment 9. Similar to our study, Richardson and Waller (2007) also studied both exocentric and egocentric distance judgments, and found a large underestimation for egocentric distance judgments, and a near veridical performance for exocentric distance judgments during the pre-test (in the other trials users received feedback). This showed that users were more accurate in estimating exocentric distances, which is in line with previous studies (Cutting & Vishton, 1995). Although in our study exocentric distances were also estimated slightly more accurate, both exocentric and egocentric distances were underestimated. In the study by Richardson and Waller (2007), participants estimated distances using blind folded walking, and saw both target objects at the same time during the exocentric distance judgments (in contrast to our study where participants saw the starting point and the destination point in sequence and verbally estimated the distances). Previous literature showed that participants are less accurate in estimating distances verbally, compared to blindfolded walking (Cutting & Vishton, 1995). In addition, our sequential presentation of the target objects was more ambiguous and therefore more difficult to estimate. Both factors could explain why the results of Experiment 9 are different from those of the study of Richardson and Waller (2007).

When interpreting the under- and overestimations in terms of accuracy another interesting finding emerged. For estimations in virtual environments a gain of one resulted in more accurate distance estimations, whereas on a display the highest accuracy was found for a gain of 5. This result can be interpreted in two ways. First, higher gain levels may give rise to smaller distance estimations. Therefore, on regular 2D displays where distances are generally overestimated, a higher gain level leads to improved accuracy. In virtual environments on the other hand, distances are generally underestimated. Thus, a lower gain level leads to more accurate estimations. However, an alternative interpretation is related to the way we are used to interact with displays and in the (virtual) world. On computer displays the gain is normally larger, since this leads to more efficient display work (Johnsgard, 1994; Casiez et al., 2008). For immersive virtual environment, such as wearing a HMD, realism and naturalness are often paramount, and therefore a one-to-one correspondence between our movements in the real and virtual world would appear most sensible. This presents yet another trade-off between naturalness and efficiency one needs to consider in designing (3D) interaction.

In the current chapter we also explored whether interaction method influenced distance estimates. We hypothesized that with more direct interaction, participants would rely more on arm and hand movements for their estimations than during interaction via a device. Results of Experiment 9, however, showed similar distance estimations for participants holding a tool and those interacting with a glove. A reason might be that users did not see a virtual representation of the tool in the virtual environment, resulting in the same visual stimuli in both tool and glove conditions. Perhaps results would have shown an effect of tool use on participants distance estimations, if the tool itself had been visualized within the virtual environment. In addition, we tested whether reaching distance influenced the perception of distances between objects. Our hypothesis was that when participants had to reach further, their estimations of distances would increase. Results of Experiment 9 confirmed this hypothesis, showing that distances between objects were perceived as larger when objects were further away (i.e., 65 and 80 cm), compared to estimations 50 cm away from the observer. Since effect sizes were small ($d \approx .15$) more research is needed. Nevertheless it shows a subtle trend that when the reaching distance increased, distances between objects were perceived as larger. This result might serve as an alternative explanation of results found by Witt et al., (2005), who argued that a stick extends our reaching area and therefore distances are perceived as closer than estimations without a stick. However, when pointing to an object while holding a stick, the reaching distance is shorter than pointing to that object without a stick, which can be a potential underlying mechanism explaining the results. This opens new research questions studying the relation between reaching distance and distance estimations in more detail.

In contrast to Chapters 3 and 4, experiences in terms of both hedonic and pragmatic quality were similar for the two interaction methods (glove and device-based interaction). However, in both Chapters 3 and 4 participants did not wear gloves while interacting. Although participants can also freely move their hands and fingers while wearing gloves (similar as for gesture-based interaction), the experience of wearing a glove or interacting without a glove might be different. In addition, experienced hedonic and pragmatic quality did not differ between the two gain levels. Furthermore, we asked participants to reflect on their sense of embodiment and feeling of fatigue in relation to the interaction method. Whereas in terms of embodiment no differences emerged between glove and device-based interaction or between the two gain levels, participants did experience slightly more fatigue when using a tool ($d = .45$), although this trend did not reach significance. Since we used a between-subjects design, the lack of a reference (i.e., participants were not able to compare the two interaction methods or gain levels with each other) may lead to a tendency to mostly use the center of the scale, which might explain the findings in the questionnaire data. If we had allowed participants to compare the interaction methods with each other, as was done in Chapters 3 and 4, results might have revealed differences between the two interaction methods and gain levels. Therefore, more research is needed to study whether, and to what extent, gain level and wearing gloves change the experience of users when interacting with digital content.

A limitation of Experiment 9 was that the estimation of egocentric distances and object sizes was always performed during the second half of the experiment; i.e., after completing exocentric distance estimation. In Experiment 8, results suggested that the gain that participants were presented with during the first block may have been used to build up a frame of reference, which was then used throughout the experiment. Similarly, experiences during the exocentric distance estimation task may have influenced estimations of egocentric distances and object sizes. However, since both Experiments 8 and 9 revealed the same trend, indicating that larger movements resulted in larger distance estimates, we do not expect effects will be different when the order of the estimation tasks is reversed.

A second limitation could lie in the fact that participants only saw one object (rectangle, or cube) at a time, to ensure they interacted with the environment before estimating the distance. This procedure may have made estimation more difficult, and rendered it more sensitive to gain manipulation. On the other hand, naturalistic environments and tasks – e.g. medical imagery – often present information of high ambiguity, perhaps also resulting in a relatively strong dependence on motor system in perceptual judgments. In future research it would also be interesting to explore participants' perceptions in complex – naturalistic – environments after prolonged interaction with this content.

5.4.1 Practical implications

In this chapter, we learned that perception of distances in mediated space is affected by the movements we make during interaction. In the real world our physical movements and perceptual changes are generally directly and unambiguously related as one-to-one, whereas in virtual environments this relation depends on the gain level implemented. Findings in this chapter demonstrate that the size of our hand movements influences our distance estimations. The effect sizes demonstrated a medium effect of gain on distance estimates in both Experiments 8 and 9. This is a relevant consideration when developing interactive virtual reality applications, as such design choices will impact both task performance (Johnsgard 1994; Casiez et al. 2008), and our perception and experience of the environment we are interacting with. For many applications, such as product development, training simulations, computer-aided design, computer-aided manufacturing, and medical surgery a veridical perception of the environment is important, and therefore potential misperception in virtual environments should be taken into account when designing virtual reality applications.

CHAPTER 6

General discussion

"Machines that fit the human environment instead of forcing humans to enter theirs will make using a computer as refreshing as taking a walk in the woods"
(Marc Weiser, 1991, p. 89)

"I believe we will look back on 2010 as the year we expanded beyond the mouse and keyboard and started incorporating more natural forms of interaction such as touch, speech, gestures, handwriting, and vision--what computer scientists call the "NUI" or natural user interface." (Steve Ballmer, CEO Microsoft, 2010)

We are living in an era of spectacular technological progress. Since the dawn of the industrial revolution, technology has penetrated almost all realms of life and is supporting a vast array of activities and tasks. Media technologies, including computers, mobile phones and televisions, have become an intrinsic part of our domestic, leisure and work environments. At present, three significant trends are of particular importance to this thesis. First, we are witnessing a significant diversification of media technology, ranging from small, wearable displays (mobile computing), to large-scale ambient and immersive 3D environments. Secondly, new sensing technologies are offering opportunities for user interactions to move away from the constraints of the traditional keyboard/mouse combination, towards gesture-controlled interfaces, multitouch surfaces, face and voice recognition, activity sensing, context sensing, and natural-language. These interfaces have in common that they increase the bandwidth of human-machine interaction, engaging the body to a greater extent, and potentially affording more intuitive interactions than previously possible. At the same time, a third trend can be observed within the domain of human-computer interaction: A shift in application purpose from productivity-oriented technologies where performance is a key objective, to applications for everyday life, that aim for user experiences through leisure, play, culture and art. Such a shift in application purpose is reflected in the concepts and metrics that are being used to describe and measure the relevant user experiences, and to optimize technology accordingly. The current thesis is located at the intersection of these trends, studying a broad range of user experiences in relation to 3D display environments and 3D interaction technologies.

As a point of departure, we have argued in this thesis that interaction methods should match our abilities, needs and preferences, such that the interaction becomes transparent and our focus can be on the task at hand and not on the interface technology (see also Winograd & Floris, 1986). When looking specifically at 3D interfaces, the focus of previous research to date has primarily been on the performance benefits that 3D interaction affords. Whereas 3D displays allow for more intuitive and realistic visualization of 3D datasets than their 2D counterparts, their recent introduction in people's living rooms, as part of the users' need for leisure and relaxation, also calls for a reconsideration of the relevant interaction quality metrics in relation to 3D interfaces. Before we can fully benefit from the third dimension in the displayed environment, better interaction methods should be developed, such that the spatial nature of the representation and the input device are intuitively mapped. In this dissertation we have argued that instead of focusing on what is technologically feasible, interaction technologies should be designed and studied from a user-centered perspective. Taking the perspective of embodied interaction, we extended currently applied methodologies and studied stereoscopic visualizations and both traditional, device-based interaction methods as well as novel, deviceless (i.e., gesture-based) interaction methods. The work presented in this thesis has implications at both the theoretical and the methodological level, and presents recommendations for the design of 3D displays and interaction technologies.

6.1 Main contributions

The types of contributions of this thesis are threefold: we have added to discussions on research methodology in relation to 3D interfaces, developed new insights in perception effects as a consequence of 3D interaction methods, and formulated implications for the design of 3D displays and interaction technologies.

First, we noted that evaluation methodologies currently applied when evaluating 3D displays and interaction technologies are often limited to efficiency measures such as completion times and accuracy, and to pragmatic qualities when considering usability. In Chapter 2, we applied the concept of perceived workload in addition to completion times and accuracy, to better understand the benefits of stereoscopic visualization in a performance-oriented context. Results showed that completion times benefits most from stereo visualizations. In addition, stereoscopic visualizations may reduce workload, yet the disparity level used to visualize the content was proved to be an important factor reducing workload. In Chapter 3, we applied a broader perspective of user experience studying embodied interaction. User experience factors such as hedonic quality and fun were shown to be relevant measures in addition to usability items when studying users' experience of interaction with 3D content. In Chapter 4, we replicated and extended findings of Chapters 2 and 3 showing that hedonic quality, fun and affect are relevant measures not only when studying embodied interaction, but also when watching content on 3D displays.

Second, we demonstrated that embodied interaction not only affects the experience of users interacting with technology, but also impacts fundamental processes of perception through the integration of visual and physical (motor) information. In Chapter 5, we measured participants' distance perception on a 2D display and in a 3D virtual environment. Participants interacted with content using different gain settings, allowing us to manipulate the amplitude of their hand and arm movements, while leaving the visual feedback unchanged. Results showed that visual distance estimates were affected by body movements, such that larger hand and arm movements elicited larger estimates for the same distances than did smaller hand movements. This finding not only demonstrates the dependency of action and perception, but also has implications for designing mediated environments as we will discuss next.

The third contribution concerns the implications for the design of 3D display applications and embodied interaction technologies. For 3D displays, we established the importance of using a range of disparities when studying whether stereoscopic displays can yield a performance advantage. We subsequently were able to determine the optimal level of disparity when performing a complex spatial task (Experiment 1) and showed that disparity levels between 10 - 20 min of arc were preferred. In addition, we demonstrated that a combination between motion (object motion or movement parallax) and stereoscopic visualization yields the most efficient performance (Chapters 2 and 4). In Experiment 9 (Chapter 5) we showed that in embodied interaction, certain parameters that are under the control of interaction designers may influence the perception and interpretation of visualized content. In particular, we demonstrated that the fundamental variable controlling the ratio of output to input, that is, the gain of the interaction device, impacts distance estimates and should therefore be taken into account when accurate distance and size estimates are critical.

6.2 Limitations and future directions

The current thesis has some limitations regarding the generalizability of the results. The first limitation concerns the population sample that was used in the studies reported in this thesis, which consisted of university students (and junior employees) that were roughly between 18 and 30 years of age. Therefore, results cannot be generalized to groups of people such as children or seniors, who might experience these technologies differently than the student population (as one would also predict based on Gibson's notion of affordances). Participants that took part in our experiments had little to no experience with gesture-based interaction with stereoscopic displays (as mentioned previously, at the time of the studies stereo displays were not yet widely available and Microsoft had not yet launched their Kinect). Second, in both Chapters 2 and 4 we used a path-tracing task to measure the benefits of stereo visualization compared to monoscopic visualization. In this particular task, no pictorial depth cues were available, and therefore it is unclear whether results found for this task can be generalized to more realistic settings such as angiography, in which the bends of the blood vessels are less extreme and therefore better corresponds to the law of good continuation than stimuli used in our experimental task. In addition generalization across other tasks containing a wider variety of pictorial depth cues remains as yet unexplored. Moreover, earlier studies have indicated that experienced users, such as medical doctors, have learned to use pictorial depth cues in such a way that stereo did not always increase performance, whereas for inexperienced users stereo was always advantageous (see e.g., Beurden, 2012). Therefore, additional research is needed to find out which persons, in which situations, benefit from stereo visualization, measured in terms of task performance, workload, and user experience.

A third limitation pertains to the way in which we assessed user experience and perceived workload. In the current thesis, we used questionnaires when assessing experienced and workload, whereas other methods – e.g., physiological measures or a secondary task (O'Donnell & Eggermeier, 1986) – are also available. Also in terms of user experience additional measures can be applied, including behavioral data such as smiling, pressure exerted on the interface, or finger tapping (e.g., van den Hoogen, IJsselsteijn & de Kort, 2008), or physiological measures assessing participants' emotions such as pupil diameter or heart rate (Fairclough, 2009). We chose to use questionnaires since these do not intrude upon the task and provide an accurate view on how users experience the technology. In future research, these methods can be applied in combination with questionnaires to give more insight in participants' perceived workload and user experience during embodied interaction and while performing tasks on 3D displays.

The gesture technology used in the current thesis was developed before the introduction of commercial body trackers such as the Microsoft Kinect. The gesture technology applied in this thesis was developed within the HELIUM3D project and had not yet reached its optimal level of performance. This tracker sometimes faltered, which may have distracted participants from the task, affecting their performance, experience and perception of workload. Although we expect user experience and performance to increase with more robust tracker technology in future experiments, user expectations and abilities may also transform as commercial trackers become widely available. This could affect the evaluation of gesture-based technologies in different ways. First, with more matured gesture-based interaction methods, either through standardization or through convention, users may be more able to interact efficiently using gesture-based technologies, thus improving the usability and pragmatic elements of the interface. On the other hand, as discussed by Karapanos and colleagues (2009), hedonic values of an interface, such as novelty, surprise and mystery, may be appreciated to a lesser extent as the interface becomes more 'mundane' through everyday use. As the novelty of gestures as interaction style is likely to wear off though extended use, its human factors costs may rise, since the physical effort one needs to expend in interacting with a gesture-based interface may give rise to physical fatigue and discomfort (as shown in this thesis), and even repetitive strain injury (Bonis, 2007). Although we expect gestures to find their appropriate place in the repertoire of interaction systems, the balance between perceived costs and benefits may still shift significantly in the future, thereby necessarily limiting the generalizations that can be made based on the work reported in this thesis.

6.3 Frequently asked questions

During the years I have been working on this thesis various questions were frequently asked, either during project work, conferences, or from family and friends. Based on these questions I formulated five questions which I will use to discuss the results and implications of this thesis work.

6.3.1 Do we need three-dimensional displays?

Studies from this thesis, as well as previous literature have shown that stereo clearly increases task performance. In addition, image quality factors such as naturalness and viewing experience score better when images were viewed stereoscopically. Over the last few years, stereoscopic displays have become more affordable and various applications support stereoscopic visualizations, both in entertainment as well as in professional settings. In Chapters 2 and 4 we showed that stereoscopic displays contribute to a better task performance and lower levels of cognitive load. For entertainment settings, previous studies revealed that images appear more natural (Lambooij et al., 2010) and that our sense of presence increases when images are visualized in stereo (IJsselsteijn, 2004). In line with these findings we found that users also experienced higher levels of enjoyment and positive affect when interacting with stereoscopic content (Chapter 4, experiment 7). The higher level of perceived positive affect can also be relevant for performance-oriented settings, since previous studies have shown that a positive mood can increase creativity and cognitive flexibility (Ashbly, 1999; Davis, 2009; Isen, 2001). Also in the current thesis this positive effect may have contributed to a better performance, however this cannot be determined based on our findings and should therefore be addressed in future research. When participants were able to interact with content by means of object motion or movement parallax, task performance increased compared to static visualizations. A combination of motion and stereo decreased completion times, without losing any task accuracy. Also perception of workload was lower when combining motion with stereo, although the results of this thesis indicated that the disparity level applied in the visualizations is an important factor. In sum, results showed that for both entertainment as well as performance-oriented contexts people can benefit from stereoscopic presentations, either in terms of more intuitive visualization or higher levels of fun, hedonic quality and positive affect.

6.3.2 *What is unique in the embodied interaction perspective?*

Interaction with computers has long been dominated by interaction devices such as the mouse and keyboard. These days, new interaction methods are introduced on the market that rely on our body movements, such as touch screen interfaces, Nintendo Wii and Microsoft Kinect. In HCI literature these interaction methods are seen as more natural, since they are easy to learn and the interaction has become transparent (Bowman et al. 2012; Weiser, 1991; Winograd & Flores, 1988). However, in our view, these interaction methods are not only more natural but also more embodied, since they rely on bodily information such as sensory-motor dependencies and body representations. However, what makes an interaction more or less embodied? Contrary to what one might expect, the level of embodiment is not necessarily related to the size or number of body movements performed during the interaction. In our view, proposed in Chapter 1, an embodied interaction perspective implies interaction in which interaction is not only purposeful by means of accomplishing a task, but also impacts cognition, perception and overall experience of users. Therefore embodied interaction will go beyond what has been traditionally the main goal when interacting with computers - i.e., making interaction more efficient - by making interaction personally relevant in terms of overall experience, cognition and perception. For example, playing a game using the Wii may be more fun than using a traditional controller, or people who use the Wii combine game play with doing a physical workout. Both the experience of fun and doing physical exercises are personally relevant factors, although not contributing directly to in-game performance. In this thesis, we hypothesized that embodied interaction may affect our overall experience since our body is more involved during the interaction and therefore makes interaction more personal. In addition, since our body is more involved during the interaction it might also have repercussions on our perception of the environment. In paragraphs 6.3.3 and 6.3.4 we discuss the findings and implications of embodied interaction found in this thesis.

6.3.3 *Does our experience change when interacting in an embodied fashion?*

In three experiments (Experiments 5, 6, and 7), we studied the experience of embodied interaction and results showed that embodied methods of interaction increased users' experiences in terms of positive affect and enjoyment (Chapters 3 and 4). Results further showed that holding a device (e.g., Wii), or not (e.g., gestures), both changed people's experience in terms of hedonic quality and fun equally. However, in terms of pragmatic quality and perceived performance, hand-held devices such as the mouse and the Wii were preferred. It should be noted that the accuracy of recently developed tracking technologies is promising, and in the near future these technologies will likely further improve such that experiences of pragmatic quality of deviceless interaction may increase and become comparable with device-based interaction. Taking into account these developments, embodied interaction could be applied in a variety of application areas. Gaming is a substantial consumer market were embodied interaction already found its relevance, since it makes games more enjoyable and increases players' sense of being immersed in the game world (Mcgloin et al., 2011; Skalski et al. 2011). Furthermore, since higher levels of affect and enjoyment can increase cognitive flexibility and creativity, also tasks that require a high level of creativity may benefit from embodied interaction as we will discuss next.

6.3.4 Is the 'fun factor' the only merit of embodied interaction?

Our definition of embodied interaction suggests that also cognitive and perceptual processes might be influenced when interacting using an embodied interface method. In Chapter 5 we showed that embodied interaction affects our perception of the environment and objects in the environment. When using larger hand or arm movements, participants interpreted distances between objects as larger. This finding is interesting both for applications that require an accurate representation of the environment (e.g., medicine, military applications, tele-operation), and for applications with which one would exaggerate environmental characteristics (e.g., gaming or virtual worlds). In addition, results are also relevant for designers, showing that settings, such as interaction gain, normally implemented in a fairly ad hoc fashion, can influence users' perception of the displayed environment. Therefore, when developing applications for mediated environments, the gain in which users interact with these environments should be considered more carefully. In addition to these effects, embodied interaction might decrease cognitive load, however our results did not confirm this hypothesis. Perhaps this is due to the technological limitations of the system we used, since the interaction was not yet optimal and the resulting inaccuracies, unresponsiveness and/or errors may have added to the cognitive load instead of decreasing it. However, previous literature showed that movements can support children learning math (Goldin-Meadow, 2010) and that embodied interaction can enhance learning processes (Malinverni, Lopez Silva & Pares, 2012). Therefore, effects of embodied interaction on cognitive processes will remain an interesting theme - in addition to hedonic aspects - for future research as interaction technology progresses.

6.3.5 What makes gesture -based interaction special?

Gesture-based interaction is a special category of embodied interaction, during which people interact with technology using their limbs, usually their arms and hands, without being constrained by a controller. Since different joints in our finger, hand and wrist, can be used during the interaction, it theoretically supports a much wider range of movements and postures than is available when holding a controller. Gesture interfaces may result in more embodied interaction, since users can use hand gestures that are meaningful to them. Nevertheless as discussed in Chapter 1, gesture-based interaction will not necessarily be the most natural interaction in all contexts, and consequently are not likely to replace interaction methods such as the mouse for interaction with a desktop computer. However, since gesture interaction does not require any devices or screens to be touched, it can be useful in a variety of environments, such as in environments with stringent sterility (e.g., medical operating theatre), or environments where controllers are not available (e.g., screens in public areas). At the same time, for specific tasks such as changing the volume, switching channels, turning a light on and off, gestures could elegantly replace the range of different controllers required for a variety of devices at home. Second, as mentioned above, previous research showed that the use of gestures can promote learning (Goldin-Meadow, 2010) and may therefore prove useful in teaching applications. In addition, gestures may also decrease cognitive load, which is most prominent when gestures are meaningful (Cook et al. 2011). Experiment 7, in which participants performed a complex spatial task using gesture or mouse-based interaction, did not reveal a lower level of cognitive load for gestures. Nevertheless, improved tracking accuracy could support a wider range of meaningful gestures, which may lower cognitive load in the future. Results further showed that body fatigue should be taken into account when designing gesture-based systems. As the results of Experiment 7 showed, gestures are currently able to compete with the mouse in terms of task accuracy and completion times, and outperform the mouse in terms of hedonic quality and positive affect. We therefore conclude that, at present, gesture interaction has important benefits, rendering its application in various disciplines advantageous, and making it more likely that it will be more frequently applied in the future.

6.4 Conclusion: embodied interaction in the future

Recent innovations such as the Nintendo Wii, Microsoft Kinect and touch-screens, already illustrate the increasing popularity and application of embodied interaction. In this thesis we learned that interacting with technology is more than providing input to technological applications or systems. The interaction in itself can affect our experience, perception and cognition, such that the motivation of using a particular interaction method is personally relevant and not only driven by efficiency. Gesture-based interactions are one of the important means of interacting with 3D content, supporting an intuitive mapping between the dimensionality of the displayed information and the dimensionality of the input devices used to navigate and manipulate this information. As demonstrated by the progress in media technology over the past 20 years – the era encapsulated between the remarks of Marc Weiser (1991) and Steve Ballmer (2010) quoted at the start of this chapter – we are moving towards a future in which a more holistic, human interaction of people with technology is enabled. As a consequence, the distinction between the real world and virtual reality is becoming increasingly less meaningful. In many cases, this trend is likely to enhance our feelings of control and empowerment, our hedonic appreciation, and our efficient task performance. But like all technologies, both 3D displays and gesture-based systems come at a cost, and it is up to us, as human factors researchers, to establish the optimal design parameters and boundary conditions that ensure that 3D interfaces will enhance our interactions with an increasingly digital world.

References

Accot, J., & Zhai, S. (1997). Beyond Fitts' law: Models for trajectory-based HCI tasks. *Proceedings of Conference on Human Factors in Computing Systems* (pp.295-302), Atlanta, GA: CHI Publications.

Allison, R. S., Gillam, B. J., & Vecellio, E. (2009). Binocular depth discrimination and estimation beyond interaction space. *Journal of Vision, 9*, 1-14.

Ammelrooy, van P., (2012, 2 August). *Met 3D helm door je virtuele villa*. Amsterdam, the Netherlands: Volkskrant.

Ashbly, F. G., Isen, A. M., & Turken, A. U. (1999). A neuropsychological theory of positive affect and its influence on cognition. *Psychological Review, 106*, 529-550.

Baas, M., de Dreu, C. K. W., & Nijstad, B. A. (2008). A meta-analysis of 25 years of mood-creativity research: hedonic tone, activation, or regulatory focus? *Psychological Bulletin, 134,* 779-806.

Ballmer, S. (2010). A transforming trend - The natural user interface. The Huffington Post, January 12, 2010, from http://www.huffingtonpost.com/steve-ballmer/ces-2010-a-transforming-t_b_416598.html

Barfield, W., & Rosenberg, C. (1995). Judgement of azimuth and elevation as a function of monoscopic and binocular depth cues using a perspective display. *Human Factors, 37*, 173-181.

Becker, A. B., Warm, J.S., Dember, W. N., & Hancock, P. A. (1995). Effects of jet engine noise and performance feedback on perceived workload in a monitoring task. *The International Journal of Aviation Psychology, 5,* 49-62.

Beerends, J. G., & De Caluwe, F. E. (1999). The influence of video quality on perceived audio quality and vice versa. *Journal of Audio Engineering Society 47*, 355-362.

Berti, N., & Frassinetti, F. (2000). When far becomes near: Remapping of space by tool use. *Journal of Cognitive Neuroscience, 12*, 415-420.

Beurden, van M. H. P. H., IJsselsteijn, W. A., & Hopf, K. (2011). User centered design of gesture based interaction technology. *3DTV conference: The True Vision-Capture, Transmission and Display of 3D Video* (pp. 1-4). May 16-18, Antalya, Turkey.

Beurden, van M. H. P. H., IJsselsteijn, W. A., & Juola, J. F. (2012). Effectiveness of stereoscopic displays in medicine: A review. *3D Research, 3*, 1-13.

Beurden, van M. H. P. H., IJsselsteijn, W. A, & de Kort, Y. A. W. (2012). User experience of gesture based interfaces: A comparison with traditional interaction methods on pragmatic and hedonic quality. In: E. Efthimiou, G. Kouroupetroglou & S.-E. Fotinea (Eds.). *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication*. 9th International Gesture Workshop - GW 2011, Athens: Revised selected papers. LNCS/LNAI Vol. 7206, Springer

Beurden van M. H. P. H., IJsselsteijn W. A., & Kort de Y. A. W. (2011) Evaluating stereoscopic displays: Both efficiency measures and perceived workload sensitive to manipulations in binocular disparity. *Proceedings of SPIE-IS&T Electronic imaging,* 7863:786316 1 – 786316 7.

Beurden, van M. H. P. H., & IJsselsteijn, W. A. (2010). Range and variability in gesture-based interactions with medical images: Do non-stereo versus stereo visualizations elicit different types of gestures? *IEEE Virtual Reality: workshop on medical virtual environments*, Waltham, MA, USA.

Beurden van M. H. P. H., Kuijsters A., & IJsselsteijn, W. A. (2010). Performance of a path tracing task using stereo and motion based depth cues. *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop,* 176-181.

Bolt, R. (1980). "Put that there": Voice and gesture at the graphics interfaces. *SIGGRAPH Computer Graphics, 13*, 262-270.

Bonis, J. (2007). Acute Wiiitis. *The New England Journal of Medicine, 356*, 2431-2432

Borg, G. A. (1982). Psychophysical bases of perceived exertion. *Medicine and Science in Sports and Exercise, 14*, 377 – 381.

Bowman, D., Kruijff, E., LaViola, J., & Poupyrev, I. (2005). *3D User Interfaces: Theory and practice*. Boston, USA: Addison –Wesley.

Bowman, D. A., McMahan, R. P., & Ragan, E. D. (2012). Questioning naturalism in 3D user interfaces. *Communications of the ACM, 55*, 78-88.

Boxtel van, J. J. A., Wexler, M., & Droulez, J. (2003). Perception of plane orientation from self-generated and passively observed optic flow. *Journal of Vision, 3,* 318-332.

Brooke, J. (1996). SUS: a quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (eds). *Usability Evaluation in Industry, (pp. 189-194).* London, England: Taylor and Francis.

Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge Massachusetts, USA: Belknap Press.

Buxton, W. (1983). Lexical and pragmatic considerations of input structures. *Computer Graphics, 17*, 31-37.

Buxton, W. (2010). A touching story: A personal perspective on the history of touch interfaces past and future. *Society for Information Display (SID) Symposium Digest of Technical Papers, 41*, 444-448.

Buxton, B. (2012). Some milestones in computer input devices: An informal timeline. Draft version of April 4, 2012. Retrieved October 15, 2012, from: http://billbuxton.com/inputTimeline.html.

Buxton, W., & Myers, B. (1986). A study in two-handed input. *Proceedings of CHI '86 Conference on Human Factors in Computing systems* (pp. 321-326). New York:ACM.

Card, S. K., Mackinlay, J. D., & Robertson, G. G. (1990). The design space of input devices. *Proceedings of CHI'90 Conference on Human Factors in Computing systems,* (pp. 117-124). New York: ACM.

Casiez, G., Vogel, D., Balakrishnan, R., & Cockburn, A. (2008). The impact of control display gain on user performance in pointing tasks. *Human Computer Interaction, 23*, 215-250.

Cassell, J. (1998). A framework for gesture generation and interprestation. In R. Cipolla, & A., Pentland, (eds.), *Computer Vision in Human-Machine Interaction,* (pp. 191-215). New York: Cambridge University Press.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. *Proceedings of CHI '88 Conference on Human Factors in Computing systems*, (pp. 213-218). New York:ACM.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General, 140,* 102-116.

Clark, A. (2008). *Supersizing the mind: Reflections on embodiment, action, and cognitive extension.* Oxford, England: Oxford University Press.

Clark, A. (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence.* New York: Oxford University Press.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Routledge Academic.

Combe, E., & Wexler, M. (2010). Observer movement and size constancy. *Psychological Science, 21*, 667-675.

Cook, S. W., Yip, T. K., & Goldin-Meadow, S. (2011). Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes*, *26*, 1-17.

Corlett, E. N., & Bishop, R. P. (1976). A technique for measuring postural discomfort. *Ergonomics, 9*, 175-182.

Creem-Regehr, S. H., Willemsen, P., Gooch, A. A., & Thompson, W. B. (2003). The influence of restricted viewing conditions on egocentric distance perception: Implications for real and virtual environments. *Perception, 34*, 191-204.

Cumming, G. (2012). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis. New York: Routledge.

Cutting, J., & Vishton, P. (1995). Perceiving layout and knowing distances: The integration, relative potency and contextual use of different information about depth. In W. Epstein, & S. Rogers (Eds.), *Perception of Space and Motion* (pp. 69-117). San Diego, USA: Academic Press.

Davis, M. A. (2009). Understanding the relationship between mood and creativity: A meta-analysis. *Organizational Behavior and Human Decision Processes, 108*, 25-38.

De la Rosa, S., Moraglia, G., & Schneider, B. A. (2008). The magnitude of binocular disparity modulates search time for targets defined by a conjunction of depth and colour. *Canadian Journal of Experimental Psychology, 62*, 150-155.

Dinh, H. Q., Walker, N., Song, C., Kobayashi, A., & Hodges L. F. (1999). Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. *Proceedings of the IEEE Virtual Reality Los'99 conference*, 222-228.

Dourish, P. (2001). *Where the action is: Foundations of embodied interaction*. Cambridge, UK: MIT press.

Durgin, F. H., Baird, J. A., Greenburg, M., Russell, R., Shaugnessy, K., & Waymouth, S. (2009). Who is being deceived? The experimental demands of wearing a backpack. *Psychonomic Bulletin & Review, 16*, 964-969.

English, W. K., Engelbart, D. C., & Berman, M. L. (1967). Display-selection techniques for text manipulation. *Transactions on Human Factors in Electronics, 8*, 5-15.

Fairclough, S. H. (2009). Fundamentals of physiological computing. *Interacting with Computers, 21*, 133-145.

Faubert, J. (2001). Motion parallax, stereoscopy, and the perception of depth: practical and theoretical issues. In J., Bahram (Eds.), *Proceedings of SPIE CR76*, 168-191.

Fikkert, F. W. (2010). Gesture Interaction at Distance, Doctoral Dissertation, University of Twente, *Siks Dissertation* No. 2010-07.

Fishman, J. M., Ellis, S. R., Hasser, C. J., & Stern, J. D. (2008). Effect of reduced stereoscopic camera separation on ring placement with a surgical telerobot. *Surgical Endoscopy, 22*, 2396-2400.

Fitzmaurice, G. W., Ishii, H., & Buxton, B. (1995). Bricks: Laying the foundations for graspable user interfaces. *Proceedings of CHI'95 Conference on Human Factors in Computing systems* (pp. 442-449). New York: ACM.

Foley, J., Wallace, V., & Chan, P. (1984). The human factors of computing graphics interaction techniques. *IEEE Computer Graphics & Applications, 4*, 13-48.

Fovia. (2010). Stereoscopic images. Retreived January, 2010, from http://www.fovia.com

Friedrich, O. (1983). The machine of the year, the computer moves in. *Time, 121*, 14-24.

Fröhlich, B., & Plate, J. (2000). The Cubic Mouse: A new device for 3D input. *Proceedings of CHI'00 Conference on Human Factors in Computing systems* (pp.526-531). New York: ACM.

Gallagher, S. (2011). Interpretations of embodied cognition. In W. Tschacher & C. Bergomi (Eds.), *The implications of embodiment: Cognition and communication* (pp. 59-71). Exeter, England: Imprint Academic.

Getty D. J., & Green, P. J., (2007). Clinical applications for stereoscopic 3-D displays. Journal of the *Society for Information Display, 15*, 377-384.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Hillsdale, USA: Lawrence Erlbaum.

Goldin-Meadow, S. (2010). When gesture does and does not promote learning. *Language and Cognition, 2*, 1-19.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *American psychological science, 12*, 516-522

Graetzel, C., Fong, T. W., Grange, S., & Baur, C. (2004). A non-contact mouse for surgeon-computer interaction. *Technology Health Care, 12,* 245–57.

Haans, A., & IJsselsteijn, W. A. (2012). Embodiment and telepresence: Toward a comprehensive theoretical framework. *Interacting with Computers, 24*, 211-218.

Hamilton, A., Wolpert, D., & Frith, U. (2004). Your own action influences how you perceive another person's action. *Current Biology, 14, 493-498.*

Hancock, P. A. (1996). Effects of control order, augmented feedback, input device and practice on tracking performance and perceived workload. *Ergonomics, 39,* 1146-1162.

Hart, S .H., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam, The Netherlands: Elsevier.

Hartson, R., & Pyla, P. S. (2012). *The UX book, process and guidelines for ensuring a quality user experience.* Waltham, USA: Morgan Kaufmann.

Hassenzahl, M. (2004). The interplay of beauty, goodness and usability in interactive products, *Human-Computer Interaction, 19*, 319-349.

Hauptmann, A. (1989). Speech and gestures for graphic image manipulation. *Proceedings of CHI'89 Conference on Human Factors in Computing systems* (pp.241-245). New York: ACM.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). Multilevel and longitudinal modeling with IBM SPSS. New York, NY: Routledge.

Hendrix, C., & Barfield, W. (1995). Relationship between monocular and binocular depth cues for judgments of spatial information and spatial instrument design. *Displays, 16*, 103-113.

Hershenson, M. (1999). *Visual space perception: A primer.* Cambridge, England: The MIT Press.

Hincley, K., Pausch, R., Globe, J. C., & Kassell, N. F. (1994) Survey of design issues in spatial input. *Proceedings of the Seventh Annual ACM symposium on User Interface Software and Technology* (pp. 213-222). New York: ACM.

Hincley, K., Pausch,R., Proffitt, D., & Kassell, N. F. (1998). Two-handed virtual manipulation. ACM *Transactions on Computer-Human Interaction, 5*, 260-302.

Hinckley, K., & Wigdor, D. (2002). Input technologies and techniques. In A. Sears & J. A. Jacko (eds.), *The Human-Computer Interaction Handbook – Fundamentals, Evolving Technologies and Emerging Applications, Third Edition*. Londen, England: Taylor & Francis.

Hommel, B., Müssler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral & Brain Sciences, 24,* 849-937.

Hoogen, van den W. M., IJsselsteijn, W. A., & de Kort, Y. A. W. (2008). Exploring behavioral expressions of player experience in digital games. In A. Nijholt & R. Poppe (Eds.), *Proceedings of the workshop on Facial and Bodily Expression for Control and Adaptation of Games ECAG 2008,* (pp11-19). Amsterdam, the Netherlands.

Hopf, K., Neumann, F., & Przewozny, D. (2011). D7.5 Second generation gesture trackers. *Technical Report ICT-7-215280 HELIUM3D.*

Hornbaek, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies, 64*, 79-102.

Howard, I., & Rogers, B. J. (2002). *Seeing in depth: Depth perception vol. 1.* Toronto, Canada: I. Porteous.

Hu. H. H., Gooch, A. A., Creem-Regehr, S. H., & Thompson, W. B. (2002). Visual cues for perceiving distances from objects to Surfaces. *Presence: Teleoperators and Virtual Environments, 11*, 652-664.

Hubona, G. S., & Shirah, G. W. (2005). Spatial Cues in 3D visualization. *Ambient intelligence for Scientific Discovery. LNAI, 3345,* 104-128.

Hubona, G. S., Shirah, G. W., & Fout, D. G. (1997). The effects of motion and stereopsis on three-dimensional visualization. *International Journal of Human-Computer studies, 47*, 609-627.

Hubona, G. S., Shirah, G. W., & Jennings, D. K. (2004). The effects of cast shadows and stereopsis on performing computer-generated spatial tasks. *IEEE Transactions on Systems, Man and Cybernetics, 34*, 483-493.

Hutchison, J. J., & Loomis, J. M. (2006). Does energy expenditure affect the perception of egocentric distance? *Spanish Journal of Psychology, 9*, 332–339.

Hutchins, E. L., Hollan, J. D. & Norman, D. A. (1985). Direct manipulation interfaces. *Human-Computer Interaction, 1*, 311-338.

IJsselsteijn, W. A. (2003). Presence in the past: What can we learn from media history? In G., Riva F., Davide, & W. A., IJsselsteijn (eds.), *Being There - Concepts, Effects and Measurements of User Presence in Synthetic Environments,* (pp. 17-40). Amsterdam, the Netherlands: IOS Press.

IJsselsteijn, W. A. (2004). *Presence in Depth*. Dissertation, Eindhoven University of Technology, ISBN 90-386-2127-2.

Isen, A. M. (2001). An influence of positive affect on decision making in complex situation: Theoretical issues with practical implication. *Journal of Consumer Psychology, 11*, 76-85.

Ishii, H., & Ullmer, B. (1997). Tangible bits: Towards seamless interfaces between people, bits and atoms. *Proceedings of CHI'97 Conference on Human Factors in Computing systems* (pp.234-241). New York: ACM.

ISO 9241-11. (1997). *Ergonomic requirements for office work with display terminals (VDTs) part 11*: Guidance on Usability.

Jacob, R., & Sibert, L. (1992). The perceptual Structure of multidimensional input devices. *Proceedings of CHI'92 Conference on Human Factors in Computing systems* (pp.211-218). New York: ACM.

James, W. (1884). What is an emotion. *Mind, 9*, 188-205.

Johnsgard, T. (1994). Fitt's Law with a virtual reality glove and a mouse: effects of gain. *Proceedings of Graphics Interface - GI '94* (pp. 8-15). Toronto: Canadian Information.

Jordan, P. (2000), *Designing pleasurable products. An introduction to the new human factors.* London, England: Taylor & Francis.

Jordan, P. W. & Johnson, G. I. (1993). Exploring mental workload via TLX: The case of operating a car stereo whilst driving. In A.G., Gale, I.D., Brown, C.M., Haslegrave, H.W., Kruysse & S.P., Taylor (Eds.), *Vision in Vehicles-IV* (pp. 255-262). Amsterdam, the Netherlands: North- Holland.

Karam, M., & Schraefel, M. (2005). A taxonomy of gestures in Human Computer Interactions. *Technical report ECSTR-IAM05-009, electronics and computer science*, University of Southampton, 1-45.

Karapanos, E., Zimmerman, J., Forlizzi, J., & Martens, J. B. (2009). User experience over time: An initial framework. *Proceedings of the Chi'09 conference on Human factors in computing systems* (pp.729-738)New York: ACM.

Kipp, M. (2004) *Gesture generation by imitation: From human behavior to computer character animation*. Doctoral dissertation, Saarland University, Saarbruecken, Germany, Boca Raton, USA: Dissertation.

Kölsch, M., Beall, A. C., & Turk, M. (2003). The postural comfort zone for reaching Gestures. *In HFES Annual Meeting Notes,* Oktober 2003.

Kunnapas, T. M. (1955) Influence of frame size on apparent length of a line. *Journal of Experimental Psychology, 50*, 168-170.

Lambooij, M., IJsselsteijn, W. A., Bouwhuis, D., & Heynderickx, I. (2010). Evaluation of stereoscopic stills: Beyond 2D quality. *IEEE Transactions on Broadcasting, 57*, 432-444.

Lambooij, M., IJsselsteijn, W.A., Fortuin, M., & Heynderickx, I. (2009). Visual Discomfort and Visual Fatigue of Stereoscopic Displays: a Review. *Journal of Imaging Technology and Science, 53,* 1-14.

Loomis, J. M., & Knapp, J. M. (2003). Visual perception of egocentric distance in real and virtual worlds. In L. J. Hettinger & M. W. Haas (Eds.), *Virtual and Adaptive Environments* (pp.21-46). Mahwah, USA: Erlbaum.

Loomis, J. M., & Philbeck, J. W. (2008). Measuring spatial perception with spatial updating and action. In R. L. Klatzky, B. M. Whinney & M. Behrman (Eds.), *Embodiment, Ego-space and Action*. New York, USA: Taylor & Francis.

MacKenzie, I. S. (1992). Fitt's law as a research and design tool in Human-Computer Interaction. *Human-Computer-Interaction 7,* 91-139.

Malinverni, L., Lopez Silva, B., & Pares, N. (2012). Impact of embodied interaction on learning processes: Design and analysis of an educational application based on physical activity. *Proceedings of the 11th International Conference on Interaction Design and Children* (pp. 60-69). Bremen: ACM.

McGloin, R., Farrar, K. M., & Krcmar, M. (2011). The impact of controller naturalness on spatial presence, gamer enjoyment, and perceived realism in a tennis simulation video game. *Presence, 20*, 309-324.

McKee, S. P., & Taylor, D. G. (2010). The precision of binocular and monocular depth judgements in natural settings. *Journal of Vision, 10*, 1-13.

McMahan, R., Alon, A., Silva, M., Leal, A., Hagan, R., & Bowman, D. (2010). Evaluating natural interaction techniques in video games. *Proceedings of the IEEE symposium on 3D user interfaces*, 11-14.

McNeill, D. (2005). *Gesture and thought*. Chicago, USA: University of Chicago Press.

McWhorter, S., Hodges, L., & Rodriguez, M. (1991). Evaluating of display parameters affecting user performance of an interactive task in a virtual environment. *Technical Report GIT-GVU-91-31*, Atlanta: Georgia Institute of Technology.

Merritt, J. O. (1991). Evaluation of stereoscopic display benefits. In L. Hodges, D. McAllister, & J. Merritt (Eds.), *Introduction to Stereoscopic Displays and Applications, Short Course Notes*. Washington, DC: SPIE The International Society for Optical Engineering.

Miall, R. C., & Wolpert, D. M. (1996). Forward Models for Physiological Motor Control. *Neural Networks, 9, 1265-1279.*

Mohler, B. J., Creem-Regehr, S. H., Thompson, W. B., & Bülthoff, H. H. (2010). The effect of viewing a self-avatar on distance judgements in an HMD-based virtual environment. *Presence Teleoperators and Virtual Environments, 19,* 230-242.

Naepflin, U., & Menozzi, M. (2001). Can movement parallax compensate lacking stereopsis in spatial explorative search tasks*? Displays, 22*, 157-164.

Niedenthal, P. M., Brauer, M., Halberstadt, J. B., & Innes-Ker, A. H. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional experession. *Cognition and Emotion, 15,* 853-864.

Nielsen, J. (1993). *Usability Engineering.* Boston, USA: Academic press.

Nishikawa, A., Hosoi, T., Koara, K., Negoro, D., Hikita, A., Asano, S., Kakutani, H., & Miyazaki, F. (2003). FAce MOUSe: A novel human-machine interface for controlling the position of a laparoscope. *IEEE Trans. on Robotics and Automation, 19*, 825– 841.

Noë, A. (2004). *Action in Perception.* Cambridge, USA: The MIT press.

Norman, D. A. (2004*). Emotional design: Why we love (or hate) everyday things.* New York, USA: Basic Books.

O’Donnell, R. D., & Eggermeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. Kaufman, & J. Thomas (Eds.)*, Handbook of perception and human performance: Cognitive processes and performance* (pp 1-49). New York: USA: Wiley-Interscience.

O’Regan, J. K., & Noë, A. (2001). A sensorimotor approach to vision and visual consciousness. *Behavioral and Brain Sciences, 24,* 939-973.

Okoshi, T. (1980). Three dimensional displays. *Proceedings of the IEEE, 68*, 548-564.

Osiurak, F., Morgado, N., & Palluel-Germain, R. (2012). Tool use and perceived distance: When unreachable becomes spontaneously reachable. *Experimental Brain Research, 218*, 331-339.

Pastoor, S. (1997). 3-D displays: A review of current technologies, *Displays, 17(2),* 100-110.

Pastoor, S., (1993). Human factors of 3D displays in advanced image communications. *Displays, 14(3),* 150-157.

Pavlovic, V., Sharma, R., & Huang, T. (1997). Visual interpretation of hand gestures for human computer interaction: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 10*, 677-695.

Proffitt, D. R. (2006). Embodied perception and the economy of action. *Perspectives on* Psychological *Science, 1,* 110-122.

Proffitt, D. R., Stefanucci, J., Banton, T., & Epstein, W. (2003). The role of effort in perceiving distance. *Psychological Science*, 14, 106-112.

Proske, U. (2006). Kinesthesia: The role of muscle receptors. *Muscle Nerve, 34*, 545-558.

Quek, F. (2004). Gesture and Interaction. *Encyclopedia of Human-Computer Interaction, 1*, 288-292.

Quek, F., (1996). Unencumbered gestural interaction. *IEEE Multimedia, 3,* 36-47.

Quek, F., McNeill, D., Bryll, R., Duncan, S., Ma, X-F., Kirbas, C., McCullough, K. E., & Ansari, R. (2002). Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction, 9*, 171-193.

Richardson, A. R., & Waller, D. (2007). Interaction with an immersive virtual environment corrects users' distance estimates. *Human Factors, 49*, 507-517.

Riskind, J. H., & Gotay, C. C. (1982). Physical posture: Could it have regulatory or feedback effects on motivation and emotion? *Motivation and Emotion, 6*, 273–298.

Rogers, B., & Graham, M. (1982). Similarities between motion parallax and stereopsis in human depth perception. *Vision Research, 22*, 261-270.

Roscoe, S. N. (1984). Judgements of size and distance with imaging displays. *Human Factors, 26*, 617-629.

Rosenberg, L. B. (1993). The effect of interocular distance upon operator performance using stereoscopic displays to perform virtual depth tasks. *Virtual Reality Annual International Symposium*, Seattle: IEEE, 27-32.

Roufs, J. A. J., & Boschman, M. C. (1991). Visual comfort and performance. In J.A.J. Roufs, (eds.), *The man-machine interface, volume 15*(24-40). London, England: MacMillan Press.

Schneider, W., & Deubel, H. (2002). Selection-for-perception and selection-for-spatial-motor-action are coupled by visual attention: A review of recent findings and new evidence from stimulus-driven saccade control. In W. Prinz & B. Hommel (eds.), *Attention and performance XIX: Common mechanisms in perception and action, number 19 in Attention and performance* (pp. 609-627). Oxford, England: Oxford University Press.

Sedgwick, H. A. (1986). Space perception. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of Perception and Human Performance, vol. 1*. New York, USA: John Wiley and Sons.

Servos, P., Goodale M. A., & Jakobson L. S. (1992). The role of binocular vision in prehension: A kinematic analysis. *Vision Research, 32*, 1513-1521.

Sexton, I. & Surman, P. (1999). Stereoscopic and autostereoscopic display systems. *IEEE Signal Processing Magazine, 16*, 85-99.

Skalski, P., Tamborini, R., Shelton, A., Buncher, M., & Lindmark, P. (2011). Mapping the road to fun: Natural video game controllers, presence, and game enjoyment. *New Media & Society, 13*, 224-242.

Smith, D. C., Cole, R. E., Merritt, J. O., & Pepper, R. L. (1979). Remote operator performance comparing mono and stereo TV displays: The effects of visibility, learning and task factors. *Technical report Naval Ocean Systems Center (NOSC/TR-380).* San Diego, USA.

Sollenberger, R. L., & Milgram, P. (1993). Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. *Human Factors, 35*, 483-499.

Sturman, D., (1991). *Whole Hand Input*, Doctoral Dissertation, Massachusetts Institute of Technology.

Sutherland, I. (1996). Technology and courage. *Sun Microsystems Laboratories Technical report series.*(pp. 1-33). Burlington: Sun Microsystem Laboratories.

Thompson, W. B., Willemsen, P., Gooch, A. A., Creem-Regehr, S. H., Loomis, J. M., & Beall, A. C. (2004). Does the quality of the computer graphics matter when judging distances in visually immersive environments? *Presence: Teleoperators and Virtual Environments, 13,* 560-571.

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers, 13*, 127-145.

Wachs, J. P., Stern, H. I., Edan, Y., Gillam, M., Handler, J., Feied, C., & Smith, M. (2008). A gesture-based tool for sterile browsing of radiology images. *Journal of the American Medical Informatics Association, 15*, 321–323.

Wallach, H., & O'Connell, D. N. (1953). The kinectic depth effect. *Journal of Experimental Psychology, 45*, 205-217.

Waller, D. (1999) Factors affecting the perception of interobject distances in virtual environments. *Presence: Teleoperators and Virtual Environments, 8,* 657-670.

Ware, C. (2008). *Visual Thinking for design*. Burlington, MA: Morgan Kaufmann Publishers.

Ware, C., Hui, D., & Franck, G. (1993). Visualizing object oriented software in three dimensions. *Proceedings of the 1993 conference of the Centre for Advanced Studies on Collaborative research: software engineering* (pp. 612-620), Toronto, Canada.

Ware, C., & Mitchell, P. (2008). Visualizing graphs in three dimensions. *ACM Transactions on Applied Perception, 5,* 1-15.

Wartenberg, C., & Wiborg, P. (2003). Precision of exocentric distance judgements in desktop and cube presentations. *Presence: Teleoperators and Virtual Environments, 12,* 196-206.

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form.* Ames, Iowa: The University of Iowa.

Weiser, M. (1991). The computer of the 21st century. *Scientific American, 265,* 94-104.

Wexelblat, A. (1998). Research challenges in gesture: Open issues and unsolved problems. *Proceedings of the International Gesture Workshop on Gesture and Sign Language in Human-Computer Interaction, 1371,* 1-11.

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition, 68,* 77-94.

Wexler, M., & van Boxtel, J. J. A. (2005). Depth perception by the active observer. *Trends in Cognitive Sciences, 9,* 431-438.

Willemsen, P., Coltin, M. B., Creem-Rehehr, S. H., & Thompson, W. B. (2009). The effects of head-mounted display mechanical properties and field of view on distance judgements in virtual environments. *ACM Transactions on Applied Perception, 6*, 1-14.

Willemsen, P., Gooch, A. A., Thompson, W. B., & Creem-Rehehr, S. H. (2008). Effects of stereo viewing conditions on distance perception in virtual environments. *Presence: Teleoperators and Virtual Environments, 17*, 91-101.

Williges, R. C., & Wierwille, W. W. (1979). Behavioral measures of aircrew mental workload. *Human Factors, 21*, 549-574.

Winograd, T., & Flores, F. (1988). *Understanding computers and cognition: A new foundation for design*. Norwood, NY: Ablex.

Witmer, B. G., & Kline, P. B. (1998). Judging perceived and traversed distance in virtual environments. *Presence: Teleoperators and Virtual Environments, 7*, 144-167.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionaire. *Presence: Teleoperators and Virtual Environments*, *7,* 225-240.

Witt, J. K., Linkenauger, S. A., Bakdash, J. Z., & Proffitt, D. R. (2008). Putting to a bigger hole: Golf performance relates to perceived size. *Psychonomic Bulletin and Review, 15,* 581–585.

Witt, J. K., & Proffitt, D. R. (2005). See the ball, hit the ball: Apparent ball size is correlated with batting average. *Psychological Science, 16*, 937-938.

Witt, J. K., Proffitt, D. R., & Epstein, W. (2005). Tool use affects perceived distance but only when you intend to use it. *Journal of Experimental Psychology*: *Human Perception and Performance, 31*, 880-888.

Wobbrock, J., Morris, M., & Wilson, A. (2009). User defined gestures for surface computing. *Proceedings of CHI '09 Conference on Human factors in Computing Systems* (pp. 1083-1092). New York, NY: ACM.

Wohlschläger, A. (2000). Visual motion priming by invisible actions. *Vision Research, 40,* 925-930.

Yeh, Y. Y., & Silverstein, L. D. (1992). Spatial Judgements with Monoscopic and Stereoscopic Presentation of Perspective Displays. *Human Factors, 24,* 583-600.

Zhai, S. (1995). *Human Performance in six degree of freedom input control*. Unpublished doctoral dissertation, University of Toronto, Canada.

Zhai, S., & Milgram, P. (1993). Human performance evaluation of manipulation schemes in virtual environments. *Proceedings of the IEEE Virtual Reality Annual International Symposium* (pp. 155-161). Seattle, Washington, USA: IEEE Press.

Zwickel, J., Grosjan, M., & Prinz, W. (2010). What part of an action interferes with ongoing perception? *Acta Psychologica, 134, 403-409.*

Zwickel, J., & Prinz, W. (2012). Assimilation and contrast: the two sides of specific interference between action and perception. *Psychological Research, 76, 171-182*.

Summary

Interaction in Depth

Nowadays, a number of technical developments, such as sensor, processing, storage and communication technologies, enable a significant increase in the availability of data, and create the need to visualize these data in an intuitive fashion. The development of stereoscopic 3D displays offers a potential improvement for visualizing 3D spatial structures more intuitively, especially when combined with more natural and embodied means of interacting with 3D datasets. At present, new interaction technologies become available, that allow human-computer interfaces to move away from the constraints of the traditional keyboard/mouse combination, towards gesture-controlled interfaces, multitouch surfaces, face recognition, and voice recognition. These interfaces have in common that they broaden the bandwidth of human-computer interaction, engaging the body to a greater extent, and potentially affording more intuitive interactions than previously possible. In this thesis, the central focus is on the evaluation of interaction methods that are potentially useful in the context of stereoscopic 3D displays. We hypothesized that natural interaction with 3D displays - correctly mapped in spatial dimensions, and corresponding to previously learned skills - will likely enhance the effectiveness, efficiency, and overall experience of the interaction. Embodied interaction, in particular gesture-controlled interaction, may positively impact users' emotions and decrease cognitive load. In addition, embodied interaction may affect a user's perception of the environment, as people perceive an environment not only in terms of its behaviorally independent visual properties, but also in terms of their ability to act in it. The central research question in this thesis is to investigate how 3D interaction maps onto 3D spaces, and to what extent interaction can optimize performance and user-experience, or influence the very nature of perception and understanding of the digital world.

Chapter 1 provides an introduction to the thesis, providing an overview of 3D display technologies, as well as human-computer interaction technologies, in particular interaction in 3D, with the human as embodied social actor.

In Chapter 2, we extended current 3D display evaluation methods by applying the concept of perceived workload, in addition to completion times and accuracy, to better understand the benefits of stereoscopic visualization in performance-oriented contexts. Results showed that stereo either yielded an advantage or resulted in no difference in performance, yet stereo never decreased performance. We further learned that the effectiveness of stereo depends on both the disparity level employed and the difficulty

level of the task. The use of motion, i.e., both object motion and movement parallax, was most effective in increasing accuracy, while, especially for the difficult tasks, a combination between motion and stereo was most effective in decreasing completion times. Results in terms of cognitive workload revealed mixed results, which might be due to the disparity levels used in these experiments.

In Chapter 3, we provided a user-centered assessment of movement-based interactions. First, we determined the range and variability of the repertoire of gestures used in 3D interaction. Subsequently, we performed experimental evaluations of the pragmatic and hedonic qualities of gesture-controlled interactions and compared this with more traditional device-based interactions. Results showed that, whereas embodied interactions were preferred in terms of hedonic qualities, they appear to be outperformed by traditional interfaces in terms of pragmatic aspects of the interaction.

Chapter 4 combines and extends the work of Chapters 2 and 3, by replicating the main experimental findings, utilizing a performance-oriented task, and including additional outcome measures. Results largely confirm and validate our previous findings, demonstrating that stereo is an effective cue in decreasing completion times, whereas motion is most important depth cue for increasing accuracy. In addition to Chapter 2, the results showed that stereo can decrease cognitive workload, as long as the right level of disparity is selected. Moreover, hedonic quality, fun and affect not only increased when interacting with gestures, but also while performing a task in 3D.

In Chapter 5, we demonstrated that embodied interaction not only affects user experience, but also impacts processes of perception at a more fundamental level. In this chapter we manipulated the amplitude of users' hand and arm movements, by changing the gain (ratio between the size of our hand movement and the movement of the cursor on the screen) of the interaction, and subsequently obtaining estimations of the distance between two objects. Results showed that the amplitude of body movements affect distance estimates, irrespective of any 'objective' visual representation. Specifically, larger hand movements elicited larger estimates for the same distances than smaller hand movements. This result suggests that people may incorporate physical (motor control) information while interpreting distances in the environment.

The findings of this thesis contribute to discussions on research methodology in relation to 3D interactions and 3D displays, and provide new insights in the effects of 3D interaction on visual perception. Furthermore, results in this thesis formulate implications for the design of 3D displays and 3D interaction technologies that are aimed to accurately display mediated environments and/or enhance users' experience and performance within such environments.

Samenvatting

Interactie in Diepte

De huidige technologische ontwikkelingen (zoals, sensortechnologie, dataververwerking, dataopslag en communicatietechnologie), zorgen voor een sterke toename in de hoeveelheid beschikbare data. Hierdoor ontstaat de noodzaak om deze data op een intuïtieve manier te visualiseren. De ontwikkeling van stereoscopische (3D) schermen biedt een potentiële verbetering om bijvoorbeeld 3D structuren intuïtiever te visualiseren, zeker in combinatie met meer natuurlijke en 'embodied' (het betrekken van het lichaam bij de interactie) manieren van interacteren met 3D-datasets. Momenteel komen er nieuwe interactiestijlen beschikbaar die het mogelijk maken, om naast traditionele mens-computer-interfaces zoals het toetsenbord en de muis, ook interfaces te gebruiken gebaseerd op gebaren, aanraakschermen, gezichtsherkenning en spraakherkenning. De overeenkomst tussen deze interactiestijlen is dat het de bandbreedte van mens-computer interactie vergroot door ook het lichaam te betrekken bij de interactie, waardoor interactie intuïtiever kan worden dan voorheen. In dit proefschrift ligt de nadruk op het evalueren van interactiestijlen die geschikt kunnen zijn voor het interacteren met 3D-schermen. Interactie met 3D schermen is natuurlijk als onze bewegingen kloppen met de dimensies van de 3D dataset en als deze overeenkomen met eerder geleerde vaardigheden. Dit alles kan de effectiviteit, efficiency en de algehele ervaring van de interactie vergroten. Daarnaast kan 'embodied interaction', met in het bijzonder gebaargestuurde interactie, een positieve invloed hebben op de emoties van gebruikers en wellicht cognitieve (werk)last verlagen. Verder kan 'embodied interaction' ook perceptie van de omgeving beïnvloeden. Hoe we de omgeving waarnemen wordt namelijk niet alleen beïnvloed door interpretatie van visuele informatie, maar ook door onze bewegingen en ons vermogen om in de omgeving te handelen. Daarom kan 'embodied interaction' wellicht ook onze waarneming van afstanden in een 3D omgeving beïnvloeden, onafhankelijk van de gepresenteerde visuele informatie. In dit proefschrift staat de vraag centraal hoe 3D interactie het beste aansluit bij 3D omgevingen en in hoeverre interactie de taakprestatie en gebruikerservaring kan optimaliseren en in welke mate het onze perceptie van onze omgeving beïnvloedt.

Hoofdstuk 1 vormt de inleiding op het proefschrift, met daarin overzichten van 3D scherm technologieën en mens-computer interactie, in het bijzonder 3D-interactie, met de mens als 'embodied' sociale actor.

In hoofdstuk 2 onderzoeken we of stereoscopische visualisaties taakprestatie kunnen verbeteren, waarbij we naast nauwkeurigheid en snelheid ook hebben gekeken naar het concept cognitieve werklast. De resultaten lieten zien dat stereo soms wel tot prestatieverbetering leidde en soms niet, maar we zagen nooit een verslechtering van de prestatie. Verder lieten de resultaten zien dat de effectiviteit van stereo afhankelijk was van zowel het diepteniveau als de moeilijkheidsgraad van de taak. Het gebruik van bewegingsparallax bleek het meest effectief voor het verhogen van de nauwkeurigheid van de taak, terwijl een combinatie van bewegingsparallax en stereo vooral in moeilijkere taken resulteerde in een snellere taakprestatie. Resultaten in termen van cognitieve werklast lieten een gemixt beeld zien, waarbij we denken dat dit veroorzaakt wordt door de aangeboden diepteniveaus.

In hoofdstuk 3 bespreken we hoe gebruikers de interactie beoordelen als de interactie is gebaseerd op lichaamsbewegingen, zoals het interacteren met handgebaren. Als eerste hebben we bepaald welke handgebaren mensen maken wanneer ze interacteren met een 3D object. Daarna hebben we interactie met handgebaren vergeleken met meer traditionele interactiestijlen, waarbij we gekeken hebben naar pragmatische en hedonische kwaliteiten van de interactie. De resultaten lieten zien dat interacties waarbij je je lichaam gebruikt hoger scoorden in termen van hedonistische kwaliteiten, maar dat traditionele interactiemethoden beter scoorden in termen van pragmatische aspecten.

In hoofdstuk 4 richten we ons op de replicatie van de belangrijkste experimentele bevindingen van hoofdstuk 2 en 3, alleen nu met een prestatiegerichte taak en met aanvullende vragen over de subjectieve ervaring van proefpersonen. De resultaten bevestigden en valideerden grotendeels eerdere bevindingen, waaruit blijkt dat stereo een effectieve cue is voor het verhogen van de taaksnelheid, terwijl bewegingsparallax de nauwkeurigheid verhoogt. Als aanvulling op hoofdstuk 2 vonden we dat stereo cognitieve werklast kan verlagen, mits het diepteniveau optimaal gekozen is. Daarnaast vonden we dat hedonistische kwaliteiten niet alleen hoger waren als we interacteerde met handgebaren, maar ook wanneer de taak is uitgevoerd in 3D.

In hoofdstuk 5 laten we zien dat 'embodied interaction' niet alleen invloed heeft op de ervaring van gebruikers, maar ook op onze waarneming. In dit onderzoek manipuleerden we de grootte van de handbewegingen door de 'gain' (verhouding tussen de armbeweging en de beweging van de cursor op het scherm) van de interactie te manipuleren. Vervolgens lieten we mensen de afstand tussen twee objecten schatten. In de resultaten zagen we dat de geschatte afstand beïnvloed werd door de grootte van de handbewegingen. Bij dezelfde afstanden werd bij grotere handbewegingen de afstand groter ingeschat dan bij kleinere handbewegingen. Dit resultaat illustreert dat mensen fysieke informatie gebruiken bij het interpreteren van afstanden in hun omgeving.

De inzichten verkregen in dit proefschrift dragen bij aan de discussies over onderzoeksmethodologie in relatie tot 3D interactie en 3D displays en ontwikkelt nieuwe inzichten in de effecten van 3D interactie op onze visuele waarneming. Verder geven de uitkomsten van dit proefschrift richtlijnen voor het ontwerpen van 3D-schermen en 3D interactie technologieën, die zijn gericht op het accuraat weergeven van gemedieerde omgevingen, en hoe de gebruikerservaring en prestaties binnen dergelijke omgevingen verbeterd kunnen worden.

Acknowledgements

Writing this last piece of my thesis allows me to think back on the period I have worked on the University. First of all, I had the privilege to work on two European projects, developing state of the art technology, where it was my task to investigate the human factor in these projects. I really enjoyed the international collaboration with different European institutes and companies. However, it was also a challenge to define research designs which provided input for the technical developments of the 3D displays, and contributed to academic research which could serve as input for my Ph.D. As a result, the work described in this thesis investigated more applied as well as fundamental research questions. Many people contributed to my thesis, which I would like to thank.

First, I would like to thank my first promotor and daily supervisor Wijnand IJsselsteijn, for his valuable comments and positive attitude. I enjoyed our, sometimes, philosophical discussion on a variety of topics in this thesis, as well as our collaboration in the two European projects. In addition I would like to thank my co-promotor Yvonne de Kort. I appreciated her fresh view and valuable comments on my thesis work, and her support given in the final writing phase of my Ph.D. I also want to thank Jim Juola, my second promotor for his support throughout the years, and valuable comments that improved my thesis work. I'm also grateful to my committee members Dirk Heylen, Jean-Bernard Martens and Huib de Ridder for their valuable feedback on the draft version of this thesis

In addition, I would like to thank all the project members of the MUTED and HELIUM3D projects, for our good collaboration and inspiring project meetings. Especially I would like to thank Gert van Hoey, Klaus Hopf, Haralambos Hatzakis, and Michiel Visser, with whom I worked together on various experiments described in this thesis.

In addition I want to thank Martin Boschman and Aart van der Spank, who always were of great help to optimize the lab settings and facilities I needed for my, sometimes, challenging experimental set-ups. I would also like to thank all my colleagues and former colleagues at the Human Technology Interaction department for creating the pleasant working atmosphere in the last few years. I would especially thank my former roommates Marc Lambooij and Daniel Lakens, for their great support during my thesis.

I would also like to thank my family and friends, for giving me mental and physical relaxation aside of my thesis work. I also want to thank my two sisters, Ingrid and Jessica, who always have been there for me. I would especially thank Karin, for her support and patience and love in times I needed it the most, thanks! Last but not least, I like to express my deepest gratitude and love to my parents Mary and Bert, for their unconditional support and trust in me, whatever choices I made. Therefore pap en mam, I will dedicate this book to you.

Maurice van Beurden

Curriculum Vitae

Maurice H.P.H van Beurden was born on 19 December 1980 in Middelburg, The Netherlands. After obtaining his HAVO diploma in 1998 at the Nassau-scholengemeenschap in Breda, he started Electrical Engineering at the Hogeschool Brabant in Breda in 1998. He changed studies in 2000 to Applied Physics at the Fontys Hogescholen in Eindhoven. During his bachelor project in 2003, he worked on an algorithm to localize epileptic activities in the brain. From 2003 to 2007 he completed the master Human Technology Interaction at the faculty Industrial Engineering and Innovation sciences at the Eindhoven University of Technology. During his master thesis, he worked on the perception of flicker in dynamic lighting applications. From 2007, he started his PhD project at Eindhoven University of Technology, and participated in two European projects (FP6-MUTED and FP7-HELIUM3D), studying to what extend 3D interaction and 3D displays can optimize task performance and user experience. In the current thesis, he also studied the role of embodied interaction on the perception and understanding of the digital world. Since 2012, he is employed as research scientist in the Training and Performance Innovation group at TNO in Soesterberg, focusing on human factor aspects in complex and dynamic working environments.