Max- Planck- Institut
für biologische Kybernetik
Spemannstraße 38 • 72076 Tübingen • Germany

Technical Report No. 77

# Vision and Action in Virtual Environments: Modern Psychophysics in Spatial Cognition Research

Heinrich H. Bülthoff and Hendrik A.H.C. van Veen[†],

December 1999

This paper will also appear (in a revised form) as a chapter in a Springer-Verlag book edited by Laurence R. Harris and Michael Jenkin and provisionally entitled *Vision and Attention*. This book originates from the successful *York Vision and Attention Conference* held in June 1999 in Toronto, Canada.

[†] To whom all correspondence should be addressed. Hendrik-Jan van Veen is now at: TNO Human Factors Research Institute, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands; e-mail: vanVeen@tm.tno.nl.

This report is available via anonymous ftp at ftp://ftp.kyb.tuebingen.mpg.de/pub/mpi-memos/pdf/TR-077.pdf in PDF-Format or at ftp://ftp.kyb.tuebingen.mpg.de/pub/mpi-memos/TR-077.ps.Z in compressed PostScript-format. The complete series of Technical Reports is documented at http://www.kyb.tuebingen.mpg.de/techr/

# Vision and Action in Virtual Environments: Modern Psychophysics in Spatial Cognition Research

*Heinrich H. Bülthoff & Hendrik A. H. C. van Veen*

**Abstract.** The classical psychophysical approach to human perception has been to study isolated aspects of perception using well-controlled and strongly simplified laboratory stimuli. This so-called cue reduction technique has successfully led to the identification of numerous perceptual mechanisms, and has in many cases guided the uncoverage of neural correlates (see chapters elsewhere in this volume). Its limitations, however, lie in the almost complete ignorance of the intimate relationship between action, perception and the environment in which we live. Real world situations are so different from the stimuli used in classical psychophysics and the context in which they are presented, that applying laboratory results to daily life situations often becomes impractical if not impossible. At the Max-Planck-Institute for Biological Cybernetics in Tübingen we pursue a behavioral approach to human action and perception that proves especially well suited for studying more complex cognitive functions, such as object recognition and spatial cognition. The recent availability of high-fidelity "Virtual Reality" environments enables us to provide subjects a level of sensory realism and dynamic sensory feedback that approaches their experiences in the real world. At the same time, we can keep the ultimate control over all stimulus aspects that are required by the rules of psychophysics. In this chapter, we take a closer look at these developments in spatial cognition research and present results from several different experimental studies that we have conducted using this approach.

*Keywords*: Virtual Reality; Virtual Environments; Human Behavior; Perception; Recognition; Navigation; Spatial Cognition; Psychophysics; Biological Cybernetics

## 1 Introduction

In recent years, the study of spatial cognition experienced a strong technology push caused by major advancements in two very different areas: brain imaging and virtual reality. Indeed, those who manage to combine the potential of both technologies receive considerable attention (e.g., Maguire, Frith, Burgess, Donnett & O'Keefe, 1998; Maguire, Burgess, Donnett, Frackowiak, Frith & O'Keefe, 1998; Epstein & Kanwisher, 1998). In this chapter, however, we focus exclusively on the role of virtual reality technology in the recent evolution of the field. In subsequent sections, we identify the major motivations behind this development, and we provide illustrative examples taken from our own laboratory. Our primary goal here is to provide useful information for making proper and effective use of virtual reality in cognitive science.

So what is exactly happening? What do we mean when we say "virtual reality technology is pushing spatial cognition research ahead?" The answer probably lies in the very nature of virtual reality (VR): VR is a technique that strives to create the illusion of experiencing a physical environment without actually being there (the concept of verisimilitude has been mentioned in this context). The virtual environments (VEs) thus created provide the researcher with a new experimental platform in addition to the natural environment and the classical highly reduced and abstracted laboratory settings. We will see further on in this chapter why VEs have earned a place alongside those other options, and why their role in studying human spatial cognition is strongly increasing (also see Péruch & Gaunet, 1998, and Darken, Allard & Achille, 1998). Some of the major factors are illustrated in the following paragraph.

The classical psychophysical methods that are used to investigate perception are characterized by the use of well-controlled but, compared to the real world, strongly simplified laboratory stimuli such as dots, plaid patterns or random dot stereograms. These abstracted stimuli often bear little resemblance to those occurring in the real world, but are nevertheless very useful for identifying low-level perceptual mechanisms. The study of higher-level cognitive behaviors such as object recognition, visual scene analysis, and navigation, requires a different methodology. At this level the intimate and extensive relationship between action, perception, and the environment plays an important role (Gibson, 1966 & 1979). To unravel the mechanisms working at this level, it is less important to understand perception itself than it is to investigate its role in guiding actions. Moreover, it is questionable whether one can study such higher-level mechanisms using the abstracted stimuli that are typically applied in psychophysics. In navigation, for instance, we repeatedly use landmarks to decide where we are and where we should go next. It is obvious that such a behavior strongly depends on the abundance and specific appearance of these landmarks. Thus, a systematic study of cognitive behaviors like navigation ideally uses a methodology that supports the control and manipulation of arbitrary complex stimuli in a reproducible way, at the same time allowing for the recording of natural behavioral responses to these stimuli. The technology that enables us to develop such a methodology has become available only recently. Advancements in computer graphics and display technology have led to the emergence of a new area of computer science and engineering, called VR. As said before, VR is essentially a technique that creates the illusion of experiencing a physical environment without actually being there. This is accomplished by intercepting the normal action-perception loop: the participant's actions are measured and used to update a computer representation of a virtual environment, which is then presented to the participant by means of visual and other displays (haptic, tactile, auditory, etc.). This technique allows us principally to manipulate all aspects of the sensory stimulation as well as the effects of the participant's actions on these sensory experiences. As such, it enables us to study fundamental questions in human cognition.

In the following sections we describe several motivations behind the increased usage of VEs in spatial cognition research. In the first section we discuss some issues from a biological cybernetics point of view. Subsequent sections deal with the technology that enables us to use VEs, discuss stimulus control in the context of VEs, point out the increased level of stimulus relevance that VEs offer compared to traditional laboratory methods, identify spatial cognition in VEs as a new interesting field and give examples from our own VE laboratory.

## 2 Biological Cybernetics

Biological cybernetics is a subfield of biology that studies the complete cycle of action and perception in organisms. More specifically, it studies how organisms acquire sensory information, how they process and store it, and how they finally retrieve this information again to generate behavior. Such behavior – like moving through the world – on its turn alters the sensory information available to the organism and in doing so closes the action-perception loop. Systematic research of this complex feedback system requires fine control over those elements of the action-perception cycle that lie outside the organism, i.e., the world with which the organism interacts. Intercepting the feedback loop by manipulating the 'world'-parameters alters the way in which action can influence perception. Such open- and modified closed-loop experiments are common practice in neuroethology (Reichardt, 1973; Heisenberg & Wolf, 1984) and sensorimotor studies (Hengstenberg, 1993). Virtual reality techniques are now for the first time enabling us to perform similar experiments in the domain of complex human behavior, such as navigation through unknown cities (Mallot, Gillner, Van Veen & Bülthoff, 1998) or manipulating virtual objects with a haptic simulator (Ernst, Van Veen, Goodale & Bülthoff, 1998; Ernst, Banks & Bülthoff, 1999).

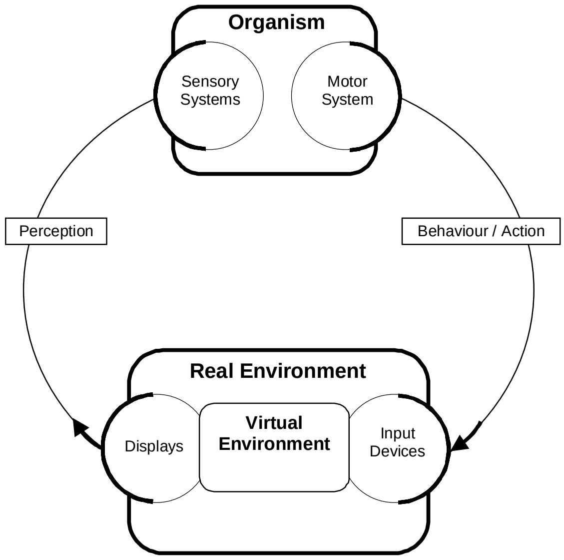


***Figure 1.*** *Information processing loop. The perception/action loop between sensory and motor systems is intercepted by a virtual environment that is encapsulated into the real environment. Adapted from Figure 9 in Distler, Van Veen, Braun & Bülthoff 1998.*

Figure 1 shows a basic diagram of the action-perception cycle from a biological cybernetics perspective. The diagram symbolizes the flow of information between organism and environment. It is strongly simplified to make it easier to concentrate on the important elements for this discussion. For instance, the homeostatic processes that take place within an organism are not included, nor do we attempt to differentiate between types of behavior that do or do not induce changes in the environment. Please refer to Bülthoff, Foese-Mallot & Mallot (1997) and references therein for more details. In this diagram we illustrate how we think VEs can be utilized to study the action-perception cycle. Inside the natural environment (the "real" one, if you wish) a second environment is created by means of human-computer interfaces. A smaller or larger part of the user's behavior is monitored by input devices such as movement trackers and is used to update a computer representation of the organism plus its virtual environment. Displays such as HMDs (Helmet/Head Mounted Displays) and earphones are used to communicate this representation to the user. Three important observations can be made by looking at this diagram. First, it is immediately clear that VEs are not substitutes for the real environment, they are merely embedded in it. Thus, we have to deal with a person experiencing two environments at the same time. Second, it is currently by no means possible – indeed it's hard to imagine being possible at all (but read Gibson, 1984) – to completely measure all behavior, to create a complete virtual world, and to stimulate all senses completely. A VE is always a reduced environment. Third, the devices interfacing the organism with the VE inherently suffer from delays, distortions, bandwidth restrictions, and limited ranges, which causes them to be distinguishable from the real environment. We discuss these observations in more detail below.

## 2.1 Two Parallel Worlds

VEs have been applied successfully for treatment of certain phobias (for an overview see Glantz, Durlach, Barnett & Aviles, 1996 and 1997), such as fear of spiders (Carlin, Hoffmann, & Weghorst, 1997), fear of height (Rothbaum, Hodges, Kooper,


***Figure 2.*** *Snapshot of Virtual Tübingen. The 3D reconstruction and rendering of a typical narrow street of historical Tübingen demonstrates the fidelity of our VR model which is achieved with few polygons but high resolution texture maps for each individual house (there are not two houses in the 700 house model of Tübingen which are the same!).*

Opdyke, Williford, & North, 1995), and fear of flying (Mühlberger, Herrmann, Wiedemann, & Pauli, 1999). In the latter case, participants interact with a VE that simulates different stages of flying. Without ever leaving the ground, and with the participants fully aware of this and of the fact that everything is just a simulation, even a relatively simple VE can be convincing enough to induce fear and generate changes in physiological parameters such as heart rate, skin conductance, and EEG. Sometimes quite the contrary happens, and in a VE that has been designed for optimal visual quality participants do not feel immersed at all, but rather start commenting on artifacts of the simulation such as that "all trees in the landscape look so similar". And sometimes it seems as if participants can mentally switch from one world to the other and back, or even can observe both worlds in parallel! The central question here is: How much of the participants perception and behavior is related to each of the worlds? What happens when the worlds provide conflicting information (such as in the aforementioned fear of flying treatment example)? Simple linear weighting models seem inappropriate here. In our eyes, the majority of the psychological and philosophical questions related to this concept of two parallel environments are yet unexplored. For some, this has been enough reason not to use VEs for spatial cognition research. A good way to see how much we can trust results obtained using VEs is to pair studies in VEs with studies in the natural environment. If the results gained in both environments are consistent with each other, further experiments can be performed in the VE taking advantage of its advanced features (see the section on Stimulus Control). We have done so, for example, in an experiment that studies mental representations of familiar environments. Inhabitants of the city of Tübingen in southern Germany were asked to point as accurately as possible towards well-known locations in their inner city, both while being present in the real city and while experiencing a very detailed virtual reality simulation of that same inner city (for details see Sellen, 1998, and Van Veen, Sellen & Bülthoff, 1998). Subjects responded very accurately in both cases: the mean absolute pointing error was 11˚ when the subject was present in the real city, and increased marginally to 13˚ when experiencing the virtual version of that city. Further analysis showed that the pattern of systematic errors was extremely similar in the two conditions, suggesting that similar mental representations were recalled in both cases. Further experiments exploiting the simultaneous presence of a real and virtual version of this city environment are underway, as well as experiments in which we make changes to the virtual city that are not possible with the real one.

### 2.2 Incompleteness

Much can be said about the incompleteness of a VE in comparison to the real environment. If we focus on the direct implications for spatial cognition research using VEs the most severe problem is probably the pitfall of superficial realism. A VE might look realistic enough for one's purposes, but still can lack certain qualities that turn out to be essential for other tasks. For instance, after going through great lengths to create a realistic (mainly visually realistic) virtual model of the city of Tübingen (see Van Veen, Distler, Braun & Bülthoff, 1998; also see figure 2), at least one subject in our experiments (an inhabitant of real Tübingen) complained about the lack of appropriate height differences between the streets. She used to find her way around the town by remembering how certain roads sloped upwards and others downwards, something none of the other subjects seemed to do. Obviously, this is information of which researchers can make good use (work on the role of height differences in navigation is now underway, see Mochnatzki, Steck & Mallot, 1999), but the potential danger is also clear. The system of validation elicited in the previous paragraph is again essential. Note, of course, that there are many other obvious forms of incompleteness with which we also have to deal, such as the lack of stimulation of certain senses (typically only visual simulations are used in VR), the incompleteness of the stimulation (e.g., limited field-of-view), the simplicity of the environment, and all the problems associated with ego-movement in VEs. Some of these points are discussed again in the section *Enabling Technologies.*

### 2.3 Delays and Distortions

An ideal interface between the participant and the VE should operate unnoticeably. If not, it's likely that participants start changing their behavior to circumvent the problems of the interface. Such change in behavior has of course implications for the validity of the experimental study. A typical problem is the feedback delay caused by the processing time required to reflect changes of the participant's behavior in changes on the displays. In vehicle simulators, for example, participants often compensate for feedback delays by reducing the speed of the vehicle (very slow speeds can mitigate the impact of feedback delays; see Cunningham &

Tsou, 1999) and by employing alternate control strategies (Sheridan & Ferrel, 1963). Short delays are essential for studies involving fast control loops such as those found in steering tasks, manual manipulation, or head tracking. Distortions are especially evident and disturbing when parts of the real and virtual worlds interact, such as when the participant tries to grab a virtual object with his real hand, or when head movements are measured to update the images displayed on the HMD. While humans can adapt to delays and distortions (for reviews of spatial adaptation, see Bedford, 1993; Harris, 1965, 1980; Welch, 1978; for temporal adaptation, see Cunningham, Billock & Tsou, 2000), this ability is limited (Bedford, 1999).

Two interesting concepts that are largely intertwined with the discussion above about the problems of the parallel worlds, incompleteness, and interfacing, are presence and immersion. Slater and Wilbur (1997) define presence as "*a state of consciousness, the (psychological) sense of being in the virtual environment.*", and immersion as "*a description of a technology, the extent to which the computer displays are capable of delivering an inclusive, extensive, surrounding, and vivid illusion of reality to the senses of a human participant.*" Their distinction between technology-related aspects and consciousness-related ones seems quite useful for better understanding why certain VEs are more effective than others.

## 3 Enabling Technologies

Given the way contemporary VEs are created, we can distinguish three different types of technologies: those that measure human behavior, those that support building virtual models, and those that display these environments to the user. We do not want to discuss these technologies here in full detail, but some key elements are worth mentioning, because they have helped to revolutionize our research.

### 3.1 Measuring Behavior

***Figure 3.*** *VRbike in front of the large screen projection screen. The panoramic image of virtual Tübingen is projected by three ceiling mounted CRT projectors in such a way that at the head position of the cyclist a realistic 180 degree view of Tübingen can be experienced while cycling through the model*

.

The most interesting class of measuring devices with respect to spatial cognition research is the equipment that tracks the participant's movements through the real world. VEs are usually simulated within the confinement of a real room, and thus any type of ego-movement of the participant in the VE must be mapped onto movements within the boundary of that room. Often the participant can not move in the real world at all, because he has to remain seated in front of a monitor or projection screen. Recent advancements in movement tracking technology now allow for accurate real-time measurements of translation and rotation of the head, trunk, and hand within room-sized enclosures. In combination with an HMD, the participant can move about in a virtual world by actually walking through a real space (e.g., see Chance, Gaunet, Beall & Loomis, 1998, and Usoh, Arthur, Whitton, Bastos, Steed, Slater and Brooks, 1999). The limited size of the real room remains a restrictive factor of course. For studies involving larger virtual spaces different solutions are applied. In our laboratory we use a specially configured exercise bicycle originally distributed by Tectrix™ and Cybergear™ (VRbike; see figure 3, and Distler, 1996) to move through large-scale virtual worlds like cities and forests (see Distler, Van Veen, Braun, Heinz, Franz & Bülthoff, 1998). The participant needs to pedal and steer, the bicycle provides appropriate pedaling resistance and tilts in curves, but the whole configuration itself does not physically translate. We are therefore able to use this bicycle in front of a large panoramic projection screen. Similar solutions involve treadmills (e.g., see Darken, Cockayne & Carmein, 1997) and car-like interfaces.

### 3.2 Building Models

The requirements of other areas such as the military and game industries have led to the development of high quality software and hardware for rapidly creating and rendering complex virtual environments. In our laboratory we make use of a very powerful graphical supercomputer (Onyx2™ InfiniteReality™, manufactured by Silicon Graphics™) to reach a high level of visual realism. Note, however, that much cheaper PC-based systems are now also reaching performance levels that seem sufficient for many VE-studies of spatial cognition. At the software level, modeling tools such as 3D Studio Max™ and Multigen™ (which we use for many projects) offer tremendous capabilities for designing virtual worlds.

### 3.3 Display Systems

Several different types of visual displays are in common use now. Simple monitors are used less and less due to the limited field-of-view that they provide and the restrictions on the participant's movements. In recent years most of the technical developments have been focused on creating high-quality HMDs and panoramic projection systems. Truly panoramic systems are very expensive but can provide very high levels of immersion by covering the whole visual field with computer-controlled imagery. HMDs are much cheaper and allow the subject to move around quite a bit more. Proper head tracking without delays is still extremely difficult though, and in practice HMDs often give disappointing results. HMDs do not cover the whole visual field with computer-generated images. Instead, their design effectively blocks sight of the real world in all directions and combine that with a small segment where the display is located. Other display types worth mentioning in relation to spatial cognition are auditory systems (for high-fidelity 3D spatial audio rendering), haptic and tactile feedback systems (for providing contact cues with virtual objects), and motion platforms. These latter systems come in many varieties and are used to simulate physical movement of the observer, mainly by combining a little bit of real motion with a lot of transient motion cues (sudden onsets and offsets of motion, acceleration cues, etc.). The basic sensory systems that these devices stimulate are the proprioceptive and vestibular senses. We have recently installed a virtual reality system incorporating such a motion platform (manufactured by MotionBase™) in our laboratory in Tübingen, and it is currently being used for research on spatial updating and scene recognition. It will also be used to validate and extend the

***Figure 4.*** *Birds-eye view of a small city with a hexagonal street raster. This artificial city (Hexatown) surrounded by global landmarks served in several experiments to study the importance of local and global landmarks in human wayfinding.*

research on driving behavior that was done in our lab (e.g., Chatziastros, Wallis & Bülthoff, 1997, 1998, and Wallis, Chatziastros & Bülthoff, 1997). Note that the VRbike mentioned above functions both as a measuring device (through its steering and pedaling sensors) and as a display (through its computer-controlled pedaling resistance and its tilting motion).

A lot of work is going on to improve all these technologies at many different levels. More immersive displays, more realistic environments, and more powerful motion trackers are under development and this will certainly improve the applicability of VEs for cognitive science.

# 4 Stimulus Control

Conducting experiments in VEs means that someone has to program or define the complete content of the environment. Everything that is in there has explicitly been put there. This ensures that a precise description of the stimulus can be reported, allowing anyone to repeat or reproduce the experiment in order to validate the results. This is certainly not always possible with experiments in real environments. The major difference between conducting experiments in real and in virtual environments, however, is that in the latter case one has in principle complete control over the environment. This has several substantial advantages:

- all subjects can participate under exactly the same conditions
- the environment is optimally designed for the experiment
- no uncontrolled external factors in the environment (traffic, weather) can disturb the experiment
- any parameter of the experiment can be varied systematically, even during the experiment
- one can switch from environment A to environment B in a split-second
- changes to the environment can be made at any time

We would like to demonstrate the power of extreme stimulus control by briefly summarizing a few experiments conducted in our laboratory.

## 4.1 Wayfinding & Dynamic City Layouts

In a series of experiments, Mallot and colleagues investigated the mental representation of spatial knowledge of structured large-scale environments (see Gillner & Mallot, 1998; Steck & Mallot, in press; Mallot, Gillner, Van Veen, & Bülthoff, 1998). Using a specially created artificial virtual city called Hexatown (named so because of its hexagonal street raster, which forces a left-right movement decision at every junction; see figure 4), they tried to unravel the building blocks of mental spatial representation. To do so, subjects first learned certain routes through Hexatown until they could repeat them flawlessly. In the subsequent testing phase, subjects were instantly put at locations somewhere along the route and were then asked to start completing those routes. Between training and testing phase, however, modifications to the city plan were made in such a way that different mental representations would correspond to different route completions. For instance, in Mallot & Gillner (1999) some of the buildings were moved to different locations. This way the researchers were able to conclude that the learned routes were stored in a graph-like representation of local elements, and not in globally consistent survey map type of representation. Certainly no one outside Hollywood would consider conducting such experiments in the real world.

## 4.2 Visual Homing in Virtual Worlds

Homing can be defined as the act of finding one's way back to a starting point after an excursion through the environment. Communication with other organisms set aside, homing can be achieved by applying a combination of two basic mechanisms. In the "environment-centered" approach, the organism navigates by combining current position information extracted from the local environment with its spatial long-term memory. In the "organism-centered" approach, the organism uses sensory information about its self-motion through the environment to continuously update its position relative to a starting point. Riecke and collaborators (Riecke, 1999; Van Veen, Riecke & Bülthoff, 1999) studied whether this latter mechanism, usually called path-integration, works properly and effectively when only visual information is present. They conducted triangle completion experiments in high-fidelity vision-only virtual environments. On each trial subjects had to return to their starting point after moving outwards along two prescribed segments using the mouse buttons. Environment-centered strategies were precluded by replacing all landmarks in the scene by others during a brief dark interval just before the subjects started the return path. The results indicated that subjects acquired a fairly accurate mental representation of the triangular paths by just optical information alone. Omitting the scene modifications before the return movement resulted in nearly perfect performance, stressing the dominant role of environment-centered mechanisms under more natural conditions. Experiments like this one are obviously extremely difficult to set up in the real world but can be done rather elegantly using VEs.

### 4.3 Scene Perception & Dynamic Scene Content

The process by which we recognize and analyze scenes remains largely mysterious. What evidence we do have, suggests that the instantaneous, full, and detailed perception of a scene which we experience, is simply illusory and that detailed analysis of objects can only be achieved in a more piecewise, serial manner. In recent years a phenomenon called change blindness has been used to estimate the accurateness of the representation of static scenes. Change blindness is the failure to detect a change in a scene, usually because the transient of the change is masked in some way (more can be found in other chapters in this volume). Wallis and Bülthoff (2000) have conducted an experiment in Tübingen in which they extended the change blindness paradigm to dynamic scenes. A person drives or is being driven along a virtual road. At regular intervals the screen blanks for a very short period during which a change to the scene near the road is being made. Their results show that change blindness also occurs in dynamic scenes. In particular they show that especially changes in object location are hard to detect when the subject moves through the environment. Although others have managed to do related experiments in the real world (see Levin & Simons, 1997, and Simons & Levin, 1998, and the chapter by Simons in this volume), the level of systematic control available when using VEs is incomparable.

### 4.4 View-based Scene Recognition

Gibson (1979) showed us a long time ago the importance of the moving observer in a natural environment but this importance extends to encoding and recognition of scenes also. If an observer knows where he is and in what direction he is looking, then by actively moving around he could build a coherent spatial representation of the immediate environment. The computer vision community has adopted the benefits of an active observer under the *'active-vision'* paradigm, which is nicely illustrated in the book by Blake and Yuille (1992). Of course, psychologists know the importance of ego-motion and interactivity already for a long time under the framework of perception for action. In a series of experiments Christou & Bülthoff (1999) investigated how we represent our immediate environment. Specifically they asked the question: If we learn to recognize a room from a limited set of directions will we recognize it also from novel views? In the experiments, participants explored a virtual attic of a house (see figure 5) by using a 6 degree-of freedom interface (Spacetec IMC Co., Massachusetts, USA) to drive a simulated camera through the environment. In the familiarization phase participants had to find and acknowledge small encoded markers in the room that only appeared when viewed close enough. The movement through the room was restricted along one major axis of the room and the viewing direction was restricted to the left or right by 60 degrees. Since the participants were only allowed to "walk" back and forth and could not turn around, they could never see the room from the other direction. After finding all the markers, each participant was shown pictures of the locations of each of the markers together with images from the other direction, which they never saw before. An equal number of distracter images taken from a similar 3D 'distracter' environment were also shown to participants. They simply had to respond when they believed the current image was taken from the original environment they had traversed during the familiarization stage. The results showed that after extensive, controlled, and yet realistic learning in a virtual environment the restrictions imposed on the content of perceptual experience are still reflected in recognition performance. The familiar views were easily recognized while the performance dropped significantly for the novel views. Performance also dropped considerably when the active familiarization phase described above was exchanged for passively watching a sequence of snapshots of the attic. Especially the ability to recognize the novel direction views deteriorated. This was not the case for the back-seat driver condition, in which the active familiarization phase was replaced by passively watching a pre-recorded movie of another subject performing the active condition. In summary, Christou and Bülthoff have shown that active vision improves recognition performance. The back-seat driver condition shows that observer ego-motion is the critical variable, not volitional movement. What they have not shown is what a more natural locomotion could provide us with. It is quite conceivable that recognition performance improves much more if observers are

***Figure 5.*** *Virtual attic. Experiments with active and passive exploration of this VR model helped us to understand the importance of the active observer in view-based scene recognition.*

totally immersed in the virtual environment by either walking or cycling through it.

Stimulus control has been the key to the success of psychophysical studies of the past century. We hope to have shown above that virtual reality techniques now allow us to greatly extend the range of problems that can be studied with this approach.

## 5 Stimulus Relevance

One of the hidden benefits of using VEs for spatial cognition research is the increased level of stimulus relevance. The classical reductionist's approach is to remove all stimulus components that are not directly relevant for the study being conducted. A single aspect of perception or behavior is singled out and studied in great detail, and all the other sensory inputs are kept to a minimum. Of course we are all very much aware of the usefulness of this scientific method. Problems emerge, however, when we try to integrate the knowledge of all these isolated aspects to understand perception and behavior in natural environments. Non-linear and dynamic interactions, a priori expectations (Bayesian vision!), inter-individual differences, new levels of stimulus complexity, and highly dynamic scenes, are only a few of the factors that often make such integration processes hopelessly complicated if not impossible. At the same time, one can ask whether the results obtained using isolated stimuli have any relevance at all for perception and behavior under natural conditions. Without claiming to have found a general solution to this problem, we would like to put forward the following consideration. In terms of perturbation theory, the classical reductionist's approach involves the systematic variation of one or a few stimulus parameters around certain control values, keeping all other parameters constant. The level at which all these other parameters are kept is often best described by 'zero'. However, perturbing the stimulus around 'zero' is not a very ecologically interesting condition. Moreover, the reduced level of sensory stimulation might cause undesired changes in behavior that go unnoticed. A much more relevant approach, at least in terms of understanding perception and behavior in natural environments, would be to set all non-varied stimulus aspects equal to a level typical of the natural environment. That obviously poses a stimulus control problem, because the number of parameters that would need to be considered is unimaginatively large. However, what we gain with such an approach is that we can assume that the perturbations in which we are interested are studied in a realistic context. In essence, we have greatly improved stimulus relevance. It is obvious that we want to conclude this consideration with expressing our belief that VEs can reach a level of sensory realism that is good enough to support such an approach. Whether or not this modern psychophysical method can live up to the promise of increased stimulus relevance remains to be proven, but for us it seems to be the only way out so far.

## 6 Spatial Cognition in VEs

An interesting and recent development that spatial cognition researchers could employ, is the increased usage of VEs in different domains. Some people spend major parts of their working time in VEs, which gives spatial cognition in VEs a whole new meaning. Not only can we apply VEs for the study of spatial cognition, we can study the spatial cognition of humans living in VEs. The problem itself is not completely new. For several decennia simulators have been used to train driving and flying skills of military personnel, and gradually this approach has been transferred to the civilian domain. Obviously, the question of transfer of training from VEs to practical situations is related to the problem of validation that studies of spatial cognition using VEs have to face. Advancements in technology and thinking are now creating new questions. What happens to the spatial mental representation of people confronted with temporally or spatially discontinuous VEs, created for instance by using hyperlinks (see Ruddle, 1999, and Ruddle, Howes, Payne & Jones, 1999)? To what extent can we keep track of rapidly changing spatial scenes, such as those that can emerge when the historical development of a city area is (virtually) played back at high speed? What are the implications for spatial information processing when the real world is augmented by overlaid spatial information generated from synchronous virtual models? Studying such and other unusual situations enabled by the new technologies might provide us with surprisingly new insights about the organization of our spatial memory and capabilities, especially with respect to plasticity and adaptability.

## 7 Concluding Remarks

We would like to point out here that we understand that VEs are not always the best way to go. Maximizing stimulus control is probably best achieved by removing all unnecessary cues from the stimulus, e.g., the classical reductionist's approach. Maximum stimulus relevance is of course only available in the natural environment. We hope to have made clear, however, that using VEs means combining the best of both, and opens up many new and exciting possibilities.

The introduction of VEs in spatial cognition research is along the same lines as the introduction of the graylevel raster display and the later extensive usage of computer graphics in perception and recognition research. The increasing availability and dropping costs of the technology will soon make these tools accessible to virtually anyone. We expect that within the next couple of years the usage of VEs for studying spatial cognition will become common practice in many labs. The promise of increased stimulus control and relevance and the emergence of exciting new questions will certainly motivate many researchers to do so. Applying VEs will drive an integration process across disciplines: perception, behavior and the (virtual) environment will reunite again.

In this light it might be worthwhile to briefly discuss the guest editorial called "Virtual Psychophysics" that recently appeared in the journal "Perception" (Koenderink, 1999). In his editorial Koenderink first shows his excitement about the new possibilities that computers and virtual worlds seem to offer. He mentions several factors that are also highlighted in the current paper, such as increased stimulus control and stimulus relevance (which he considers enormously important), and he adds to that the benefit of being able to quickly produce all kinds of stimuli that "*..would have been completely out of the scope of the old-day optical setups.*" But then he turns extremely skeptical and expresses his fear that most if not all of the modern psychophysical studies that use **virtual** worlds will turn out to be **virtual** psychophysics in a couple of decennia. His main argument is that the visual realism of contemporary virtual environments is deceiving and almost nobody realizes that. He is really worried about that "*.. present authors take familiarity with their virtual world pretty much for granted.*", or in other words, that no one seems to care about a comprehensive description of the stimulus they use. We think that this view is way too skeptical. Of course, the apparent realism of contemporary VEs is only a trick, a trick that is getting better every year. The patterns of light and dark (the example used by Koenderink) shown on our displays are not the same as those encountered in the real world, even though they look pretty realistic to the untrained eye. But how important is that? And does not every researcher know that? Sure enough, for those of us who study how the distribution of light and dark in a scene conveys to us information about the detailed spatial relationships between scene elements, an extensive knowledge of the physical laws of optics and materials is essential. Indeed, for some of the problems in this specific area there is no piece of software that simulates the necessary level of physics. But we believe that a trained researcher will recognize such a situation, and will refrain from using computer graphics in such a case. Similarly, those of us who study completely different problems, such as wayfinding, will judge the differences between the light patterns found in the virtual and real worlds as not or only marginally relevant to the task they are studying. In fact, they argue in much the same way as the reductionists argue when they remove every bit of stimulation that is not directly relevant to the task at hand, the only difference being the control point around which they conduct their perturbation studies. We believe that researchers are smart enough to realize that the computer graphics and virtual worlds they use are not the same as the real world. The patterns of light and dark are different, and so are the level of stimulus complexity and the level of sensory complexity and the naturalness of movement and etcetera etcetera.

VE-based studies will be able to survive the test of time when we pay attention to two rules. First, those of us who want to generalize their results beyond the specific virtual world used for the experiment (which would otherwise indeed be nothing more than a study of spatial cognition in that particular VE), need to find ways to validate their results. This can be done by comparing the results with other studies, thus building up a framework of mutually supporting results, or by running similar experiments in the real world, which provides a framework itself. Second, and we support Koenderink in this, it is extremely important that scientists using VEs write down in their papers as complete as possible either how the particular virtual world (the stimulus!) has been created and displayed, or, alternatively, how it differs from the real world. With the expected developments in computer graphics in mind, this latter option might become more and more popular in the decades to come.

## Acknowledgements

The authors would like to thank Stephan Braun for his help in preparing this chapter, Douglas Cunningham, Rainer Rothkegel and Sibylle Steck for useful comments on previous versions of the manuscript, and Hartwig Distler for many stimulating discussions in the past which have helped to shape the insights presented in this chapter. While in Tübingen, Hendrik-Jan van Veen was funded by the Max-Planck Society and by the Deutsche Forschungsgemeinschaft (MA 1038/6-1, 1038/7-1).

## References

**Blake**, A. & Yuille, A. L. (1992) *Active Vision*, MIT Press, Cambridge, MA.

**Bedford**, F. L. (1993) “Perceptual learning”, The Psychology of Learning and Motivation, 30: 1-60.

**Bedford**, F. L. (1999) “Keeping perception accurate”, Trends in Cognitive Science 3 (1): 4-11.

**Bülthoff**, H. H., Foese-Mallot, B. M. & Mallot, H. A. (1997) “Virtuelle Realität als Methode der modernen Hirnforschung (translation: Virtual reality as a method for modern brain research)”, in H. Krapp & T. Wägenbauer (Eds.), Künstliche Paradiese – Virtuelle Realitäten, Wilhelm Fink Verlag, München: 241-260.

**Carlin**, A. S., Hoffmann, H. G. & Weghorst, S. (1997) “Virtual reality and tactile augmentation in the treatment of spider phobia: a case report”, Behavior Research and Therapy, 35: 153-158.

**Chance**, S. S., Gaunet, F., Beall, A. C. & Loomis, J. M. (1998) “Locomotion mode affects the updating of objects encountered during travel: The contribution of vestibular and proprioceptive inputs to path integration”, Presence: Teleoperators and Virtual Environments, 7 (2): 168-178.

**Chatziastros**, A., Wallis, G. M. & Bülthoff, H. H. (1997) “The effect of field of view and surface texture on driver steering performance (Utiliser un environnement virtuel pour évaluer indicateurs qui affectent la performance du conducteur)”, *Proceedings of Vision in Vehicles VII, Sept. 1997, Marseille, France*. [To appear in 1999 as: A. G. Gale, (Ed.), I. D Brown., C. M. Haslegrave & S. P. Taylor (co-eds.), Vision in Vehicles VII, Elsevier Science B.V., Amsterdam, North-Holland].

**Chatziastros**, A., Wallis, G. M. & Bülthoff, H. H. (1998) “Lane changing without visual feedback?”, Perception, 27 (supplement): 59.

**Christou**, C. G. & Bülthoff, H. H. (1999) “View dependence in scene recognition after active and passive learning”, Memory and Cognition, in press.

**Cunningham**, D. W. & Tsou, B.H. (1999) “Sensorimotor adaptation to temporally displaced feedback”, Investigative Ophthalmology and Visual Science, 40 (4): 585.

**Cunningham**, D. W., Billock, V. A. & Tsou, B. H. (2000) "Sensorimotor adaptation to violations of temporal contiguity and the perception of causality", manuscript submitted for publication.

**Darken**, R. P., Cockayne, W. R. & Carmein, D. (1997) "The omni-directional treadmill: a locomotion device for virtual worlds", Proceedings of UIST '97, October 14-17, 1997, Banff, Canada: 213-221.

**Darken**, R. P., Allard, T. & Achille, L. B. (1998) "Spatial orientation and wayfinding in large-scale virtual spaces: An introduction", Presence: Teleoperators and Virtual Environments, 7 (2): 101-107.

**Distler**, H. (1996) "Psychophysical experiments in virtual environments", in *Virtual Reality World'96 Conference Documentation*, Computerwoche Verlag, München.

**Distler**, H. K., Van Veen, H. A. H. C., Braun, S. J. & Bülthoff, H. H. (1998) "Untersuchung komplexer Wahrnehmungs- und Verhaltensleistungen des Menschen in virtuellen Welten (translation: The investigation of complex human perception and behavior in virtual worlds)", in I. Rügge, B. Robben, E. Hornecker & F.W. Bruns (Eds.), *Arbeiten und Begreifen: Neue Mensch-Maschine-Schnittstellen*, Lit Verlag, Münster: 159-172.

**Distler**, H. K., Van Veen, H. A. H. C., Braun, S. J., Heinz, W., Franz, M. O. & Bülthoff, H. H. (1998) "Navigation in real and virtual environments: Judging orientation and distance in a large-scale landscape", in M. Göbel, J. Landauer, M. Wapler & U. Lang (Eds.), *Virtual Environments '98: Proceedings of the Eurographics Workshop in Stuttgart, Germany, June 16-18, 1998*, Springer Verlag, Wien.

**Epstein**, R. & Kanwisher, N. (1998) "A cortical representation of the local visual environment", Nature, 392: 598-601.

**Ernst**, M. O., Banks, M. S. & Bülthoff, H. H. (1999) "Touch can change visual slant perception", manuscript submitted for publication.

**Ernst**, M. O., Van Veen, H. A. H. C., Goodale, M. A. & Bülthoff, H. H. (1998) "Grasping with conflicting visual and haptic information", Investigative Ophthalmology & Visual Science, 39 (4): 624

**Gibson**, J. J. (1966) *The senses considered as perceptual systems*, Houghton Mifflin, Boston.

**Gibson**, J. J. (1979) *The ecological approach to visual perception*, Houghton Mifflin, Boston.

**Gibson**, W. (1984) *Neuromancer*, Victor Gollancz Ltd, Great Britain.

**Gillner**, S. & Mallot, H. A. (1998) "Navigation and acquisition of spatial knowledge in a virtual maze", Journal of Cognitive Neuroscience, 10: 445-463.

**Glantz**, K., Durlach, N. I., Barnett, R. C., & Aviles, W. A. (1996) "Virtual Reality (VR) for psychotherapy: From the physical to the social environment", Psychotherapy, 33: 464-473.

**Glantz**, K., Durlach, N. I., Barnett, R. C. & Aviles, W. A. (1997) "Virtual reality (VR) and psychotherapy: Opportunities and Challenges", Presence: Teleoperators and Virtual Environments, 6: 87-105.

**Harris**, C. S. (1965) "Perceptual adaptation to inverted, reversed, and displaced vision", Psychological Review, 72: 419-444.

**Harris**, C. S. (1980) "Insight or out of sight? Two examples of perceptual plasticity in the human adult", in C. S. Harris (Ed.), Visual Coding and Adaptability, Lawrence Erlbaum, Hillsdale, NJ: 95-149.

**Heisenberg**, M. & Wolf, R. (1984) *Vision in Drosophila*, Springer Verlag, Berlin.

**Hengstenberg**, R. (1993) "Multisensory control in insect oculomotor systems", in F. A. Miles & J. Wallman (Eds.), *Visual Motion and its Role in the Stabilization of Gaze*, Elsevier Science Publishers.

**Koenderink**, J. J. (1999) "Virtual psychophysics", Guest editorial, Perception, 28: 669-674.

**Levin**, D. T. & Simons, D. J. (1997) "Failure to detect changes to attended objects in motion pictures", Psychonomic Bulletin and Review, 4 (4): 501-506.

**Maguire**, E. A., Burgess, N., Donnett, J. G., Frackowiak, R. S. J., Frith, C. D. & O'Keefe, J. (1998) "Knowing where and getting there: a human navigation network", Science, 280: 921-924.

**Maguire**, E. A., Frith, C. D., Burgess, N., Donnett, J. G. & O'Keefe, J. (1998) "Knowing where things are: Parahippocampal involvement in encoding object locations in virtual large-scale space", Journal of Cognitive Neuroscience, 10: 61-76.

**Mallot**, H. A., Gillner, S., Van Veen, H. A. H. C. & Bülthoff, H. H. (1998) "Behavioral experiments in spatial cognition using virtual reality", in C. Freksa, C. Habel & K. F. Wender (Eds.), Spatial Cognition: An interdisciplinary approach to representing and processing spatial knowledge, Lecture Notes in Artificial Intelligence 1404, Springer Verlag, Berlin.

**Mallot**, H. A. & Gillner, S. (1999) "View-based vs. place-based navigation: What is recognized in recognition-triggered responses?", manuscript submitted for publication.

**Mochnatzki**, H. F., Steck, S. D. & Mallot, H. A. (1999) "Geographic slant as a source of information in maze navigation", in N. Elsner & U. Eysel (Eds.), *Göttingen Neurobiology Report 1999, Volume II*, G. Thieme Verlag, Stuttgart: abstract No. 875.

**Mühlberger**, A., Herrmann, M., Wiedemann, G., & Pauli, P. (1999) "Treatment of fear of flying by exposure in virtual reality", manuscript submitted for publication.

**Péruch**, P. & Gaunet, F. (1998) "Virtual environments as a promising tool for investigating human spatial cognition", Current Psychology of Cognition, 17: 881-899.

**Reichardt**, W. (1973) „Musterinduzierte Flugorientierung Verhaltensversuche an der Fliege *Musca domestica*", Naturwiss. 60: 122-138.

**Riecke**, B. (1998) *Untersuchung des menschlichen Navigationsverhalten anhand von Heimfindeexperimenten in virtuellen Umgebungen (translation: Studying human navigation behavior by performing homing experiments in virtual environments)*, Masters Thesis, Physics Department of the Eberhard-Karls-Universität Tübingen, Germany.

**Rothbaum**, B. O., Hodges, L. F., Kooper, R., Opdyke, D., Williford, J. S., & North, M. (1995) "Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia", American Journal of Psychiatry, 152: 626-628.

**Ruddle**, R. A. (1999) "The problem of arriving in one place and finding that you're somewhere else", *Proceedings of the workshop on Spatial Cognition in Real and Virtual Environments, April 27-28, 1999, Tübingen, Germany* : 58.

**Ruddle**, R. A., Howes, A., Payne, S. J. & Jones, D. M. (1999) “The effects of hyperlinks on navigation in virtual environments”, manuscript submitted for publication.

**Sellen**, K. (1998) *Schätzen von Richtungen in realen und virtuellen Umgebungen (translation: Estimation of Directions in Real and Virtual Environments),* Masters Thesis, Biology Department of the Eberhard-Karls-Universität Tübingen, Germany.

**Sheridan**, T. B. & Ferrel, W. R., (1963) “Remote manipulative control with transmission delay”, Perceptual and Motors Skills 20: 1070-1072.

**Simons**, D. J. & Levin, D. T. (1998) “Failure to detect changes to people in real-world interaction”, Psychonomic Bulletin and Review, 5 (4): 644-649.

**Slater**, M. & Wilbur, S. (1997) “A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments”, Presence: Teleoperators & Virtual Environments, 6: 603-616.

**Steck**, S. D. & Mallot, H. A. (1999) “The role of global and local landmarks in virtual environment navigation”, accepted for publication in Presence: Teleoperator & Virtual Environments.

**Usoh**, M., Arthur, K., Whitton, M. C., Bastos, R., Steed, A., Slater, M. & Brooks, F. P. Jr. (1999) “Walking > walking-in-place > flying, in virtual environments”, in *SIGGRAPH 99 conference proceedings,* Annual Conference Series, ACM SIGGRAPH August ‘99, Addison Wesley: 359-364.

**Van Veen**, H. A. H. C., Distler, H. K., Braun, S. J. & Bülthoff, H. H. (1998) “Navigating through a virtual city: Using virtual reality technology to study human action and perception”, Future Generation Computer Systems, 14: 231-242.

**Van Veen**, H. A. H. C., Sellen, K. & Bülthoff, H. H. (1998) “Pointing to invisible landmarks in real and virtual environments”, Investigative Ophthalmology & Visual Science, 39 (4): 625.

**Van Veen**, H. A. H. C., Riecke, B. E. & Bülthoff, H. H. (1999) “Visual homing to a virtual home”, Investigative Ophthalmology & Visual Science, 40 (4): 798.

**Wallis**, G. M., Chatziastros, A. & Bülthoff, H. H. (1997) “Even experienced drivers have the wrong concept about how to change lanes”, Perception, 26 (supplement): 100.

**Wallis**, G. & Bülthoff, H. H. (2000) “What’s scene and not seen: Influences of movement and task upon what we see”, to appear in Visual Cognition, 7.

**Welch**, R. B. (1978) *Perceptual Modification: Adapting to altered sensory environments*, New York: Academic Press.