Chapter 1
Introduction

Having thought about visual perception for five years, I have become more and more impressed by the power and ability of the visual system. It is amazing that this system is able to abstract so much information that is useful for the observer from an enormous amount of available data. In some respects the visual process is similar to the memory process: the basis of remembering is forgetting everything that is not useful and the basis of visual processing is perceiving important aspects of the environment by ignoring the rest. But how does the system work? How does it combine different features into an object? How does it integrate changes over time? And an important question for this thesis: how does the brain derive depth from two-dimensional projections on the retinae? My introduction is twofold: first of all I would like the reader to share my admiration for the processes involved in visual perception, and secondly I would like to explain why the work I have been doing is important.

I will begin by giving a short introduction to the human visual system. After looking at reference frames and depth cues, I will look at the traditional visual space research and go on to discuss the newer paradigms that have been proposed. Finally I will introduce my own work that is described in detail in the following chapters of this thesis.

1.1 The Visual System

The visual system consists of sensory organs (the eyes) and brain-structures that process the signals that have entered the eyes. However, visual processing already begins in the retinae. I have no intention of discussing every aspect of the visual system here, since the purpose of this introduction is merely to introduce the reader to the basics of the system.

Light entering the eyes

Light enters the eye through the lens. The shape of the lens causes the beam of light to bend. This bending of the light rays enables the eye to focus on objects at different distances from the observer. If an object is close to the observer, the ciliary muscle contracts to make the lens spherical. The light-rays are therefore bent to a considerable degree. If an object is far away, the lens has to be flat in order to have the object in focus. This happens when the ciliary muscles relax, leading to less bending of the light-rays.

The light hits the retina after entering the eye. The outer layer of the retina contains the photoreceptors, the rods and cones, which contain pigments that absorb photons. Bipolar cells transmit the signal from the photoreceptors to the ganglion cells. These ganglion cells have long axons that leave the eye at the blind spot. The blind spot got its name because this area of the retina has no photoreceptors and thus gives no output to the visual system. In the bipolar layer, there are neurons that contribute to the communication between adjacent bipolar cells. The initial processing of visual information (lateral inhibition) takes place in these cells. The light has to enter through this layer of cells to reach the photoreceptors.

The fovea is the spot on the retina on to which the part of the visual field that is in focus is projected. This area of the retina contains a large amount of cones. The cells that are used for further processing are not in front of the photoreceptors here. This is why we have a particularly sharp image of the part of the visual field that is projected on to the fovea. Figure 1.1 gives a schematic picture of the eye.

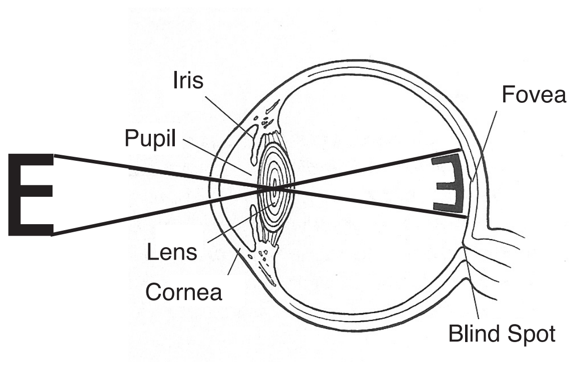


Figure 1.1 *Horizontal cross-section of the eye*

Visual processing

We can distinguish multiple types of ganglion cells. The most widely known ones are the M and P cells. The M cells have a larger cell-body and also have a larger receptive field. These cells are distributed fairly evenly throughout the retina. The P cells, on the other hand, are smaller and have small receptive fields. They are located in and close to the fovea. The axons of the ganglion cells together form the optic nerve as they leave the eye through the blind spot. The optic nerves of both eyes cross in the optic chiasm. Here the ganglion cells from the nasal areas of the retinae cross to the other side of the brain (see Figure 1.2). This causes the information from the left visual field to be projected on to the right side of the brain and vice versa. From the optic chiasm the optic tract passes through the lateral geniculate nucleus. From this point we can speak of two separate visual pathways: the magnocellular pathway and the parvocellular pathway; these are projected mainly by the M and P ganglion cells, respectively. The parvocellular visual pathway is responsible for color and form perception and the recognition of objects. It transmits detailed information that can only be derived from objects that are in focus, due to the presence of P ganglion cells primarily near the fovea. The magnocellular pathway is concerned with the perception of depth and movement, and derives information mainly from the M ganglion cells. These two pathways are not as strictly separated as is often assumed, as they interact at many stages of processing.

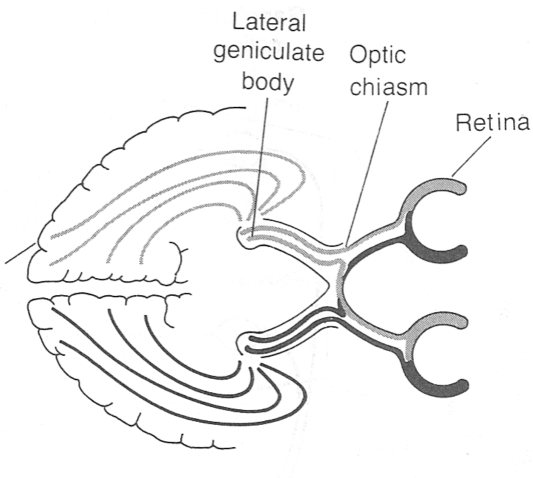


Figure 1.2
Horizontal cross-section of the brain showing the visual pathway from eyes to visual cortex

The visual pathways connect the geniculate nucleus to the primary visual cortex (V1) of the occipital lobe and then lead on to higher visual areas like the secondary visual cortex (V2). Some processing of depth-information takes place in V1, however spatial processing takes place primarily in the parietal lobe of the cortex, which is part of the magnocellular or dorsal visual pathway. There has been much discussion about whether this dorsal pathway is concerned mainly with visual control of action (Goodale, Milner, Jakobson, & Catey, 1991) or with a more general spatial representation of the world. However, according to Creem and Proffitt (2001), the discussion could be redundant if one thinks in terms of reference frames.

They suggest that in the parietal lobe different areas are involved in different reference frames. Since visual control of action and spatial representation of the world are based on different reference frames, this might explain the two theories of parietal lobe function (Creem, & Proffitt, 2001). However, others have suggested that the dorsal pathway is mainly concerned with egocentric representations (Neggers, van der Lubbe, Ramsey, & Postma, In press). Several authors suggest that it is the ventral pathway (the parvocellular pathway) rather than part of the dorsal pathway that is involved in the formation of allocentric representations.

1.2 Reference frames

The concept of reference frames is important in visual space research, and indeed for perception in general. By ‘reference frame’ we mean the locus (or set of loci) with respect to which the spatial position of an object is defined (Pick, & Lockman, 1981). Reference frames can be based on different aspects of the environment and are mainly divided into two types: egocentric and allocentric reference frames. By an egocentric reference frame we mean the relationship between objects and the observer. If observers use an egocentric frame of reference, they store information about positions of objects relative to their own position.

In an allocentric reference frame, on the other hand, one does not relate the spatial information to the observer, but to other objects in the environment. Of course, this frame of reference is dependent on the location of the observer. We will focus on the allocentric reference frame that could be used by the observers in our experiments. The frame consists mainly of the properties of the experimental room and the objects in the room. The experimental room has walls, a floor, a ceiling and some doors and windows. These elements together form several straight lines in the visual field of an observer. Based on a few assumptions about the usual shapes of objects, the visual input provides a basis for an allocentric reference frame. The assumptions people use are that windows and doors, for example, usually have rectangular shapes and that the walls are orthogonal to each other and to the floor and ceiling of the experimental room. If this is not the case, observers tend to make errors in judging the positions of objects. These reference frames provide a structure that an observer can use to make judgments about the positions of objects. Whether they actually use the available structure is a another question.

1.3 Depth cues

Humans have a fairly good sense of depth. However, the fact that we perceive depth is quite intriguing, since the image that is projected onto our retinae is two- instead of three-dimensional. Somehow, we derive the third dimension from these two-dimensional images. A number of possible sources of information (cues) have been described in the literature. However, the fact that these sources of information are present for observers does not mean that they actually use them to deduce depth in the scene. So we will have to be careful about assuming a certain degree of effectiveness when some cue is present. We can divide the various depth cues into three groups. I will discuss some physiological depth cues and pictorial depth cues below. Another important group of depth cues is derived from motion. I will not go into detail about these motion-related cues since they will not be discussed in the rest of the thesis.

Physiological depth cues

By physiological depth cues, I mean aspects of the physiology of the sensory organs that can contribute to the perception of depth. Ever since scientific interest in depth perception began, physiological depth cues have been regarded as very important. In fact, they are important for seeing depth in the range of 1 to 4 meters from the eyes. However, when distances from objects are larger, these cues are not as effective as people often assume.

> ***Box 1 Experiencing binocular disparity***
>
> Close your right eye and hold your right index finger at an arm's length from your nose. Hold the left index finger somewhere halfway between your eye and the right finger, so that your right index finger is blocked from view by the left one. Then close your left eye and open your right one. You'll notice that both fingers are visible now! This is due to binocular disparity.

An important physiological depth cue is binocular disparity. We have two retinae onto which two images of the world seen from slightly different positions are projected. These two images differ slightly from each other depending on the positions of the objects that are projected onto the retinae. Thus by combining the information of these images we obtain information about the relative distances between the observer and the objects in the scene. This difference between the two images of the retinae is called binocular disparity (see box 1 for a demonstration).

Another physiological depth cue is called accommodation. Accommodation occurs when the lenses of our eyes deform when the distance of the objects in focus changes. Accommodation of the lens is required to produce a sharp image of the object. This cue is derived from the muscle tension that is needed to have an object in focus; hence it can produce an absolute distance estimation as a function of muscle-tension.

A third physiological depth cue involves the vergence movement of the two eyes. If a person focuses on an object, then the two eyes will rotate so that the object is projected onto the fovea of the retinae. If this object is close to the observer, the two eyes will rotate inwards, but they will rotate outwards if the object moves further away from the observer. The tension of these muscles will therefore also give information about the absolute distance to the object that is in focus.

Pictorial depth cues

Although many disciplines already recognized the importance of pictorial depth cues, for many years the scientific community focused its attention mainly on physiological depth cues. However, pictorial depth cues are very important for perceiving depth. By pictorial depth cues we mean the cues that we can use to see depth in a picture. However, they are very useful for estimating depth in the actual world. There are lots of different pictorial depth cues, so I do not intend to name them all.

One of the oldest known depth cues is occlusion; when an object A is occluded by object B, object B is closer to the observer than object A. This is not an absolute cue to distance; thus one cannot give an accurate judgment of the distance from one of the objects solely on the basis of this cue. However, occlusion is a very reliable source of relative distance information for multiple objects.

Another well described pictorial depth cue is linear perspective. This depth cue makes use of the fact that parallel lines in the physical world around us seem to converge to one single point on the horizon(when not in a frontoparallel plane). A closely related cue is the relative size cue, which involves the fact that objects further away from an observer produce smaller retinal angle sizes than objects that are closer to the observer. Together with some knowledge of the sizes of familiar objects, this cue can even give absolute estimates of the distances of single objects. Gibson (1950) introduced the term texture gradient to refer to an effective depth cue in our environment. This cue is related to both the relative size cue and linear perspective. Texture gradients involve the fact that textures on surfaces change in size and compactness as the distance from the observer changes. Other pictorial depth cues are aerial perspective, height in the visual field etc.

Prior knowledge

There are numerous cues that contain information about depth. How an observer chooses which sources of information to use is also subject of debate. First, some scientists reasoned that observers weigh up the reliabilities of cues and will rely more on cues that have proved to be most useful in the past (Ames, 1953). In the Ames' room example (Box 2), linear perspective apparently is considered a more reliable cue than relative size (of the people in the room). Another possibility is that observers choose the possible lay-out of the scene that has been encountered most often before (Gibson, 1966). A rectangular shaped room is more familiar to most people than a trapezoidal shaped room; thus the Ames' room of Box 2 is seen as a rectangular shaped room. According to Yang and Purves (2003), our brains use Bayesian calculations to let us perceive the most probable scene that could produce the retinal image. These two views on the weighing up of cues or situations need not be contradictory, they can complement each other.

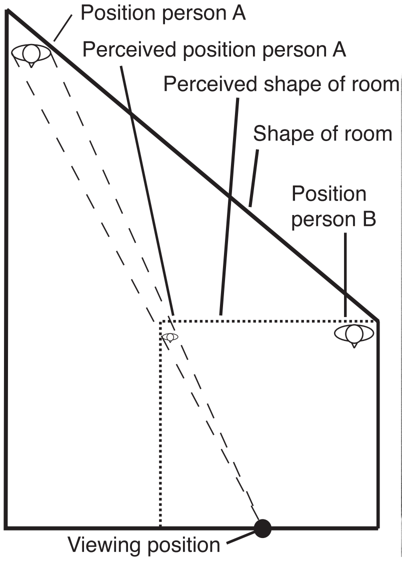
As well as finding out what kind of information sources we use to see depth, we can also investigate the accuracy or veridicality of our depth perception. I will refer to this work as visual space research. By visual space (short for visually perceived space) I mean the impression we have of our environment (physical space) on the basis of visual information.

1.4 Visual space research

Visual space research has a long tradition. The first systematic experiments were done under rather restricted conditions: they were done in dark rooms with small point-lights as stimuli and the heads of the observers were fixated. In this way scientists reduced the available cues to the physiological depth cues described above. This is why they talked about binocular visual space. Examples of these experiments are those performed by Hillebrand and Blumenfeld (Battro, di Pierro Nettro, & Rozestraten, 1976). They let observers make visual alleys consisting of points of light. These alleys were based either on parallelity (two rows of lights had to be placed parallel) or on equidistance (each pair of points is equidistant from the other pairs of points). In a Euclidean geometry these two tasks should result in the same settings, the only difference being the explanation given to the observers. However, differences were found between these two tasks. The difference between settings for these two tasks inspired Luneburg (1947, 1950) to formulate his conjecture that binocular visual space has a homogeneous non-Euclidean Riemannian metric. In other words there is a single visual space that is constant over viewing conditions, distances and tasks. If the curvature of visual space is zero, this would mean that visual space is Euclidean, which is the description

Box 2 Ames' room

One way of introducing the notion of depth cues is to explain the effects of Ames' room. Ames was an artist who was interested in depth perception. He wrote a book with lots of small experimental set-ups that clarified issues that had not been solved at the time (Ames, 1953). His most famous set-up is the "room" he created. To an observer, who views the room from a certain viewpoint, it seems to be a normal rectangular shaped room. However, the people in the room seem to be substantially different in size. When you look at the ground-plan of the room, you can see that the room is not rectangular-shaped at all and the people in it are approximately the same size but are at different distances from the observer. This effect is due to the strong effect of familiarity with rooms. Usually rooms are rectangular shaped, and windows and doors are also rectangular shaped. Thus a trapezoid that is extended in depth is easily mistaken for a square that is in another orientation.



we give to the physical space. In a Euclidean metric, parallel lines never intersect and remain equidistant from each other. If the curvature is positive (the metric is said to be spherical) or negative (the metric is said to be hyperbolical) parallel lines do not remain at the same distance from each other. Blumenfeld, however, observed that equidistance alleys always lay outside parallel alleys and did not look parallel. This observation led Luneburg to conclude that the Riemannian metric of visual space had a negative (hyperbolic) curvature. Most of the early scientists agreed with Luneburg about the curvature of this metric (Blank, 1961; Zajaczkowska, 1956); however, there is still no consensus on this matter.

In order to generalize our knowledge to everyday vision, we need to take into account the other cues that contribute to our sense of depth in the visual world. Other cues like motion-related cues and pictorial cues provide a rich source of information about depth. For example, people with one eye can see depth, although they have no binocular disparity or convergence information available. Due to some behavioral adaptations, however, they have no difficulty in dealing with the world. Therefore we need to look at all depth cues in order to try and understand human depth perception. So far, most research into visual space that was done under less restricted conditions has taken place in outdoor settings. Larger distances were used than in the experiments discussed earlier, pictorial information was present (due to testing in daylight) and often observers could move their heads freely. Gilinsky (1951), for example, used a bisection task to study the deformation of visual space. In her experiments, a line was stretched across a lawn, starting a few inches from the observer. The task was to bisect the line in two equally long parts. She concluded that visual space can be described with a compressing distance function that is asymptotic to a constant (c in Equation 2.1). This distance function was derived from Luneburg's metric of visual space, the principles of linear perspective and the law of size constancy. The distance function has constant c that varies between observers and experimental conditions. Gilinsky theorized that c was also dependent on the availability of cues to distance.

Battro et al. (1976) did experiments with different tasks in large open fields. They compared the curvature of visual space using three tasks: visual alleys, horopters and triangles. Battro and colleagues (1976) concluded that visual space can be described by a Riemannian metric with a variable curvature. The curvature varied with scale and between observers. They demonstrated that visual space is not homogeneous in nature and thereby they falsified one of Luneburg's assumptions. Foley (1972) and Koenderink, van Doorn and Lappin (2000) came to the same conclusions.

The work described in this introduction mainly concerns measurements in horizontal planes. Indow and Watanabe (1984, 1988) examined not only the horizontal sub-space, but also the frontoparallel sub-space. They found no systematic deviations in this plane. Hence, they concluded that visual space is deformed in different ways for different sub-spaces: hyperbolic for the horizontal sub-space and Euclidean for the frontoparallel sub-space.

The aim of scientists nowadays is not to falsify Luneburg's conjecture, but to find new ways of describing visual space. For example, Kelly, Loomis and Beall (2004) used other tasks (a body pointing task and a collinearity task) to measure visual space in an open field. They concluded that systematic misjudgment in exocentric direction could not be caused by misperceived egocentric distances. However, Koenderink, van Doorn and Lappin (2003) were able to explain the misjudgments of egocentric direction, measured with an exocentric pointing task, in terms of a misperception of egocentric distances. Thus, we can conclude that not all the work done so far can be explained by one existing theory. Experiments have even produced conflicting results. Thus, one cannot speak of a single visual space. Instead, the structure of visual space seems to depend on numerous factors like the kind of task that is being used (Koenderink et al., 2000; Koenderink, van Doorn, Kappers, & Lappin, 2002), the viewing conditions under which the experiments are done (Wagner, 1985), the distances that are used and individual differences between observers (Battro, et al., 1976; Koenderink, et al. 2002).

Cuijpers and colleagues did laboratory research on visual space in an indoor setting. In a laboratory environment they tested whether observers showed systematic deviations from veridical settings under normal lighting conditions. Their observers were seated in an artificially illuminated room. To prevent the observers from deriving pictorial depth information from structures besides the actual objects used in their experiments Cuijpers and colleagues covered the walls of the experimental room with wrinkled plastic. They had observers seated in a small cabin that restricted the vertical field of view. These manipulations led to the situation where the observers could not see the floor, ceiling or walls of the experimental room. In addition to these restrictions, the heads of the observers were fixated with a chin-rest. And lastly, the objects that were used to do the task were scaled with distance from the observers, so that the observers always received an equally sized projection onto their retinae.

Cuijpers and colleagues did their research via three different tasks. The first task was an exocentric pointing task (Cuijpers, Kappers, & Koenderink, 2000A). In this task, a pointer could be rotated with a remote control. The task for the observer was to direct the pointer towards a target, which was a small sphere. The second task they used was a parallelity task (Cuijpers, Kappers, & Koenderink, 2000B). During this task, two rods are in the visual field of the observer. The experimenter placed one of the rods in a certain orientation. The observer's task was to place the other rod in the same orientation as the first one, i.e. to put it parallel. The last task was the collinearity task, in which also two rods were placed in the observer's visual field (Cuijpers, Kappers, & Koenderink, 2002). The task was to rotate them both so that they were in line, i.e. collinear. By means of these experiments Cuijpers and colleagues produced data that led them to the conclusion that the structure of visual space is dependent on the task that is given to a certain observer. Therefore, in their view one cannot speak of a geometry of visual space in general (Cuijpers, et al., 2002).

1.5 This thesis

Our first experiment was based on Cuijpers' work. Cuijpers and colleagues investigated whether a single visual space could also be defined in an illuminated environment. Although they found structural deviations, they found varying patterns for different tasks. We began by extending this work by introducing free viewing conditions (Chapter 2). Secondly, we changed the 2D exocentric pointing task into a 3D version, and tried to extrapolate to 3D scenes the knowledge we had gained from experiments in the horizontal plane (Chapters 3 and 4). Thereafter, we investigated in detail an interesting pattern we had found in the earlier chapters (Chapter 5). And finally, we compared our exocentric pointing task with a new task that we developed (Chapter 6).

Two lines of research can be distinguished in this thesis. One concerns spatial parameters and their effects on the perception of the positions of objects (Chapters 2, 3 and 6). The other line mainly concerns the effects of contextual information and reference frames (Chapters 4 and 5). This work should help us to understand which spatial and contextual parameters influence our perception of depth.

Chapter 2

In this chapter we discuss three experiments that we did to test whether systematic deviations of the visual perception of the positions of objects still occur under free viewing conditions. By free viewing conditions we mean that the observers had an unobstructed view of the walls, floor and ceiling of the experimental room. Furthermore, they could rotate their heads and upper-bodies freely. We used the exocentric pointing task, the parallelity task and the collinearity task as described above. All measurements were made in the horizontal plane on eye-height. We looked at the effect that relative distance and the horizontal separation angle had on the observers' settings and compared our findings to the results reported by Cuijpers and colleagues.

Chapter 3

In Chapter 3 the 3D exocentric pointing task is introduced in order to explore visual space in three dimensions. The pointer and ball could be positioned at different heights. In addition to rotating the pointer in the horizontal plane, the observer could also rotate it in the vertical plane. Thus, we used two dependent variables in this research: the slant and the tilt (the orientation in the horizontal and vertical orientation). We varied the horizontal and vertical separation angles and the relative distance. If visual space were isotropic, then the tilt would be dependent on the vertical separation angle as the slant is dependent on the horizontal separation angle. Moreover, both variables should be dependent on the relative distance.

Chapter 4

In this chapter we describe the experiments that were conducted to test whether external and internal references were used during a 3D exocentric pointing task. We compared a condition in which observers directed the pointer towards a ball while the pointing-direction was parallel to one of the walls with a condition that the pointing-direction was not parallel to one of the walls. Besides this, we compared conditions in which the pointing-direction was frontoparallel or not. By frontoparallel we mean in a plane perpendicular to the line of sight when one is looking straight ahead.

Chapter 5

In all our experiments, we found a difference for trials in which the pointer was positioned far from the observer (and the ball close by) and trials in which the pointer was positioned close to the observer (and the ball far away). The deviations we found were larger in the latter condition than in the first condition. In this chapter, we tried to find an explanation for this phenomenon. We tested two different explanations. First, we tested whether the position of the observer was used as a reference when the pointer was far from the observer. The position of the observer does restrict the pointing angle when the pointer is far from the observer. In our experiments we restricted the pointing angle to the same degree for the other condition (when the pointer is close to the observer) by means of poster-boards. In another experiment, we tested whether the different views of the pointer could explain the difference in deviations. If a pointer gives less information about its orientation when the pointer is close by and the ball far away, this could result in asymmetry in the amount of information an observer can use to orient the pointer.

Chapter 6

In the last chapter of this thesis we investigate visual space by means of an entirely different task. In this task, three red balls are hanging at different heights from the ceiling. A fourth ball can be adjusted in height by the observer. The task is to hang it in a plane that is defined by the three red balls. In this chapter, we describe the effect that the orientation in which the plane is tilted has on the deviations. Furthermore, we tested this with the red balls forming three different triangles; namely an acute, an obtuse and an equilateral triangle.