Chapter 2
Perceptual Aspects of VR

Ralf Doerner and Frank Steinicke

Abstract Virtual Reality (VR) has the special ability to provide the user with the illusion of presence in a virtual world. This is one aspect of the valuable potential that VR possesses concerning the design and realization of human–machine interfaces. Whether and how successfully this potential is exploited is not only a technical problem. It is also based on processes of human perception to interpret the sensory stimuli presented by the virtual environment. This chapter deals with basic knowledge from the field of human information processing for a better understanding of the associated perceptual issues. Of particular interest in VR are the perception of space and the perception of movement, which will be dealt with specifically. Based on these fundamentals, typical VR phenomena and problems are discussed, such as double vision and cybersickness. Knowledge of human perception processes can be used to explain these phenomena and to derive solution strategies. Finally, this chapter shows how different limitations of human perception can be utilized to improve the quality and user experience during a VR session.

2.1 Human Information Processing

The way that people perceive and process information is essential for the design of virtual environments and the interaction within them. Ultimately, every virtual environment is used by humans. For this reason, it is useful to study the basic functions of human information processing to better understand the various effects and phenomena of VR and to be able to take advantage of possible limitations.

Dedicated website for additional material: vr-ar-book.org

R. Doerner (✉)
Department of Design, Computer Science, Media, RheinMain University of Applied Sciences, Wiesbaden, Germany
e-mail: ralf.doerner@hs-rm.de

R. Doerner et al. (eds.), *Virtual and Augmented Reality (VR/AR)*,
https://doi.org/10.1007/978-3-030-79062-2_2

Humans perceive their environment through different senses. In the context of today's VR technologies, the most important senses are:

- the visual sense,
- the acoustic sense, and
- the haptic sense.

In most of today's VR systems, other senses, such as the olfactory (smelling) or gustatory (tasting) senses, are not stimulated. Thus, most information presented in the virtual environment is perceived through the eyes, ears, or skin. At first glance, perception in a virtual environment does not differ from perception in a typical desktop environment and the associated senses and sensory impressions. The virtual worlds presented on the screen or from the loudspeakers act as visual or acoustic stimuli; haptic impressions are conveyed via mouse and keyboard. An important aspect of the VR experience is the possibility to explore the virtual world in an immersive way. In contrast to desktop-based environments, in VR this is not only done by mouse and keyboard but by 3D input devices or by movements of the user in real space, which are mapped to corresponding movements in the virtual world. In addition to these inputs into the VR system, there are other forms of input, such as speech, gestures, and other forms of human expression (Preim and Dachselt 2015).

To better understand the complexities of human perception and cognition, it is helpful to imagine humans as an information processing system (see Fig. 2.1). In this metaphor from the field of computer science, all physical characteristics of humans are assigned to hardware and all psychological characteristics to software. The chain of information processing starts with an *input,* which is *processed* in the computer and finally presented as *output* on the output media. In human information processing, stimuli from the external world are thus first transferred to the perceptual system as input and perceived there (Card et al. 1986a). This *perceptual processor* has access to memory (e.g., visual memory) and processor (e.g., for pre-filtering) similar to the input to the computer. The processing of the resulting perceived stimuli then takes place in the *cognitive processor.* Here, further memories, i.e., the working and long-term memories, can be accessed to interpret the stimuli and plan appropriate action. The actual action then takes place in the *motor processor,* which initiates corresponding movements.

These partly substantial simplifications only approximate the much more complex biological processes, but they allow us to make predictions about human information processing. For example, Card et al. (1986a) were able to predict the time required for a whole series of human interaction tasks. This model makes it clear, among other things, why tasks that require the cognitive processor to be run through several times (e.g., comparisons) require more time than those tasks in which the cognitive processor is only run through once (e.g., simple response to stimulus).

In this context, a whole range of other models, such as *GOMS* or the *Keystroke-level Model* (KLM), can be mentioned, which are used in the field of human–computer interaction (Card et al. 1986b; Sharp et al. 2019; Shneiderman et al. 2018). In the following, we want to give a more detailed insight into the individual components of human information processing.

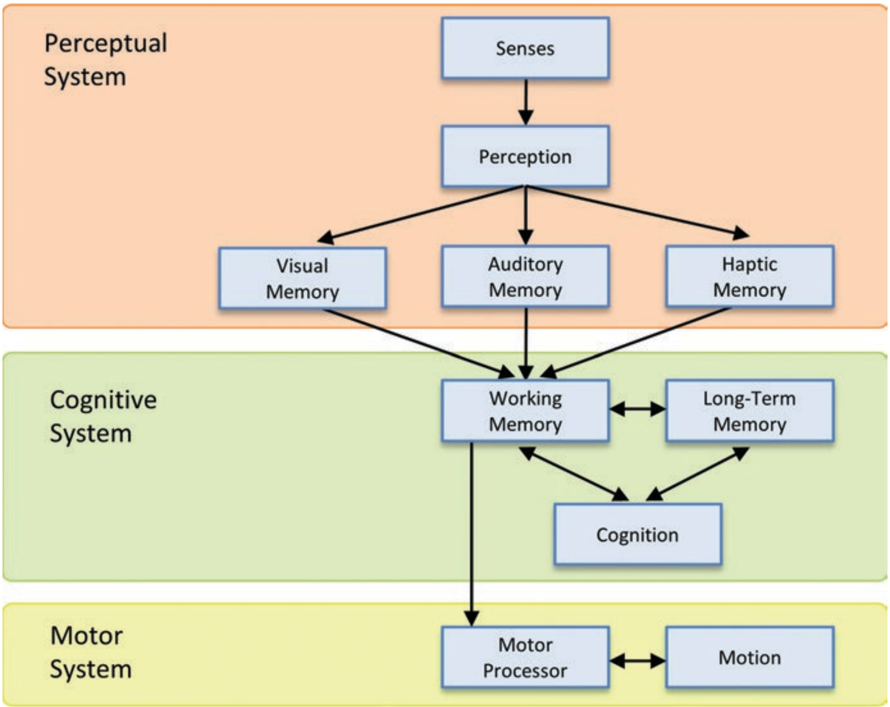

Fig. 2.1 Model of human information processing. (According to Card et al. 1986a)

2.2 Visual Perception

The visual system is the part of the nervous system responsible for processing visual information. The structure of the human eye allows light to be projected through the lens onto the inner *retina*. There are about 120 million photoreceptor cells. These are divided into the *rods*, which only perceive brightness, and the approximately 7 million *cones*, which are responsible for color vision. The cones, in turn, can be divided into three types, each of which reacts to blue, green or red hues. The optical apparatus of the eye produces an upside-down and reversed image on the retina. For the perceived image to arrive sharply on the retina, the lens must be correctly adjusted by muscles depending on the distance of the object being viewed. This process is called *accommodation*. The *fovea* is the retina area with the highest image sharpness and the highest density of photoreceptor cells. Although the eye has an aperture angle of approximately 150° (60° inside, 90° outside, 60° above, and 75° below), only 2° to 3° of the field of vision is projected onto the fovea. Under ideal conditions, the resolving power is about 0.5–1 min of angle. This means that a 1 mm spot can be perceived from a distance of about 3–6 m. The eye only remains at such a fixation point for a period of about 250 ms to 1 s before rapid, jerky eye movements (known as *saccades*) occur. These saccades serve to complement peripheral perception, in which the resolution is only about one-fortieth of the foveal resolution, and thus enable us to perceive a complete high-resolution image.

In particular, visual perception enables us to identify objects. For this purpose, the projected image of the scene is already analyzed in the retina (e.g., brightness, contrast, color and motion) and processed (e.g., brightness compensation and contrast enhancement). During transmission via the optic nerve, the spatial relationships of the photoreceptors are retained in the nerve tracts' positional relationships and synapses. This positional relationship can be detected in the visual cortex as a neural map and supports, for example, the identification and differentiation of objects (Marr 1982). The recognition of individual elements and their meaning is probably done by comparison with already stored experiences (scenes linked to body sensation, emotions, smell, sounds, and much more).

2.2.1 Stereo Vision

As an example of how human perception works and how it can be manipulated by a VR system to create presence in the virtual environment, we consider a phenomenon important for VR: *stereopsis*, also called stereo vision. Humans have two eyes but do not perceive two separate images of reality. In addition, the visual system succeeds in obtaining a three-dimensional impression of the environment from the light stimuli impinging on the two-dimensional retina of the eyes.

Let us consider point *A* in Fig. 2.2a. If we assume the eyes have fixated on point *A*, then they have been adjusted so that light from point *A* enters both the fovea of the left eye (and impinges on the retina at point A_L) and the fovea of the right eye

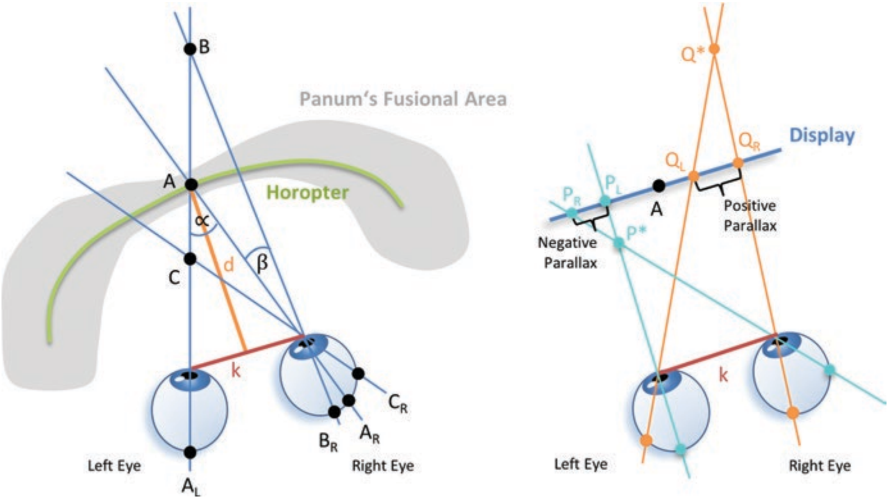

Fig. 2.2 (**a**) Stereopsis. (**b**) Manipulation of the stereopsis with a stereo display

(there at point A_R). Adjusting means that the eye muscles are moved accordingly. The closer the point A between both eyes is to the observer, the more the eyes must be turned inwards towards the nose to fixate on A. This movement of both eyes is called *convergence*. As the visual system has information on how big the convergence is, the angle α can be estimated in the triangle A, A_L and A_R, because the bigger the convergence, the bigger α is. With the knowledge of α and the distance k of both eyes, which is constant for a person, the distance d of point A from the observer can be concluded. By simple trigonometry, the following relationship between d and α can be established: $d = k/ (2 \cdot \tan \alpha)$. With this triangulation of A, which is only possible with two eyes, the visual system can thus perceive the distance of A.

The points A_L and A_R are called *corresponding points* of the retina. They would be in the same place if the two eyes were thought to be superimposed. The visual system is able to determine this correspondence. All points in reality that are mapped onto corresponding points on the retina form the *horopter*. It has the shape of a surface curved around the head, which contains the fixation point. Let us now look at point B in Fig. 2.2, which is not on the horopter. In the left eye, light from B still strikes at point A_L, while in the right eye, it strikes at point B_R. The points A_L and B_R are not corresponding points. The difference between B_R and the point A_R corresponding to A_L is called the *disparity* created by B. Disparities are often given as angles; in our example in Fig. 2.2 this would be the angle β. The larger β is, the more the point B is away from the horopter. The disparity generated by B thus provides a point of reference for determining the distances of points like B, which, unlike A, are not fixated on and whose distance cannot be determined directly based on eye convergence alone.

Two Small Experiments on Convergence and Disparity

1. Hold a pen at a distance of about 1 m in front of a person's face. Ask the person to fixate on the tip of the pen and leave it fixed. Now move the pen towards the person's nose so that you can easily observe the convergence: the eyes are directed inwards towards the nose.
2. Sit in front of a rectangular object (e.g., a monitor), close your right eye and hold your index finger so that the left index finger points to the left edge of the object and the right index finger to the right edge. Now open the right eye and close the left one. The object seems to jump relative to the fingers – the right and left eyes perceive a slightly different image; there are disparities.

Retinal disparities also allow us to obtain information about the distance of points that are in front of the horopter from the observer. Point C in Fig. 2.2 is such a point, and while light from C in the left eye also arrives at point A_L, this happens in the right eye at point C_R. The disparity now exists between A_R (the point corresponding to A_L) and C_R. The point C_R lies to the right of A_R, while B_R lies to the left of A_R. B creates an *uncrossed disparity* and C a *crossed disparity*. Whether a point lies behind or in front of the horopter can be distinguished by the fact that in the first case uncrossed disparities are generated and in the second case crossed disparities are generated.

If the disparity becomes too large, i.e., the point generating the disparity is too far away from the horopter, the visual system is no longer able to fuse the image information of both eyes into one image. As a result, one no longer sees one point but two points. All points in the world that create disparities small enough to allow a fusion of the image information from the left and right eye form *Panum's fusional area*. This area has the smallest extension around the point the eyes fixate on.

In a virtual environment, stereopsis can be manipulated with the aim of creating a three-dimensional impression, even though only a two-dimensional display surface is used. Figure 2.2b shows that a *display surface* is viewed by an observer. Viewing means that the observer fixates on a point A on the display surface with the eyes. We now illuminate two points P_L and P_R on the display surface. By taking the technical precautions described in detail in Chap. 4, we ensure that light from P_L only hits the left eye and light from P_R only the right eye. The distance between P_L and P_R on the display surface is called *parallax*. The visual system can react to this situation in two ways. First, two different points are perceived. In reality, it happens all the time that light from points in the world only enters one of the eyes. The visual system can also spatially arrange such points in relation to points from which light falls into both eyes and whose location could already be deduced (*DaVinci-stereopsis*). Secondly, the visual system explains the light stimuli at points P_L and P_R by the fact that the light comes from a single point P^* located in front of the display surface. P^* is the *fusion* of P_L and P_R. Which of the two cases actually occurs depends on a number of factors, such as how far the apparent point P^* is located from the display surface. If the visual system merges P_L and P_R, then a point outside the display surface is successfully displayed. It is also possible to create points behind the display surface by reversing the order of the points for the left and right eyes on the display surface. This is shown in Fig. 2.2 at point Q_L and Q_R, where the two points shown on the display could be fused to form a point Q^* behind the display. When P_L and P_R are displayed, this is called *negative parallax*, while in the case of Q_L and Q_R one speaks of *positive parallax*.

In VR, it is, therefore, possible to create a *stereo display* by exploiting the peculiarities of human perception. The visual system creates not only a two-dimensional but also a plastic three-dimensional image impression, in which objects appear in front of or behind the screen based on an appropriate selection of the parallax. This must be distinguished from true three-dimensional displays (*volumetric displays*), in which, for example, a display surface is moved in space.

2.2.2 Perception of Space

Not only disparities are used by the visual system to perceive spatiality and the arrangement of objects in space. This can be seen by the fact that there are people who are unable to evaluate information from disparities ('*stereo blindness*') but nevertheless develop a three-dimensional idea of the world. There are no exact figures, but it is estimated that about 20% of the population is stereo blind. A test can be used to determine stereo blindness in the same way as a test for color vision defects. It is recommended to perform such a test, especially for people who are active in the field of VR. Many people are not aware that they are stereo blind.

Today we know a whole series of clues, called *depth cues*, which are used by the brain for the perception of space. Disparity is an example of a depth cue. If a car covers a tree, the visual system can derive the information that the car is closer to the observer than the tree. This information does not require the interaction of both eyes. Thus, this clue is called a *monocular depth cue*. As it is still possible to obtain depth cues even from 2D images, this is also referred to as a *pictorial depth cue*. Disparity, on the other hand, is a *binocular depth cue*. With depth cues, one can distinguish whether they help to estimate the spatial position of an object absolutely or only relative to another object. Convergence, for example, allows an absolute position determination, whereas occlusion only permits a determination relative to the occluded object.

The informative value and reliability of the various depth cues depend in particular on the observer's distance to the respective object. While occlusion provides reliable information in the entire visible range, this is not the case for disparity. The further away a point is from the observer, the lower the disparity it generates. A point at a distance of 2–3 m produces a very small disparity. From a distance of 10 m, the disparity is de facto no longer perceptible. For VR, this means that for virtual worlds where significant objects are within arm's reach, the effort to use stereo displays should be invested. Disparity is essential in this area. For virtual worlds, however, where objects are more than 3 m away from the viewer, the use of a stereo display does not contribute much to the perception of space and may be superfluous.

Table 2.1 lists various depth cues and gives details of the area of action and the information content (indications of relative arrangement or absolute distance determination), as well as the category (monocular depth cue, binocular depth cue or *dynamic depth cue*, the latter being understood as depth cues that the observer receives through movement). The depth clues mentioned in the list are all of a visual nature, but the brain can also obtain cues from other senses, e.g., by interpreting information from touch or by analyzing the pitch of a moving object's sound. As it is essential for a good perception of a virtual world to give as many depth clues as possible in VR, we go through the list below. *Occlusion*, *disparity* and *convergence* have already been discussed. Similar to convergence, where muscle tension is taken into account to align the eyes, the brain also uses the muscle tension necessary for *accommodation*, the adjustment of the refractive power of the eye lens, as a depth cue. To see nearby objects clearly, the eye lens must be pressed together with more muscle power than is the case with distant objects. If a person fixates on an object at a certain distance, other objects appear sharp only in the vicinity of this object (e.g., in the distance range 75 cm to 1.5 m if the fixed object is 1 m away from the observer). Objects that are too far away or too close to the observer appear blurred. From the *image blur*, it is, therefore, possible to draw conclusions about the distance

Table 2.1 List of depth cues (with range of action and classification)

Depth cue	Range of action	Classification	Positioning
Occlusion	Complete range	Monocular	Relative
Disparity	Up to 10 m	Binocular	Relative
Convergence	Up to 2 m	Binocular	Absolute
Accommodation	Up to 2 m	Monocular	Absolute
Image blur	Complete range	Monocular	Relative
Linear perspective	Complete range	Monocular	Absolute
Texture gradient	Complete range	Monocular	Relative
Relative size	Complete range	Monocular	Absolute
Known quantity	Complete range	Monocular	Absolute
Height in the field of view	Over 30 m	Monocular	Relative
Atmospheric perspective	Over 30 m	Monocular	Relative
Shape from shading	Complete range	Monocular	Relative
Shadows	Complete range	Monocular	Relative
Motion parallax	Over 20 m	Dynamic	Relative
Accretion	Complete range	Dynamic	Relative

of objects. *Linear perspective* is a depth indication based on perspective distortion. Objects further away appear smaller; in reality, parallel lines seem to converge at a vanishing point (see, for example, the street in Fig. 2.3a).

Also, with textures, the texture elements become smaller with increasing distance. Thus, the *texture gradient* can serve as a depth cue. For similar objects, such as the three squares in Fig. 2.3a, which have different sizes in the image, the visual system assumes that the differences in size can be explained by different distances (and not by the fact that the objects themselves are of different size: assumption of *size constancy*). This depth cue is called *relative size*. However, the *known size* also contributes to distance estimation. We get a good impression of the size and orientation of the triangle in Fig. 2.3a because a person is standing next to it – and thus an object of which we know the size and the usual orientation in space. Moreover, the *height in the field of view* is an indication of depth. In Fig. 2.3a, square *C* is arranged higher in the image than square *A* and thus closer to the horizon line. This indicates that square *C* is further away. Connected to this is also the direction of view. If one has to look straight ahead or raise the head, the object is assumed to be further away (Ooi et al. 2001). Very distant objects do not appear so rich in contrast and have a slightly bluish coloration (cf. Fig. 2.3b), because more air and the particles it contains lie between the observer and the object (*atmospheric perspective*). The illumination of objects gives clues about their arrangement in space. On the one hand, shaded objects appear more spatial (*shape from shading*, cf. left pyramid with shading, right pyramid without in Fig. 2.3c); on the other hand, the *shadows* cast give cues about the spatial arrangement of objects (cf. shadows of spheres in Fig. 2.3d). It is especially effective when shadows are cast from above on a base surface because the visual system is used to a light source from above: the Sun. If the object is in motion, the shadow of this object is particularly useful for depth perception.

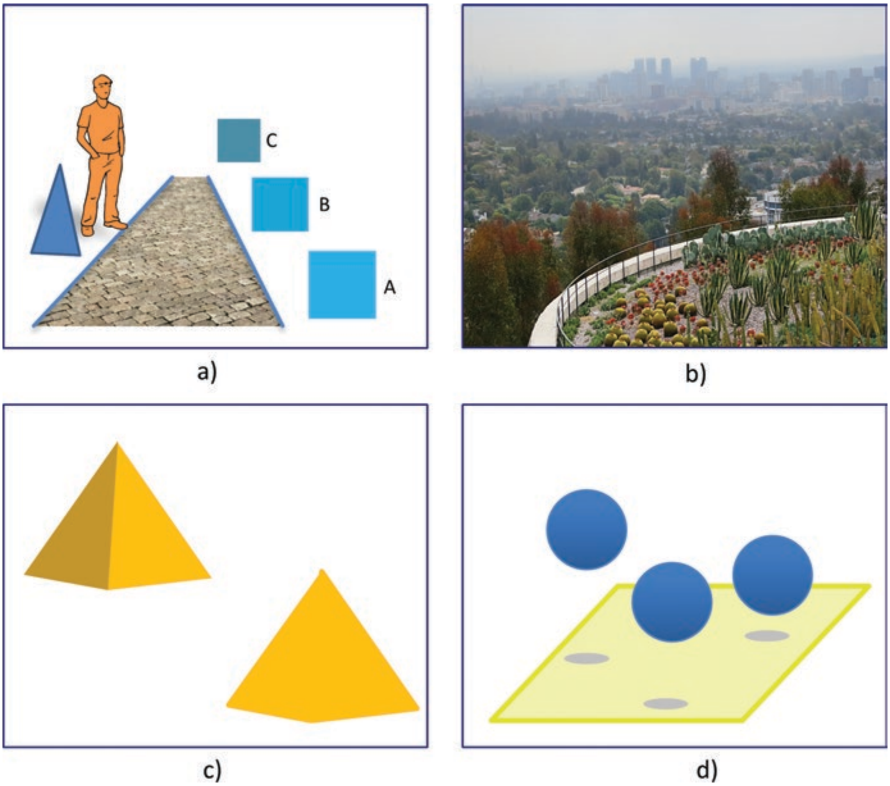

Fig. 2.3 Examples of depth cues

Finally, certain depth cues are based on movement: movement of objects or movement of the observers themselves. This includes *motion parallax*: the light stimuli from near objects move faster across the retina than those from farther away. If we drive through an avenue in a car, the nearby trees pass us quickly while the mountains in the background move only slowly. Through movement, objects suddenly become occluded or reappear behind the objects that are obscuring them. This change, called *accretion*, also gives cues to the spatial arrangement of the objects.

Depth cues are not to be considered independently of each other. For example, accommodation and convergence depend on each other (Howard 2002). Also, depth cues are of varying strength. For example, while accommodation is a weak depth cue, occlusion is a strong depth cue. All depth cues are considered for spatial perception in the form of a weighted sum. How much weight is given to a depth cue is flexible and depends on the distance of the object to be assessed. One theory (Wanger et al. 1992) assumes that the weights also depend on the current task the observer is engaged in. If the task is to estimate the spatial arrangement of distant objects, then motion parallax, linear perspective, texture gradient and shadows have a high weight. If the task is to grasp an object, disparity, convergence and accommodation are important. According to this, the depth cues in the brain are not used to form a single model of the 3D world, which is then used for different tasks, but rather task-dependent models are formed. Therefore, if not all depth cues can be generated in a VR, then a prioritization should be made depending on the task the user has to perform.

2.3 Multisensory Perception

Even though the visual sense is undoubtedly the most important source of information in the perception of virtual worlds, the auditory and haptic senses also play an increasingly important role (Malaka et al. 2009). In this respect, these two senses will also be examined more closely in the context of this chapter. Other senses, such as smell and taste, play more of an exotic role and are currently mostly used as prototypes in research laboratories. At this point, it should be noted that perceptions via the individual sensory organs are by no means processed separately, but rather an integration of the different impressions is created. For further literature, please refer to Ernst (2008).

2.3.1 Auditory Perception

The ears enable humans to perceive air movements. Such air and pressure fluctuations generate mechanical waves that hit the ear, which is made up of the outer, middle and inner ear. The auricle (outer ear) collects sound waves and transmits them to the middle ear. In the middle ear, sound waves are converted into vibrations of the eardrum. The eardrum vibrations are transmitted to the cochlea via the ossicles (anvil, malleus and stapes). The sensory cells in the cochlea then convert the mechanical energy into electrical signals. Finally, these electrical nerve impulses are transmitted to the brain via the auditory nerve. The different frequencies can be perceived by hair cells in the inner ear. The waves perceived by humans have lengths of about 0.02–20 m, which correspond to audible frequencies in the range of about 18–0.016 kHz (Malaka et al. 2009). In contrast to the visual sense, the spatial resolution is much lower. The *Head-Related Transfer Function* (HRTF) or outer ear transfer function describes the complex filter effects of the head, outer ear, and trunk. The evaluation and comparison of the amplitudes are, along with the transit time differences between the ears, an essential basis of our acoustic positioning system. However, the absolute distinguishability of intensity and frequency has clear limits, so that two noise sources are only distinguished if they are several degrees apart. In contrast, the temporal resolution is much better and acoustic stimuli can be distinguished already at 2–3 ms temporal discrepancy. The principle of localizing noise sources at different receiver positions is also used in acoustic tracking systems (see Chap. 4).

2.3.2 Haptic Perception

Haptics, or haptic perception, describes the sensory and/or motor activity that enables the perception of object properties such as size, contours, surface texture and weight by integrating the sensory impressions felt in the skin, muscles, joints and tendons (Hayward et al. 2004). The senses that contribute to haptic perception are divided into:

- tactile perception (element of surface sensitivity),
- kinesthetic perception/proprioception (depth sensitivity) and
- temperature and pain perception.

These senses thus enable the perception of touch, warmth and pain. Such perception phenomena are based on receptors in the skin. The more such receptors are available, the more sensitive the respective body part (e.g., hand, lips or tongue) is. The most important receptors are the mechanoreceptors (e.g., pressure, touch or vibration), the thermoreceptors (heat, cold) and the nociceptors (e.g., pain or itching). The mechanoreceptors, for example, convert mechanical forces into nerve excitation, which are transmitted as electrical impulses to the sensory cortex, where they are processed. As a result, shapes (roundness, sharpness of edges), surfaces (smoothness and roughness), and different profiles (height differences) can be perceived.

Haptic output devices stimulate the corresponding receptors, for example, by vibration (see Chap. 5).

A small experiment on the spatial resolution of haptic perception: take a compass or two sharp pencils and test with somebody else or yourself where in your upper extremities you can best distinguish between two points of contact and where you can distinguish least.

2.3.3 Proprioception and Kinaesthesia

In contrast to surface sensitivity, depth perception describes the perception of stimuli coming from inside the body. Depth perception is essentially made possible by proprioception and kinaesthesia. Both terms are often used synonymously. However, we will use the term *proprioception* to describe all sensations related to body position – both at rest and in motion – whereas *kinaesthesia* describes only those sensations that occur when active muscle contractions are involved. Proprioception thus provides us with information about the position of the body in space and the position of the joints and head (sense of position) as well as information about the state of tension of muscles and tendons (sense of strength). Proprioception enables us to know at any time what position each part of our body is in and to make the appropriate adjustments. Kinaesthesia (sense of movement) enables us to feel movement in general and to recognize the direction of movement in particular.

These two senses are essential, considering that interaction in a virtual environment is largely carried out by active movements of the limbs. In VR, various devices are available to stimulate these senses, such as haptic joysticks, complete exoskeletons or motion platforms (see Chaps. 4 and 5).

2.3.4 Perception of Movement

Movement is a fundamental process in real and computer-generated environments. We move through the real world, for example, by simply walking, running, or driving a car or bicycle. In addition to the user's actual movements, most virtual worlds contain a multitude of movements of other objects. From a purely physical point of view, motion is defined as a change of location over time. In visual perception, the movement of a stimulus leads to a shift in the corresponding retinal image. Provided it has the same speed, the further away the stimulus is, the smaller is the retinal shift. Still, we mostly perceive the physical and not the retinal speed. This ability is called *speed constancy* (analogous to size constancy; see Sect. 2.4.5). The human body has elementary motion detectors available for the visual perception of movement, which detect local movements in a certain direction at a certain speed. More complex, global movements are composed of local movement stimuli.

Another essential sense in the perception of movement is the *vestibular sense*. Hair cells in the inner ear detect fluid movements in the archways of the organ of equilibrium. This then makes it possible to perceive linear and rotational accelerations. To stimulate the vestibular sense, motion simulators (platforms) are used in some VR systems. It is also possible, however, to create the illusion of an own movement by visual stimuli only. This illusion is called *vection* and is created, for example, in a standing train when looking at another train that starts moving next to it. This illusion is mainly based on the perception of the *optical flow*. The optical flow can be modeled as a vector field, i.e., each point *P* on an image is assigned a vector – whereby the image is not isolated but is part of a sequence of images in which pixels corresponding to *P* can be found. The direction of this vector indicates the direction of movement of the pixel *P* in the sequence of images. The speed of the movement can be determined from the length of the vector. In this respect, the optical flow is a projection of the 3D velocity vectors of visible objects onto the image plane. Accordingly, when we humans move, we receive a whole series of different movement cues, which are all integrated to derive a final perception of movement (Ernst 2008).

2.3.5 Presence and Immersion

As described at the beginning of this chapter, an essential potential of VR lies in the possibility to create in the user the illusion of *presence* in a virtual world. For example, the user should get the feeling of complete immersion in the virtual world. The term presence (cf. Chap. 1) describes the associated subjective feeling that one is oneself in the virtual environment and that this environment becomes real. Stimulus from the real environment is thereby faded out. On the other hand, immersion describes the degree of inclusion in a virtual world caused by objective, quantifiable stimuli, i.e., multimodal stimulations of human perception. Various studies have shown that presence occurs, particularly when a high degree of immersion is achieved. Presence is achieved when the user feels located in VR and behaves as in the real world. Various studies have shown that various virtual environment parameters have the potential to increase the presence of test subjects, such as a large field of vision, activated head-tracking and real walking (Hendrix and Barfield 1996). There are several questionnaires to measure the subjective feeling of presence (Witmer and Singer 1998; Slater et al. 1994). However, it is also possible to determine the degree of presence based on physiological data or human behavior. For example, a user with a high degree of presence in an apparently hazardous situation occurring in VR will respond physiologically, e.g., with increased skin conductance or heart rate (Slater et al. 1994).

2.4 Phenomena, Problems, Solutions

When using VR, one can observe surprising phenomena. From 1 s to the next, the presentation of a virtual world in a stereo display no longer succeeds. The viewer no longer sees the world plastically but sees everything twice. Users of VR start to complain about headaches or even vomit. Although the car's interior appeared spacious when first viewed in VR, the space in the real car is then perceived as disappointingly tight, even though the virtual car and the real car are identical in terms of proportions. With knowledge of human perception, one can try to explain these phenomena and develop solution strategies to avoid or at least mitigate the resulting problems. With today's VR, we are not able to reproduce reality 1:1; there are always deviations. For example, the two images required for stereopsis for the right and left eye may have been generated at a distance between the two virtual cameras that does not correspond to the actual eye distance of an individual observer. Is that bad? Knowledge of human perception helps us to assess the magnitude of the problem associated with these deviations. The following eight subsections deal with VR-typical phenomena and problems. In each subsection, the currently known attempts at explanation are also presented as well as approaches to solutions that can be derived from them.

2.4.1 Deviating Observation Parameters

Let us assume that we recreate the Eiffel Tower and its surroundings in a virtual environment. With a virtual camera, we create an image and show it to a human observer. Light stimuli from this image are projected onto the retina in the eyes of the observer and create a visual sensation. Ideally, the image of the virtual Eiffel Tower creates the same impression that viewers would have if they were standing in front of the real Eiffel Tower. However, aberrations usually occur, which can be explained by deviations in the viewing parameters. The virtual camera generates images on a plane, while human retinas are curved. The angle of view of the virtual camera can deviate from the *field of view* of the observer. The observer does not necessarily look at the image from the same place where the virtual camera was standing – the observer might be closer or further away, perhaps not looking perpendicularly at the image but from the side. As a result, enlargements or reductions, as well as distortions of image impressions, occur. This affects the estimation of distance or the perception of the inclination of objects (Kuhl et al. 2006).

However, the distortions caused by looking at the image of the virtual world from a different perspective are surprisingly not experienced as bothersome. One speaks of the *robustness of linear perspective* in human perception (Kubovy 1986). This phenomenon can also be observed in a cinema – if the viewer sits in the first row on the very outside, he or she is very likely to have a completely different perspective than the camera that shot the film. There is, if at all, only one place in the whole cinema where the perspective of the film camera is maintained. Although this means that almost all viewers see the film in a distorted way, they do not mind. One explanation for this phenomenon is that the viewer's visual system actively corrects the distorted image impression. This correction is based, among other things, on the deviation of the viewing direction from the normal of the image plane (Vishwanath et al. 2005). Conversely, this active correction could be responsible for the fact that images taken with a wide opening angle of the virtual camera ('wide-angle perspective') may appear distorted even when viewed from the correct position.

Although deviating viewing parameters are not experienced as particularly irritating, it is advisable to strive to minimize the deviation. This is especially true for applications where the correct estimation of distances or orientation of objects in space is of high importance. It is particularly relevant if the virtual world is not only viewed passively, but active actions (grasping objects, movement) are performed. Moreover, the virtual world and one's own body should not be perceived simultaneously from different viewing positions. An approach to minimization of such deviations frequently pursued in VR consists of determining the current viewing parameters (e.g., by head-tracking, see Chap. 5), such as position and direction of gaze. If these are known, they can be transferred to the virtual camera. Another approach is to simulate long focal lengths in the virtual camera, i.e., to realize almost a parallel projection. This reduces the distortions caused by a deviating viewer position (Hagen and Elliot 1976).

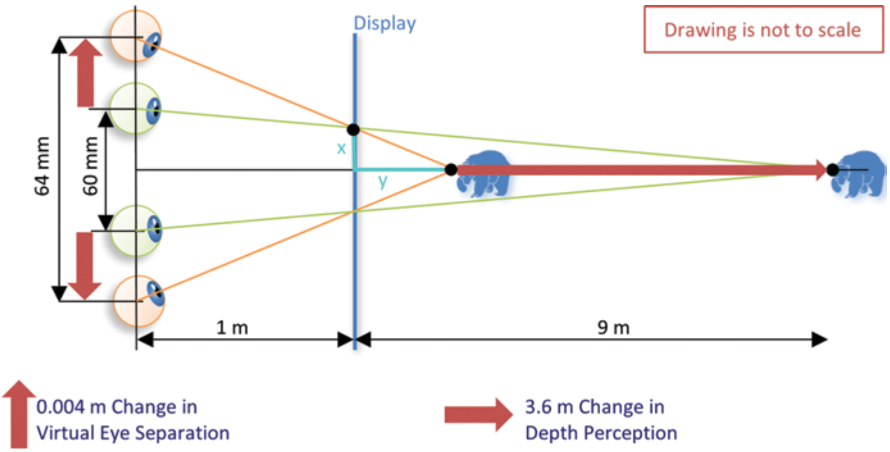
Display
Drawing is not to scale
64 mm
60 mm
x
y
1 m
9 m
0.004 m Change in Virtual Eye Separation
3.6 m Change in Depth Perception

Fig. 2.4 Geometric effect of changing the virtual eye separation (drawing is not to scale). The geometric effects also influence perception (Bruder et al. 2012a)

Stereo displays can cause additional deviation because the two virtual cameras that generate the image for the left and right eyes have a distance (called *virtual eye separation*) that may differ from the distance of the viewer's pupils. On average, the *pupil distance* is 64 mm, but the individual range is large and lies approximately in the interval from 45 mm to 75 mm. Figure 2.4 shows an example that small changes in pupil distance can result in large changes in depth perception. In this example, the pupil distance is initially 64 mm and the object shown on the projection surface appears to be 9 m behind the projection surface. If the distance between the eye points is reduced by 4 mm, it follows from the set of beams that the virtual object moves forward by 3.6 m. But as with deviations in the viewing position, deviations between virtual eye separation and pupil distance are compensated by adaptation in such a way that they do not irritate the viewer. In fact, the distance between the virtual cameras can be changed several times in 1 s without the viewer even realizing it. In VR, it is therefore not absolutely necessary to first measure the distance between the two eyes of the viewer and then adapt the distance between the two virtual cameras accordingly. However, side effects such as nausea (see Sect. 2.4.7) can occur, even if the user does not consciously notice the difference.

2.4.2 Double Vision

If the viewer of a stereo display is not able to fuse the two different images shown to the left and right eyes, *diplopia* occurs. This is a severe problem in VR, as it is perceived as extremely irritating and has a negative effect on the feeling of presence in VR. Thus, diplopia should be avoided at all costs.

The reason for diplopia has already been explained in Sect. 2.2.1: the point to be merged lies outside Panum's fusional area. Since accommodation always occurs to the display plane, the visual system tends to move Panum's fusional area near the display surface of the stereo display as well (see vergence-focus conflict, Sect. 2.4.4). This means that a stereo display cannot make objects appear arbitrarily far in front of or behind the display surface. So, if one wants to display a virtual world with the help of a stereo display, there is only a limited area available in which the virtual objects can be placed in front of or behind the display (*parallax budget*) without diplopia. Williams and Parrish (1990) state that −25% to +60% of the distance from the viewer to the display surface are the limits for the usable stereo range (in the case of an HMD, the virtual distance of the display is to be used). Here, Panum's fusional area has its thinnest extent in the area of the point that the eyes fixate on. In the worst case, it has only a width of 1/10 degree viewing angle. At a distance of 6° from the fixated point, Panum's fusional area increases in width. Then, it has a visual angle of about 1/3 degree. If a display is at typical monitor distance and has 30 pixels per cm, then points can only be arranged in a depth range of 3 pixels before diplopia occurs (Ware 2000). The situation is aggravated by the fact that the entire Panum's fusional area should not be used, since only in a partial area can fusion be achieved without effort even over longer periods of time. This partial area is called *Percival's zone of comfort* and it covers about one-third of Panum's fusional area (Hoffmann et al. 2008).

One strategy to avoid diplopia is to enlarge Panum's fusional area. The size of this area depends, among other things, on the size and richness of detail of the objects being viewed and on the speed of moving objects. By blurring the images to be fused, the amount of detail is reduced. This way, Panum's fusional area can be enlarged. Another strategy is to bring virtual objects closer to the display area and thus into Panum's fusional area. With virtual eye separation, we have already learned a technique for this. If one reduces the distance between the virtual cameras, objects meant to appear behind the display can be brought closer to the display surface. Since human perception is robust against this manipulation, changing the virtual eye separation is useful to avoid diplopia. Ware et al. (1998) propose the following formula: virtual eye separation $v = 2.5\ \text{cm} + 5\ \text{cm} \cdot (a\ /\ b)^2$, where a is the distance of the point in the scene closest to the viewer and b is the distance of the point furthest away. Another technique to bring the virtual world into Panum's fusional area is the *cyclopean scale* (Ware et al. 1998). Here, the whole scene is scaled by one point between the two virtual cameras (cf. Fig. 2.5). The cyclopean scale can be combined with the manipulation of virtual eye separation, where scaling should be performed first. Such scaling is not only useful to bring a virtual world that is too spatially extended into Panum's fusional area, but also in the opposite case: a virtual world that does not use the limited area around the stereo display can be made to appear more three-dimensional by extending it. In VR, it is useful to be clear about the available parallax budget and its use. In a stereo display, the parallax that can be displayed cannot be arbitrarily small. The lower limit is the width of one pixel.

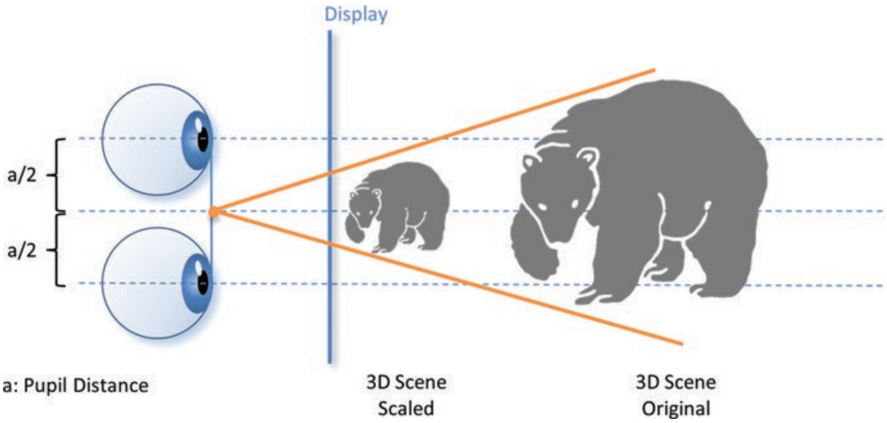

Fig. 2.5 Cyclopean scale

2.4.3 Frame Cancellation

The displays used for the presentation of virtual worlds usually have several imperfections, e.g., they cannot display the brightness levels found in reality, such as in sunlight. Also, the surface of the display is usually recognized as such and can be distracting. In particular, the edge of a display surface can be perceived as irritating. Let us assume we use a stereo display to make an object appear in front of the display plane. In case this object approaches the edge of the display and finally touches it. The following phenomenon can be observed. The illusion that the object is in front of the display is suddenly lost and the object snaps back to the level of the display. Moreover, diplopia can also be observed. This phenomenon is called *frame cancellation*, *paradoxical window* or *stereoscopic window violation* (Mendiburu 2009).

This phenomenon can be explained by the fact that the object has conflicting depth cues. According to the disparities, the object is in front of the display. However, the edge of the display seems to occlude the object, which suggests that it is behind the display. Occlusion is a stronger depth cue than disparity, which is why the object is perceived to be behind the display. Other explanation attempts point out that the object can only be seen by one eye when it is at the edge.

Keeping objects with negative parallax away from the edge or moving them quickly at the edge so that they are either completely visible or completely invisible on the image are simple strategies to avoid frame cancellation. Another strategy is to darken objects at the edge of the display and color the edge itself black so that the contrast between the edge and the object is small. Finally, black virtual stripes can be inserted in the depth of the object in the scene, thus seemingly bringing the display edge forward. The virtual stripes cover the virtual object when it approaches the display edge.

2.4.4 Vergence-Focus Conflict

In contrast to reality, some depth cues may be completely missing in VR, e.g., because the VR system's performance is not sufficient to calculate shadows in real time. Depth cues can also be wrong, e.g., the image blur might not be displayed correctly because it is difficult to determine the exact point the observer fixates on. While in reality the depth cues are consistent, they can be contradictory in VR, as the frame cancelation example shows. Contradictory depth cues not only have consequences such as a misjudgment of the spatial arrangement of objects in space or the loss of presence because the virtual world appears unnatural; other negative consequences can include eye stress, exhaustion and headaches. An example of this is the *vergence-focus conflict* (Mon-Willams and Wann 1998), also called *accommodation-convergence discrepancy* or *vergence-accommodation conflict*.

No matter whether a virtual world is viewed on a computer monitor, a projection or a head-mounted display (see Chap. 5), the viewers must adjust their eyes so that the display surface is seen sharply to easily perceive what is shown there. If a stereo display is used and an object appears in front of or behind the display surface due to disparity, the convergence is not set to the distance of the display surface but the apparent distance of the virtual object. Therefore, if the viewer wants to focus on a virtual object that appears to be in front of the display surface, the viewer must increase the convergence. As a result, however, the object suddenly appears unexpectedly blurred, as the eyes no longer focus on the display surface. This can also cause a contradiction between convergence and image blur. Convergence and focus information are therefore in conflict. As a result, headaches can occur. The risk of this increases with the duration of viewing of the virtual world (Hoffman et al. 2008).

The contradiction between the above depth cues can be reduced by bringing the virtual objects as close as possible to the display surface. For this purpose, the already discussed techniques, such as the cyclopean scale or the change of virtual eye separation can be used. These techniques can have side effects, such as falsification of depth perception. These side effects must be weighed against phenomena like fatigue or headache. There is no way to avoid the viewer's eyes converging on the display surface, as this is the only way to ensure that the image shown on the screen can be perceived sharply. The approach of subsequently introducing *depth of field* into the image (computer calculations of images allow the creation of images that are sharp everywhere – in contrast to real imaging systems such as a camera or the human eye) by blurring parts of the image and thus adapting the focus information to the convergence has not proven to be successful (Barsky and Kosloff 2008).

2.4.5 Discrepancies in the Perception of Space

In applications from the fields of architecture, CAD, urban visualization, training, simulation and medicine, three-dimensional spaces are presented. In these applications, it is essential that the users correctly perceive the virtually presented space, so that they can draw conclusions about their actions and decisions in the real world. Discrepancies between the perception of size and distance in the virtual and real worlds are particularly critical in this application context. For example, a physician simulating an operation in the virtual world should not train wrong movements due to misjudgments of space. The correct perception of sizes and distances is essential for many applications in the field of VR.

Unfortunately, many studies show that there can always be discrepancies in the perception of virtual space. For example, it has often been shown that users tend to underestimate distances in the virtual world by up to 50% (Interrante et al. 2006; Steinicke et al. 2010a). A common approach to measuring distance estimation is, for example, blind or imaginary walking. Here the user is shown a mark at a certain distance (e.g., 4 m, 6 m or 8 m) on the floor, and the user must then walk to this mark with eyes closed. In the real world, this task is easy to accomplish, and we walk almost exactly up to the mark. A user in the virtual world who sees the same scene (geometrically correct) on a head-mounted display, for example, will most likely walk much too short a distance; in some cases by up to 50%. This effect can be observed with many techniques for evaluating spatial perception (e.g., triangular completion, blind throwing, imaginary walking or verbal assessment). Many studies have shown the influence of some factors (such as stereoscopic imaging, limited field of view, realistic lighting or shading) on this distance underestimation, but up till now, there is no complete explanation for this phenomenon.

According to *Emmert's law*, there is a clear connection between sizes and distances. In this respect, the phenomenon of underestimating distances can also be observed as a phenomenon of overestimating sizes. The law states that the perceived size is proportional to the product of perceived distance with retinal size, i.e., the size of the image on the retina. The resulting law of *size constancy* is used by humans already in infancy. If, for example, a mother distances herself from her child, the projection of the mother on the retina of the child becomes smaller, but the child is aware that the mother is not shrinking, but merely moving further away. It is also the case that the more of the above-mentioned depth cues are missing, the more the angle of vision is used for size estimation. Misjudgments in the real world can also occur. These can be exploited in perspective illusions, for example. However, such misjudgments result not only from perceptual errors but also from cognitive processes. Distances are considered to be greater, for example, when subjects carry a heavy backpack (Proffitt et al. 2003) or are asked to throw a heavier ball (Witt et al. 2004). Thus, not only optical stimuli and their processing play a role in depth perception but also the intended actions and the associated effort. Furthermore, studies have shown that presence influences the perception of distances. The more present we feel in the virtual world, the better our assessments of distances become (Interrante et al. 2006). This illustrates that the correct assessment of space can be a complex task even in the real world, depending on perceptual, cognitive and motor processes.

Various approaches exist to improve the estimation of distances or sizes in the virtual world or to make the space presented or the objects displayed in it appear larger or smaller. For example, one could simply scale the entire geometry. Now the test persons would perceive the space as they would in the real world, but this does not solve the problem. Similar effects can be achieved by enlarging the *geometric field of view*. The geometric field of view refers to the area presented by the virtual scene, which is defined by the horizontal and vertical opening angle of the virtual camera. If this is enlarged, the viewer sees a larger area of the virtual world. However, since the same physical display is still used, this larger area must be mapped to the same area of the screen. Thus, the scene is minified, and objects appear further away (Kuhl et al. 2006). This is illustrated in Fig. 2.6. Similar effects can be achieved by changing the pupil distance. However, these approaches have the disadvantage that they actually present a different space utilizing, for example, perspective distortion. Subjects now continue to walk further, but they do so in another room that is projected with different geometric properties (see Fig. 2.6).

Alternative approaches are based on the idea of exaggerating the given depth cues to give the users clearer indications for the assessment of distances. For example, artificial shadows created by drawing lines to the base surface can give just as effective depth indications as stereoscopy. By using fog to desaturate the colors of distant objects, atmospheric depth can be imitated. This helps the user to better estimate distances, for example in virtual city models.

As already indicated above, cognitive factors also influence the assessment of space. It has been shown that the estimation of distances is significantly better in virtual space that is an exact representation of real space (Interrante et al. 2006). Follow-up studies have shown that this is not only due to the knowledge of real space, but especially to the higher sense of presence in such virtual worlds. This improved ability to assess distance can even be transferred to other virtual worlds.

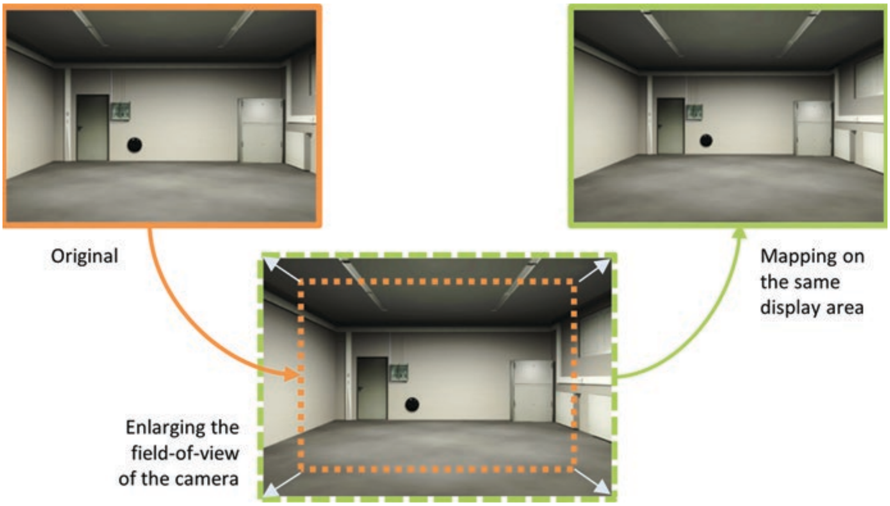
Original
Mapping on the same display area
Enlarging the field-of-view of the camera

Fig. 2.6 Presentation of the same virtual space with (left) small and (right) large geometric field of view. (According to Steinicke et al. 2009)

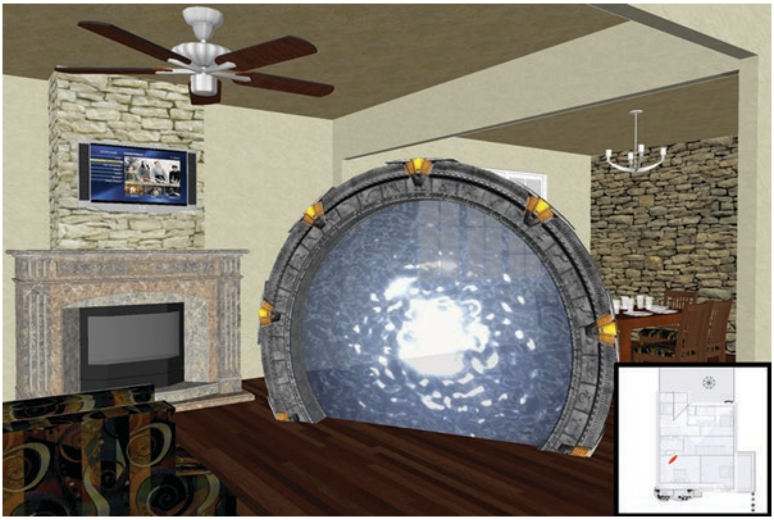

Fig. 2.7 Representation of a virtual portal through which users can travel to different virtual worlds. (According to Steinicke et al. 2010b)

For instance, a transfer can succeed if one is teleported from a virtual space exactly simulating real space to these other virtual worlds through a portal (see Fig. 2.7).

2.4.6 Discrepancies in the Perception of Movement

A similar effect as with distance underestimation can also be observed in the perception of movement, such that speeds of movement or distances covered are over- or underestimated. For example, many studies have shown that forward movements along the line of sight are underestimated (Lappe et al. 2007; Loomis and Knapp 2003). This is particularly true if the movement is only visually presented, and the user essentially perceives only the optical flow. Even if the user moves simultaneously and the movements are mapped 1:1 onto the virtual camera, this underestimation of forward movements along the line of vision occurs. In contrast to virtual straight-line movements, virtual rotations often lead to an overestimation (Steinicke et al. 2010a).

In principle, these discrepancies in the perception of movement can be resolved relatively easily by applying *gains* to the tracked movements. For example (t_x, t_y, t_z) is a measured vector that describes the head movement of a user from one frame to the next. By means of a gain g_T, this movement can now be scaled simply by $(g_T \cdot t_x, g_T \cdot t_y, g_T \cdot t_z)$. If $g_T = 1$ no scaling occurs; for $g_T > 1$ the motion becomes faster; and for $g_T < 1$ the motion becomes slower. Psychophysical studies have shown that, for example, forward movements must be slightly accelerated (approx. 5–15%) to be considered correct by users. In contrast, rotational speeds should be reduced slightly (by approximately 5–10%).

These manipulations now lead to the fact that the virtually represented movements are correctly perceived, i.e., the visually perceived movements match the vestibular-proprioceptive as well as the kinesthetic feedback. However, the users now actually perform different movements in the virtual and real environments, with the effect that, for example, certain distance estimation methods, such as counting steps, no longer work. More recent approaches by Bruder et al. (2012b) prevent such discrepancies between real and virtual movements by manipulating the optical flow. Such optical illusions only manipulate the perception of the movement but not the movement itself.

2.4.7 Cybersickness

Users of a VR/AR application may experience undesirable side effects: headaches, cold sweat, paleness, increased salivation, nausea and even vomiting, *ataxia* (disturbance of movement coordination), drowsiness, dizziness, fatigue, apathy (listlessness) or disorientation.

It is generally known that the use of IT systems is not free of health side effects. Just working at a screen can lead to headaches, for example, because the eyes are overstrained by focusing on one plane for a long time, or the visual system is stressed by flickering at low refresh rates or blurred images. These visual disturbances, known as *asthenopia* (eye strain), can also occur in VR/AR applications because they also use monitors. The symptoms can be more severe, e.g., because the displays in an HMD may be closer to the eyes or fusion may still be necessary for stereo vision. An early study (Stone 1993) concluded that 10 min of use of an HMD is as stressful for the visual system as sitting in front of a computer monitor for 8 h. The situation is worse for individuals who suffer from vision disorders and, for example, have problems with eye muscle coordination.

Side effects can also be expected when users are moving or being moved within an application, e.g., by means of a motion platform, or by simply walking. The syndrome of symptoms known as *seasickness* (more generally: *motion sickness*) has been known for a long time and has also been the subject of research. It is possible to characterize movements that cause seasickness – for example, it is known that low-frequency vibrations (which may also occur in VR installations) lead to seasickness. In flight simulators, which move an entire replica of a cockpit, it was observed early on that a significant proportion of pilots complain of feeling unwell (*simulator sickness*).

It is noteworthy that in VR/AR applications, the physiological symptoms mentioned at the beginning, which sometimes also occur in motion sickness or simulator sickness, can be observed even when the users are not moving at all. Just seeing images seems to cause discomfort. Therefore, a separate term has been coined: *cybersickness* (sometimes also called *VR sickness*). Cybersickness can occur not only during VR/AR use but also for some time afterward. Usually, the symptoms disappear by themselves. However, users may still be sensitized even after the symptoms have subsided, i.e., they may suffer from cybersickness more quickly if they repeatedly use VR/AR systems within a certain period.

The exact causes of cybersickness are not known today. Probably there is also no single cause, but it is a multifactorial syndrome. One theory often used to explain cybersickness and motion sickness is the *sensory conflict theory*: problems occur when sensory perceptions are inconsistent. If, for example, a passenger is below deck while heavy seas are moving the ship, the brain receives information via the vestibular sense that strong movements are present. In contrast, the visual sense suggests precisely the opposite when no movement is detected in the cabin. Treisman (1977) motivates the sensory conflict theory by means of evolution: in the past, such inconsistencies in sensory perception only occurred if one had eaten the wrong mushrooms – and it is a sensible protective mechanism to quickly get rid of the poisoned stomach contents. Although in motion sickness inconsistencies between the visual sense and the sense of balance are particularly important in explaining symptoms, in cybersickness inconsistencies within a sense (e.g., contradictory depth cues in the visual sense, as in the vergence-focus conflict) are also considered, or even inconsistencies between the expected sensory impressions of a user and what is actually perceived. However, the sensory conflict theory cannot explain all phenomena in the area of cybersickness, and in particular, the extent to which symptoms occur can only be predicted with difficulty. Other attempts at explanation are therefore being sought. For instance, the *postural instability theory* (Riccio and Stoffregen 1991) assumes that people cannot cope with unfamiliar situations (such as those that can occur in a virtual environment) and that there is a disruption in the control of body posture that causes further symptoms.

Even though cybersickness's exact causes cannot be explained, factors have been identified that promote cybersickness's occurrence. The first group of factors depends on the individual. Age, gender, ethnicity and also individual previous experiences with VR and AR can influence the occurrence of cybersickness. Remarkable are significant individual differences in the susceptibility to cybersickness. People who frequently suffer from motion sickness are also more susceptible to cybersickness. The second group of factors is related to the VR/AR system. Influencing factors include image contrast and associated flicker, refresh rate, tracking errors, quality of system calibration and use of stereo displays. The larger the field of view (and the more peripheral vision is involved), the more frequently the occurrence of cybersickness is observed. Other essential factors are latencies, e.g., the time offset between head movement, the new head position's detection, and the correct image display of this new head position. A rule of thumb says that latencies above 40 ms are too high and that latencies below 20 ms should be aimed for. Finally, there is a third group of factors that are related to the application. Does the user spend a long time in the application? Does the user have to move the head frequently? Does the user rotate, perhaps even more than one axis at a time? Is the head tilted off the axis around which the user is rotated (*Coriolis stimulation*)? Is the user standing instead of sitting or lying down? Do users look directly down at the area in front of their feet and cannot see far in the scene in general? Is it difficult to orientate in the scene, e.g., because a static frame of reference is missing? Is there much visual flow? Do users move quickly and a lot in a virtual world? Are there frequent changes in speed, are movements oscillating rather than linear, and are there abrupt movements? Does the user jump often or climb stairs? Are there unusual movements? Are users anxious? The more questions are answered in the affirmative and the more emphatic the agreement, the more cybersickness can be expected. Another factor is the degree of control (combined with the anticipation of movement) that a user has when navigating through a virtual environment. This is consistent with the phenomenon that the driver of a car or the helmsman of a ship suffers less often from motion sickness. Finally, a further factor is whether the application favors *vection*, i.e., the illusion of moving even though no movement is taking place.

If one wants to reduce the risk of cybersickness, one can minimize the influence of the factors mentioned, such as reducing latencies by improving the technical realization, reducing movements of the user by increased use of teleportation, or by inserting artificial blurring during the rotation of the user. Individually, one can avoid the occurrence of cybersickness by slowly getting used to VR/AR applications (McCauley and Sharkey 1992). Chewing gum and adequate fluid intake are recommended. In extreme cases, one can take medication against motion sickness. As a herbal remedy, ginger does not prevent cybersickness, but it does counteract nausea and vomiting. Ultimately, it must be accepted that the occurrence of cybersickness cannot be prevented with certainty. Consequently, users should be given an easy way to terminate a VR/AR application at any time. It is also important to inform users about the possible side effects and to obtain the explicit consent of users, especially in user tests.

Whether and to what extent cybersickness occurs is usually determined by observing or asking users. For this purpose, it makes sense to use standardized questionnaires. Although not intended for cybersickness, the *Simulator Sickness Questionnaire* (*SSQ*) and the *Motion Sickness Assessment Questionnaire* (*MSAQ*) are often used (Kennedy et al. 1993). Alternatively, users can be watched to detect symptoms – but this is sometimes difficult, e.g., headaches are difficult to detect, but vomiting is easy. Physiological body values (e.g., heart rate, skin conductivity) are sometimes measured. Here, especially, the interpretation of the measured values is difficult. Based on such measurements, studies such as Lawson (2015) conclude that 60–80% of users of a VR application show symptoms of cybersickness. Around 15% show symptoms so severe that they have to stop using the application. However, such figures should be applied with great caution to a specific VR/AR application – there are many possible influencing factors and, therefore, strong fluctuations in the values. Individual differences among users are also considerable; the same user can react very differently to a scenario repeated several times during each repetition. Nevertheless, these figures show that cybersickness is not a marginal problem, but a real barrier to the use of VR and AR. Consequently, cybersickness should be taken into account in the development of every VR/AR application.

2.4.8 Vertical Parallax Problem

One problem with the technical implementation of stereo vision is that the virtual projection plane used in rendering cannot be brought into alignment with the display's real plane if the two are not parallel to each other. This leads to *vertical parallax,* which the viewer perceives as a strain and can lead to errors in depth perception, blurring at specific image points or double images. Let us look at Fig. 2.8a. An observer fixates on point P, and thus the eyes are aligned accordingly – the directions of gaze are no longer parallel and convergence occurs. If we reproduce this when rendering the images, i.e., if we apply the *toe-in method*, the two projection planes intersect at point P and are not parallel to each other. Most of the time, it is technically not possible to realize that, for each of the two projection planes, there is a separate display available that can be aligned accordingly. Instead, a common real display is used for both projection planes. The point *A* has the distance *v* from the display. This is the unwanted vertical parallax. The further point *A* is from point *P*, the greater the vertical parallax, and the more blurred or distorted the image appears. As with horizontal parallax, you can distinguish between negative parallax (located before the display plane, such as point *A*) and positive parallax (located behind the display plane, such as point *B*).

Because of the problem of vertical parallax, the toe-in method is avoided, and the *off-axis method* is used instead. This is shown in Fig. 2.8b. Here, each eye has a fictitious point of view P′ or P″, so that both projection planes lie on top of each other. This means that both projection planes can also be mapped exactly onto a single display plane. As a result, the viewing volumes are no longer symmetrical. Accordingly, an *asymmetrical viewing volume* must be set during rendering. This is

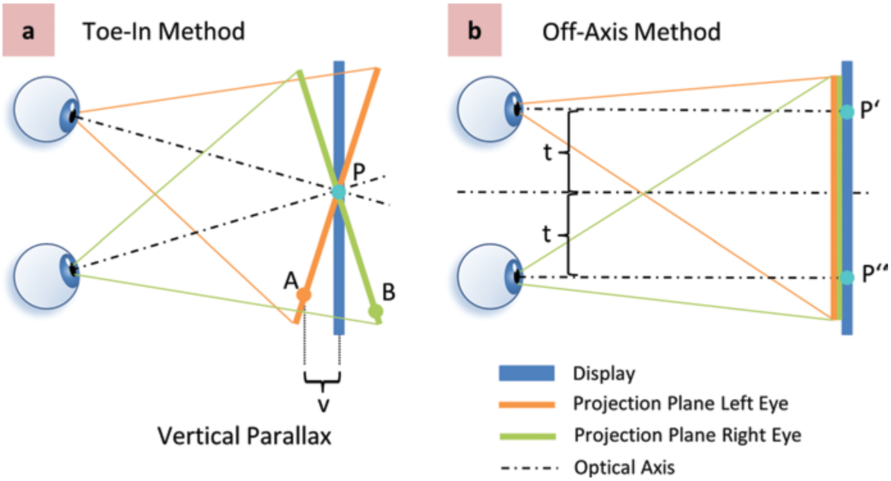

Fig. 2.8 (**a**) The toe-in method leads to the occurrence of vertical parallax. (**b**) The off-axis method solves this problem

shifted by the distance *t* from the center axis ('off-axis'). The exact size of the view volumes can be calculated through a set of rays if the distance between the projection plane and the eyepoint is known. This solves the problem of vertical parallax.

2.5 Use of Perceptual Aspects

With knowledge of human perception, we can not only explain problems occurring in VR. Knowledge about the operation of human perception can also be useful to improve a VR experience or to use available resources well. In Sect. 2.4.1 we have already seen an example of how the ability of the human visual system to adapt makes complex technical solutions superfluous: we do not have to measure the distance between the pupils of an observer to adjust the virtual cameras correctly. On the contrary, we can manipulate virtual eye separation to prevent diplopia because we know that human perception reacts robustly to changes in virtual eye separation. Besides adaptation, there are two other important perceptual aspects of VR that are exploited in VR: salience and user guidance.

2.5.1 Salience

Human perception does not have the capacity to process all environmental stimuli in equal detail. Priorities are set, and people can focus *attention* on certain aspects. In the human visual system, for example, differentiation is already inherently built-in through the uneven distribution of sensory cells on the retina of the eye – humans can align the fovea in such a way that light stimuli from environmental objects classified as particularly relevant hit this point in the retina, which possesses a high number of sensory cells.

VR makes use of this characteristic of human perception because VR systems often do not have the capacity to artificially generate all environmental stimuli equally well. If you know what the user of a VR system is focusing his or her attention on, you can adjust the quality of the rendering (e.g., simulation of surface materials, quality of the object models, effort invested in anti-aliasing), sound quality, quality of the animation or accuracy of the world simulation. Conversely, one does not need to invest any or only a few resources of a VR system in areas that are not the focus of attention. In extreme cases one can even observe *inattentional blindness*. In an experiment, Simons and Chabris (1999) showed nearly 200 students 75 s long videos in which basketball players throw a ball at each other. The viewers had the task of counting how many passes a team makes – attention was thus focused on the ball. The video showed an unusual event for 5 s, e.g., a person dressed as a gorilla walking across the field. About half of the viewers did not notice this at all. So why go to the trouble of creating images of a gorilla in a VR version of this scene if the viewer does not notice it?

There are two obstacles to exploiting these phenomena of human perception. On the one hand, while it is possible to make statements about probabilities, it is not possible to predict with certainty which environmental stimuli are considered important for an individual in a concrete situation. Hence, we could make mistakes. For example, we leave out the gorilla in our VR scene even though the viewer would have seen it in the concrete situation. Here it is essential to weigh up the likelihood of making a mistake and the consequences. Due to the limited performance of VR systems, one may have no choice but to set priorities to meet real-time requirements. Violating real-time conditions (e.g., the virtual environment reacts with a noticeable delay to a user's action; see Chap. 7) can have more serious consequences than choosing the wrong priorities.

On the other hand, there is the issue that the information is needed on which the viewer's attention is currently focused. There are different approaches to obtaining this information. Firstly, technical systems can be used to determine where the observer is currently looking (eye-tracking; see Chap. 4). Secondly, through knowledge about the application and the current goals and tasks of the user of VR, it can be estimated which objects of the virtual world are likely to attract a high level of attention (Cater et al. 2003). In the gorilla example, we could deduce from the task given to the viewers that the ball is the center of attention. Myszkowski (2002) creates *task maps* that assign each object a priority for rendering, with moving objects automatically getting a higher priority. A third approach (Treisman and Gelade 1980) is based on the *feature integration theory*. This approach is attractive for VR because it does not require any additional knowledge about the application or the viewing direction of the viewer but can work solely on the images of the 3D scene: the *salience* (also called *saliency*) of objects is determined as a measure of their importance.

Salience describes how strongly an object stands out from its surroundings (e.g., in color, orientation, movement, depth). If one shows a person a picture with 50 squares of equal size, 47 of which are grey and 3 are red, the 3 red squares stand out and are immediately noticed. The person can easily and quickly answer the question of how many red squares can be seen in the picture. Even if the number of gray squares is quintupled, the person can just as quickly recognize that there are 3 red squares present. The feature integration theory explains this observation by postulating that human perception works stepwise. In the first stage, all incoming image stimuli are processed in parallel and examined for specific features. This happens subconsciously. It is called *preattentive processing* (see Fig. 2.9). Anatomically, receptive fields have already been identified, i.e., groups of nerve cells in the brain that are responsible for these tasks of feature extraction. The result of preattentive processing then serves as the basis for the decision in the next stage as to which regions in the image are to receive attention.

If one wants to take advantage of this in VR, one must first calculate an *attention map* (*saliency map*) of an image in which every pixel of an image is assigned a salience value. Today's algorithms for this purpose are based on the work of Itti et al. (1998). The procedure consists of first splitting the input image into feature images, e.g., extracting a luminance image that contains only brightness values.

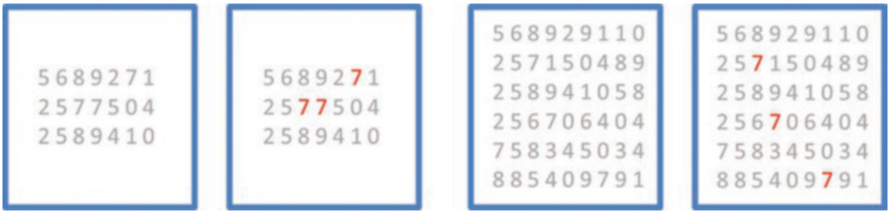

Fig. 2.9 Example of preattentive processing: the time required to find the number of occurrences of the digit '7' in a series of numbers can be reduced considerably if the number '7' is displayed in a different color. This is processed in a preattentive stage. If the number series size increases, the time for the task completion increases if the number '7' is not highlighted; otherwise it remains the same

These feature images are examined in parallel with image processing methods, whereby the operation of the receptive fields in the brain is modeled mathematically. Receptive fields that recognize orientation in a feature image can be described, for example, by *Gabor filters.* A Gabor filter is constructed from a Gaussian function modulated by a sinusoidal function and can thus map the sensitivity for different frequencies and orientations. The results of processing the individual feature images are normalized. The salience values are determined from this by weighted summation. The weighting can also be chosen depending on the current task of the observer. It is often determined by machine learning, e.g., utilizing neural networks. In this processing step, another phenomenon of human perception can be mimicked: *inhibition.* Inhibition means that nerve cells can not only be stimulated but also inhibited by stimuli, which increases differences. Algorithmically, this can be realized, for example, with a *winner-takes-all approach*, i.e., the greatest value is used for salience, while salience in the vicinity of the greatest value is reduced to enhance its significance further. The saliency map finally obtained then serves as a basis for decisions on how to use resources of the VR system, e.g., for areas with high salience 3D models with a high level of detail are used. Further data can also be obtained, e.g., *fixation maps* (Le Meur et al. 2006), which predict what an observer is likely to fixate on. Since saliency maps are two-dimensional, a relatively complex back-calculation into the 3D scene is necessary to assign a salience value to virtual 3D objects. Therefore, approaches are also being considered that directly examine characteristics of 3D objects and derive a salience value from them (Lee et al. 2005).

2.5.2 User Guidance

The area covered by the virtual environment's hardware platform in which users can move around is usually much smaller than the virtual world represented in it. Clearly, without additional input devices, the users can only explore a very small part of the virtual world by their own movements. There is a variety of so-called locomotion devices that prevent the user from moving from one place to another in the real world while walking. Examples are omnidirectional treadmills or the *Virtuix Omni* (see Chap. 4). Another approach is based on the idea of manipulating users in such a way that they walk on different paths in the real world than those perceived in the virtual world. If, for example, a small virtual rotation to one side is introduced during a user's forward movement, the user has to compensate for this rotation in the real world to be able to continue walking virtually straight ahead. This results in the user walking on a curved path in the opposite direction. Thus, users can be guided on a circular path in the VR setup while they think they are walking straight ahead in the virtual world. Investigations have shown whether and from when on test persons can detect such manipulations through *re-directed walking* (Steinicke et al. 2010a). For instance, test persons who walk straight ahead in the virtual world can be guided on a circle with a radius of about 20 m in the real world without noticing this.

2.6 Summary and Questions

In this chapter, you have acquired basic knowledge in the field of human information processing. In particular, we have dealt with some of the most important aspects in the field of spatial perception and the perception of movement. Based on this, you have learned about relevant phenomena and problems of VR. You have also seen examples of how different limitations of human perception can be exploited to improve the quality and user experience during a VR session. To design effective virtual worlds, it is essential to take findings from perceptual psychology on human information processing into account. Aspects related to perception have become increasingly important in recent years, which is reflected in the increased number of research projects in this field.

Check your understanding of the chapter by answering the following questions:

- Why is the response time for a subject longer when deciding whether a stimulus displayed on the screen matches a previously displayed stimulus than when the subject only has to respond when the stimulus appears?
- Compare a photo of a beach in the Caribbean and a photo of the streets of Manhattan. What pictorial depth cues are present in the photos?
- How does the object in Fig. 2.4 move if the virtual eye separation is not reduced from 64 mm to 60 mm, but instead increases to 70 mm?
- Why should a cyclopean scale be performed before virtual eye separation?
- Take a stereo display and conduct experiments to determine Panum's fusional area of the stereo display. Try using the techniques presented in Sect. 2.4 to fit a 3D scene that initially protrudes over the panorama area.
- Find more examples of conflicting depth cues in VR.
- You would like to build a light rail simulator with which a learner can drive a streetcar through a virtual city. Think about where perceptual aspects need to be considered. Which problems can potentially arise? Where can perceptual aspects be exploited in the technical realization of the simulator?

Recommended Reading[1]

Goldstein EB (2016) *Sensation and Perception* (10th edn). Cengage Learning, Belmont – *Standard work from the psychology of perception which is not limited to visual perception. Very informative and with many examples*.

Thompson WB, Fleming WF, Creem-Regehr SH, Stefanucci JK (2011) *Visual Perception from a Computer Graphics Perspective*. CRC Press, Boca Raton – *Textbook which also explains essential aspects of perception for VR and always makes the connection to computer graphics*.

References

Barsky BA, Kosloff TJ (2008) Algorithms for rendering depth of field effects in computer graphics. In: Proceedings of 12th WSEAS international conference on computers, pp 999–1010

Bruder G, Pusch A, Steinicke F (2012a) Analyzing effects of geometric rendering parameters on size and distance estimation in on-axis stereographic. In: Proceedings of ACM Symposium on Applied Perception (SAP 12), pp 111–118

Bruder G, Steinicke F, Wieland P, Lappe M (2012b) Tuning self-motion perception in virtual reality with visual illusions. IEEE Trans Vis Comput Graph 18(7):1068–1078

Card SK, Moran TP, Newell A (1986a) The model human processor: an engineering model of human performance. In: Handbook of perception and human performance. Vol. 2: cognitive processes and performance, pp 1–35

Card SK, Moran TP, Newell A (1986b) The psychology of human–computer interaction. CRC Press

Cater K, Chalmers A, Ward G (2003) Detail to attention: exploiting visual tasks for visual rendering. In: Proceedings of Eurographics workshop on rendering, pp 270–280

Ernst MO (2008) Multisensory integration: a late bloomer. Curr Biol 18(12):R519–R521

Hagen MA, Elliott HB (1976) An investigation of the relationship between viewing conditions and preference for true and modified perspective with adults. J Exp Psychol Hum Percept Perform 5:479–490

Hayward V, Astley OR, Cruz-Hernandez M, Grant D, La-Torre GR-D (2004) Haptic interfaces and devices. Sens Rev 24(1):16–29

Hendrix C, Barfield W (1996) Presence within virtual environments as a function of visual display parameters. Presence Teleop Virt 5(3):274–289

Hoffmann DM, Girshick AR, Akeley K, Banks MS (2008) Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. J Vis 8(3):1–30

Howard IP (2002) Seeing in depth: Vol. 1. Basic mechanisms. I Porteous, Toronto

Interrante V, Anderson L, Ries B (2006) Distance perception in immersive virtual environments, revisited. In: Proceedings of IEEE virtual reality 2006, pp 3–10

Itti L, Koch C, Niebur E (1998) A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20:1254–1259

Kennedy RS, Lane NE, Berbaum KS, Lilienthal GS (1993) Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. International Journal of Aviation Psychology 3(3):203–220

[1] The ACM Symposium on Applied Perception (SAP) as well as the journal *Transaction on Applied Perception* (*TAP*) deal with multisensory perception in virtual worlds.

Kubovy M (1986) The psychology of linear perspective and renaissance art. Cambridge University Press, Cambridge

Kuhl SA, Thompson WB, Creem-Regehr SH (2006) Minification influences spatial judgement in immersive virtual environments. In: Symposium on applied perception in graphics and visualization, pp 15–19

Lappe M, Jenkin M, Harris LR (2007) Travel distance estimation from visual motion by leaky path integration. Exp Brain Res 180:35–48

Lawson B (2015) Motion sickness symptomatology and origins. In: Hale KS, Stanney KM (eds) Handbook of virtual environments: design, implementation, and applications. CRC Press, pp 532–587

Le Meur O, Le Callet P, Barba D, Thoreau D (2006) A coherent computational approach to model the bottom-up visual attention. IEEE Trans Pattern Anal Mach Intell 28(5):802–817

Lee CH, Varshney A, Jacobs DW (2005) Mesh saliency. In: Proceedings of SIGGRAPH 2005, pp 659–666

Loomis JM, Knapp JM (2003) Visual perception of egocentric distance in real and virtual environments. In: Hettinger LJ, Haas MW (eds) Virtual and adaptive environments. Erlbaum, Mahwah

Malaka R, Butz A, Hußmann H (2009) Media informatics – an introduction. Pearson, Munich

Marr D (1982) Vision: a computational investigation into the human representation and processing of visual information. MIT Press, Cambridge

McCauley ME, Sharkey TJ (1992) Cybersickness: perception of self-motion in virtual environments. Presence Teleop Virt 1(3):311–318

Mendiburu B (2009) 3D movie making: stereoscopic digital cinema from script to screen. Focal Press, New York

Mon-Williams M, Wann JP (1998) Binocular virtual reality displays: when problems do and don't occur. Hum Factors 40(1):42–49

Myszkowski K (2002) Perception-based global illumination, rendering and animation techniques. In: Spring conference on computer graphics, pp 13–24

Ooi TL, Wu B, He ZJ (2001) Distance determination by the angular declination below the horizon. Nature 414:197–200

Preim B, Dachselt R (2015) Interaktive Systeme (Band 2). Springer Vieweg, Berlin, Heidelberg

Proffitt DR, Stefanucci J, Banton T, Epstein W (2003) The role of effort in distance perception. Psychol Sci 14:106–112

Riccio GE, Stoffregen TA (1991) An ecological theory of motion sickness and postural instability. Ecol Psychol 3(3):195–240

Sharp H, Preece J, Rogers Y (2019) Interaction design: beyond human–computer interaction. Wiley, Indianapolis

Shneiderman B, Plaisant C, Cohen M, Jacobs S, Elmqvist N, Diakopoulos N (2018) Designing the user interface – strategies for effective human–computer interaction. Pearson Education Ltd, Harlow

Simons DJ, Chabris CF (1999) Gorillas in our midst: sustained inattentional blindness for dynamic events. Perception 28(9):1059–1074

Slater M, Usoh M, Steed A (1994) Depth of presence in virtual environments. Presence Teleop Virt 3:130–144

Steinicke F, Bruder G, Kuhl S, Willemsen P, Lappe M, Hinrichs KH (2009) Judgment of natural perspective projections in head-mounted display environments. In: Proceedings of VRST 2009, pp 35–42

Steinicke F, Bruder G, Jerald J, Frenz H, Lappe M (2010a) Estimation of detection thresholds for redirected walking techniques. IEEE Trans Vis Comput Graph 16(1):17–27

Steinicke F, Bruder G, Hinrichs KH, Steed A (2010b) Gradual transitions and their effects on presence and distance estimation. Comput Graph 34(1):26–33

Stone B (1993) Concerns raised about eye strain in VR systems. Real-Time Graph 2(4):1–13

Treisman M (1977) Motion sickness: an evolutionary hypothesis. Science 197:493–495

Treisman AM, Gelade G (1980) A feature integration theory of attention. Cogn Psychol 12(1):97–136

Vishwanath D, Girshick AR, Banks MS (2005) Why pictures look right when viewed from the wrong place. Nat Neurosci 8(10):1401–1410

Wanger LR, Ferwander JA, Greenberg DA (1992) Perceiving spatial relationships in computer-generated images. IEEE Comput Graph Appl 12(3):44–58

Ware C (2000) Information visualization – perception for design. Morgan Kaufmann, San Francisco

Ware C, Gobrecht C, Paton M (1998) Dynamic adjustment of stereo display parameters. IEEE Trans Syst Man Cybern 28(1):56–65

Williams SP, Parrish RV (1990) New computational control techniques and increased understanding for 3-D displays. In: Proceedings of SPIE Stereoscopic Display Applications, pp 73–82

Witmer BG, Singer MJ (1998) Measuring presence in virtual environments: a presence questionnaire. Presence: Teleoperators Virtual Environ 7(3):225–240

Witt JK, Proffitt DR, Epstein W (2004) Perceiving distance: a role of effort and intent. Perception 33:577–590