Int. J. Human–Computer Studies (1996) **44,** 829–847

What does virtual reality NEED?: human factors issues in the design of three-dimensional computer environments

JOHN WANN AND MARK MON-WILLIAMS

Department of Psychology, University of Reading, 3 Earley Gate, Whiteknights, Reading, RG6 6AL, UK

Virtual reality (VR) has invaded the public's awareness through a series of media articles that have promoted it as a new and exciting form of computer interaction. We discuss the extent to which VR may be a useful tool in visualization and attempt to disambiguate the use of VR as a general descriptor for any three-dimensional computer presentation. The argument is presented that, to warrant the use of the term virtual environment (VE), the display should satisfy criteria that arise from the nature of human spatial perception. It directly follows, therefore, that perceptual criteria are the foundations of an effective VE display. We address the task of making a VE system easy to navigate, traverse and engage, by examining the ways in which three-dimensional perception and perception of motion may be supported, and consider the potential conflict that may arise between depth cues. We propose that the design of VE systems must centre on the perceptual-motor capabilities of the user, in the context of the task to be undertaken, and establish what is *essential, desirable* and *optimal* in order to maximize the task gains, while minimizing the learning required to operate within three-dimensional interactive displays.

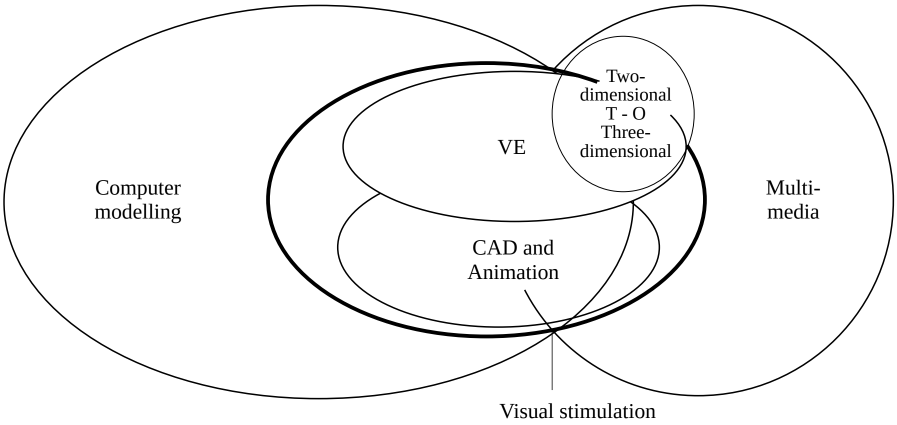
1. Who needs virtual reality?

As a starting point for considering the design of computer generated three-dimensional interactive environments (virtual reality: VR) it is perhaps necessary to ask why we need VR at all? Virtual reality has been the focus of a large number of media reports that have portrayed the more speculative aspects of the technology. This has reinforced a degree of scepticism amongst the scientific community as to the potential role of VR in science. Where serious applications of three-dimensional displays or VR have been demonstrated, such as in molecular modelling (Tsernoglou, Pesko, McQueen & Hermans, 1977; Ming, Pique, Hughes & Brooks, 1988), the technology employed in these prototypes was of a cost that placed it well beyond the reaches of most of the scientific and academic community. By the close of 1995, however, high-end desktop machines, enhanced with graphic accelerators were delivering three-dimensional graphic performance at 10% of the capital cost that would have been incurred in 1990. Mid-range graphics workstations are being marketed with three-dimensional interactive graphics as a central feature (e.g. Silicon Graphics Impact). The cost of bundled VR systems has moved within the constraints of many academic or health sector budgets (e.g. the IBM Elysium). Finally, the computer games market has reduced the cost of VR displays and peripherals to a fraction of their cost prior to 1993. As a result VR peripherals and three-dimensional graphics performance equivalent to that employed in high profile VR demonstrators in 1990 are now available to most of the scientific community.

1071-5819/96/060829 + 19$18.00/0

The question of "who really needs VR?" is not dissimilar to the question of "will everyone really use a desktop computer?" that echoed through many finance sub-committees in the 1980s. Prior to 1980 it was difficult to envisage how widespread computer use might be, particularly in occupational areas that did not hinge on "computation". Similarly, at present it is difficult to conceive of how interactive three-dimensional visualization might be used extensively in, for instance, the social sciences. To consider the issue from a broad perspective: a rapidly increasing proportion of the western population have access to computers and use them for occupational tasks. User interfaces have been introduced that allow many routine tasks to be achieved through direct-manipulation. It is no longer essential for the user to recall specific command strings or acquire semantic associations (e.g. chdir, ls -l, chmod). Windows-based interfaces present the user with icons distributed across a two-dimensional desktop, with interaction available through mouse actions such as drag-and-drop or pull down menus. There are, however, restrictions on the complexity of structure that can be presented on a two-dimensional desktop. It is commonplace to use multiple windows to display information subsets, both within and between applications. Where cross-links between respective displays are required, however, the representation is inevitably subject to two-dimensional constraints. In many cases a two-dimensional depiction of a data structure will be perfectly adequate. As a greater variety of information becomes available, however, there are applications that can benefit from breaking the bounds of two-dimensional representation. Images imported from medical scanning procedures, such as magnetic resonance imaging, can be displayed section by section, but there are additional benefits from three-dimensional reconstruction. The ability to change the user's viewpoint and inspect the size and extent of abnormal features may be crucial for effective prognosis. The use of VR can also extend beyond the representation of natural structures. Complex data-sets may also require representation outside of that available through two-dimensional depiction. Genotography software (e.g. Perkin-Elmer) uses DNA fragment data to check Mendelian inheritance and map linkages contributing to genetic disease status. In this type of application, the potential transmission links that need to be represented are likely to reach a level of complexity where their depiction will benefit from being extended across a third dimension. The use of a three-dimensional data-space, however, also necessitates the ability to move through that structure. A two-dimensional flow chart can be adequately viewed from a single viewpoint or by scrolling the display horizontally or vertically (Figure 1 upper). Levels of detail, however, are hidden within a three-dimensional structure unless the viewpoint can be changed interactively. In the example of genetic transmission, the inspection of 3rd or 4th generation influences is likely to require that *x,y* coordinate motion is supplemented with motion in depth (z). Hence a subset of computer visualization tasks may benefit from extending the representation into three-dimensions and allowing the users to traverse the structure. The areas of application where this might apply can be roughly outlined as follows.

(i) Direct three-dimensional reconstruction to display measurements of natural three-dimensional structures, e.g. non-invasive inspection of animate and inanimate structures; medical training.



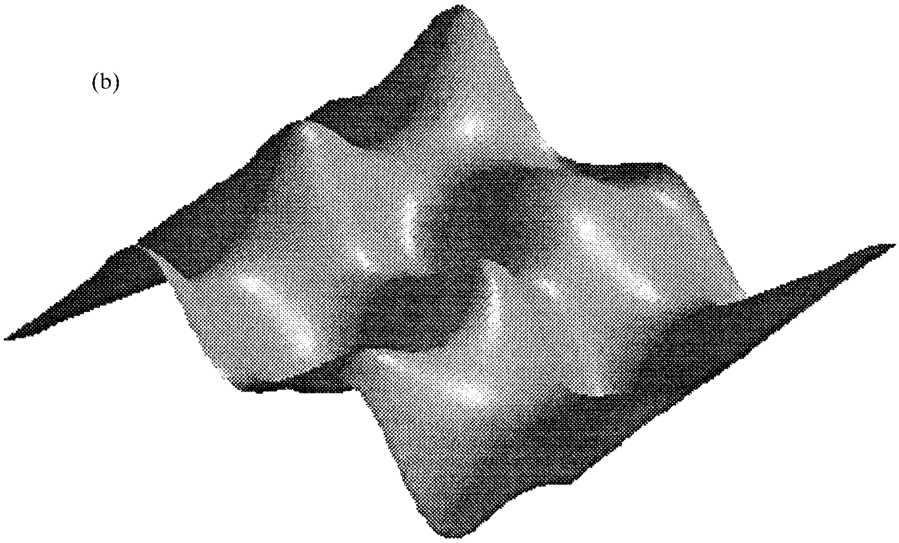


FIGURE 1. (a) Example of a flow diagram illustrating a computer simulation presented dimensions. Outputs from the simulation may be two-dimensional or three-dimensional graphics. (b) Example of three-dimensional output from a computer simulation in the form of a surface plot.

(ii) Simulation of very large or very small structures to enable a change in level of detail, e.g. molecular models; geographic/geological models.

(iii) Simulation of environments yet to be developed, e.g. architectural visualization; manufacturing design; automotive/cockpit console design.

(iv) Simulation of environmental contingencies, e.g. hazardous environment training; flight/driving simulation.

(v) Structured display of complex data-sets, e.g. communication networks with accessible levels of detail (LAN, MAN, PIPEX, EBS); stock/futures market status and dependencies.

(vi) Games and entertainment.

Applications (i)–(iv) are simulating data/features that are, in their most natural form, three-dimensional. The extension of complex data sets in three-dimensional space (v) provides the potential to display data in a form that may aid interpretation. A suitable example is representing the growth of oral bacteria of different subtypes (e.g. three-dimensional column plot) or within different oral regions (three-dimensional simulated mouth), where a critical parameter is acidicity (global colour coding of regions) and where bacterial growth will change over time. Hence we have a five-dimensional model from three spatial dimensions, supplemented with colour coding and temporal fluctuations. The model could be extended further if, for example, the scientist wanted to represent potential flaws in tooth enamel. A natural choice for this might be additional colour-coding (hot-spots), but the general status of enamel structures could be represented in the non-visual domain by three-dimensional sound. Hence, the information load may be distributed across perceptual systems, and the biochemist can explore the simulated oral environment, looking for bacterial clusters, "hot-spots" and pH fluctuations, but also hear the general deterioration of surface enamel as individual molars begin to "hum with discontent." Breaking out beyond the dimensions of the two-dimensional desktop allows the depiction of complex data in an innovative and appealing way. In this respect virtual environment systems can provide a further avenue to explore means of presentation in line with the agenda that "*The purpose of computing is insight not numbers*" (Hamming, 1962). The task for the production of efficient VR applications is to base the systems around design principles that enable intuitive and efficient interaction with three-dimensional structures.

2. What is a virtual environment?

Simply presenting data in a three-dimensional format does not, in itself, render the display as a "virtual reality". It is unfortunate that, for various reasons, some researchers, technology providers and publishers readily attach the label VR to any application that employs three-dimensional depiction. Computer aided design (CAD) has routinely used three-dimensional perspective views to depict potential structure, but does CAD have to be relabelled as VR? Some cases are clearly on the borderline: In a discussion of neural network models of locomotion, Grillner (1996) presents a series of ray-traced images of a simulated lamprey swimming in a manner determined by the network model. Either the author, or the publisher titled the section "Virtual Reality". Does a three-dimensional pictorial simulation constitute "virtual reality" and, if so, then how does VR differ from conventional cinema animation? To discuss potential definitions of systems it is essential to begin by rejecting the use of the term virtual "reality":

2.1. WHY REALITY?

In the case of the virtual lamprey cited above, it could be argued that the term VR was justified because the pictures displayed what "in reality" the model fish would be doing. The term computer animation, however, would seem to be less ambiguous. When we consider systems with which the user interacts, it is clear that current VR systems fall well short of providing coherent, high fidelity visual, auditory and haptic stimuli that are likely to fool the user into believing that the environment is ‘‘real’’. There are constrained cases where this can be achieved: flight simulation combines relatively simple distal visual stimuli with real, proximal stimuli and coherent physical motion to produce a believable training environment, likewise the passive passenger in an entertainment simulation may exhibit an elevated heart rate and wince when crashing into a simulated meteor cloud. But the delimits placed upon the level of simulation are essential to maintain belief in such settings. Irrespective of any continued escalation in computing power during the next decade, it is difficult to envisage a simulation that will support every contingency that we are afforded in our natural environment. In a VR simulation of a planned conference facility, for instance, will one be able, at a whim, to pluck a virtual satsuma from a virtual bowl on the virtual table; feel its stippled skin and smell its zest; sense the skin softly yielding as it is peeled and enjoy the sweet, soft texture in one’s mouth? How could such contingencies be catered for through digital simulation, given our impoverished understanding of the subtleties of human perception and sensation? Furthermore, why would a designer ever want to simulate an environment to this degree? Simulation of three-dimensional structures is a tool for the transmission of information. In all cases there should be a clear goal in terms of the information that needs to be supplied and the delimits that can be placed on that knowledge. The only drive towards unconstrained realism is likely to come from the entertainment industry and that is likely to be constrained by a compromise between the cost of single user interactive systems, and multi-user passive experience theatres. In summary virtual ‘‘reality’’ is an oxymoron that is misleading and unnecessary.

2.2. VIRTUAL ENVIRONMENTS

An environment provides features, information and structure that a user might explore. There is no implicit assumption that it provides all categories of information or that it perfectly mimics a natural setting. A virtual environment (VE) provides the user with access to information that would not otherwise be available at that place or time, capitalizes upon natural aspects of human perception by extending visual information in three spatial dimensions† and may supplement this information with other sensory stimuli and temporal changes. Any three-dimensional computer animation, would fit within this definition. Hence a further constraint to introduce is that a virtual environment enables the user to interact with the displayed data. At the simplest level this may be through real-time control of the viewpoint and at a more complex level the VE may facilitate active manipulation of the system parameters. Is there any benefit in calling a three-dimensional surface model, such as the lower plot of Figure 1, a VE? Even if this is rotated in response to the user’s mouse movement it is still just a three-dimensional model. But what if we allow the user to zoom in and drive across the silvered landscape, seeing it rise and fall

† There are several recurrent debates that arise within VR discussion groups, one is the evergreen retorical question of ‘‘what do we mean by reality...’’, the other is ‘‘why restrict VR to three dimensions’’. The former we consider to be irrelevant to current issues, the latter is constrained by human perception: a model may be parameterized across multiple dimensions, but human perceptual experience appears to be constrained to three Cartesian or Euler dimensions, that may be supplemented with colour or luminance and changes over time or frequency. Irrespective of the dimensions of the model, its presentation is constrained to the spatial dimensions that the user can experience.

around him or her as the model's parameters change, is this now a VE? What delineates between the two examples is the virtual structure that is perceived.

2.3. VIRTUAL STRUCTURE

The definition of *presence* within virtual environments has become embroiled in the debate about the nature of human experience (e.g. Loomis, 1992; Tromp, 1995). We rely on simple perceptual criteria, relating to the presence or absence of virtual structure, to distinguish between CAD and VEs. If one is presented with a three-dimensional depiction on a conventional screen, then irrespective of the degree of animation, the monitor surround, specularity of the screen and binocular vision informs one that this is purely a flat screen depiction. If the image is projected onto a large cinema screen, with low ambient lighting, then animation of the image will have a more profound effect and may even produce vection (an illusion of self-motion) for the stationary user. If one is free to move one's viewpoint, however, the illusion will be dispelled. Moving one's head sidewards should produce perspective changes in a natural scene, which are absent in a simple projection onto a two-dimensional surface, hence once again the illusion of a true three-dimensional structure is lost. If the display is made head-responsive, however, by changing the projection as the observer moves or by using a head-mounted display, then virtual depth is created. The optic array is consistent with a true three-dimensional space and objects appear or disappear behind others as the observer changes his or her perspective. In this way the observer perceives virtual surfaces structured in depth. The same principle holds for the use of stereoscopic images, which support the illusion of surfaces lying outside of the projection surface. There is a clear distinction between the observer perceiving that they are either within, or interacting with a (virtual) structured environment, and perceiving the display as a two-dimensional projection of a three-dimensional animated model. Perceiving oneself to be within a (virtual) structured environment is also the underpinning of a sense of *presence,* and provides a basis for identifying systems that are likely to engender a sense of immersion.

Given these criteria Figure 2 attempts a classification of computer simulations. A computer model at its most basic level will produce numerical output, although in some cases it may be appropriate to display this in a three-dimensional visual or auditory format. Whether this output is a VE, a computer-animation or CAD, depends on the level of interaction and whether it presents a true virtual structure (see previous section). There seems little value in blurring the boundaries and using the term VE/VR for all computer animations. Multimedia technology may be used to produce immersive VE displays, but many multimedia applications use conventional formats and presentation devices. Tele-operation presents an interesting case: when an operator is controlling a remote probe via a single camera-screen link, then this is the simple use of video technology. Given that the images may be relayed directly, or that a training system could recall images from disk, in response to the operator's inputs, then the difference between tele-operation and standard multimedia becomes blurred. It is feasible, however, to enhance perception of the remote environment, by providing two camera viewpoints that are presented stereoscopically, or a camera slaved to the operator's head movements. In this case the display satisfies the criteria of structured three-dimensional space, where surfaces appear to

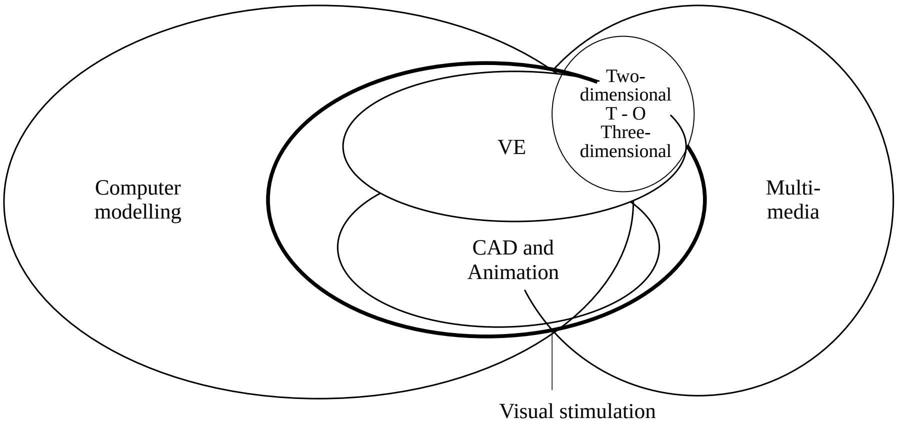


FIGURE 2. Classification of computer simulation: computer models may produce numerical or graphical outputs and some of these may support three-dimensional visual simulation. Multimedia may also contribute to computer depiction, but also has distinct areas of application. There is a poorly classified degree of overlap between virtual environments (VE) and computer aided design/depiction (CAD). Finally, tele-operation (T-O) can be classified as a VE or multimedia depending upon the extent of the simulation (two-dimensional vs. three-dimensional, see main text).

be nested in depth and tele-operation falls clearly into the bounds of a virtual environment created through multimedia techniques. Hence it is proposed that virtual environments are a subset of computer simulation and may include the use of multi-media technology. Ideally, CAD and VE should be considered as exclusive sets, but there are areas of overlap where the distinction becomes blurred.

3. What do virtual environments need?

We have proposed that the concept of VE, extends beyond simple three-dimensional depiction and VE should be used to describe systems which support salient perceptual criteria (e.g. head motion parallax, binocular vision) such that the user is able to perceive the computer generated image as structured in depth. Without such a delimit the term VE/VR becomes applicable to any interactive video display. Although some very effective virtual displays may take the observer on a passive tutorial tour, a central component of advanced VE systems is the ability to interact, and intrinsic to this are the principles of *direct manipulation.* These have previously been discussed in terms of two-dimensional computer systems, but the principles translate easily to VE systems. Hutchins, Hollan and Norman (1986) propose that the interface should minimize the *distance* between one's thoughts and the physical requirements of the system. Furthermore it should allow *engagement* with objects that can be *manipulated* through simple actions rather than through interaction at the level of programs and command structures. What the design of virtual environments needs is a principled consideration of the factors that will affect the user's perception of VE structure and allow Natural Extraction of Environmental Dimensions (NEED-1). Beyond merely perceiving the environment the system must also support Navigation, Exploration and Engagement with Data structures (NEED-2). Both of these goals have at their centre the human actor and what is required, first and foremost, is a consideration of human perception (Rushton

& Wann, 1993). A focus on the perceptual-motor capabilities of the user has been applied to more conventional interfaces. Card, Moran & Newell (1983) considered computer interaction from the perspective of the human as an information processor, making assumptions about; human visual storage time, motor output rate and the nature of working memory. Although the conclusions of Card *et al.* (1983) pertaining to the required screen update rates (10–20 Hz) and the application of Fitt's law for input devices, do not translate well to the VE setting the perspective of establishing characteristics of the user alongside the desired task is a logical starting point in establishing the requisite system performance.

Principle 1: Early in the system design process, consider the psychology of the user and the design of the user interface (Card *et al.,* 1983).

4. Natural extraction of environmental dimensions

If the goal is to support the user's perception of three-dimensional visual space what are the *essential* features that render such support, what are the *desirable* feature and what are the *optimal* features?

4.1. ESTABLISHING SPATIAL SCALE

This paper was written on a Macintosh computer, using a mouse to supplement keyboard inputs to cut, paste and delete text. When one first uses a computer mouse, or switches to a computer with a different gain on the mouse then control is erratic and iterative adjustments are required to accurately position the cursor. Once the computer becomes familiar, then the cursor can be accurately positioned with swift, deft movements of the mouse. What has been established is a spatial equivalence between movements of the mouse and the resultant movements of the cursor. The user has established a body referenced spatial scale for the computer window, such that traversing its width equates to a specific extent of limb motion.† This in turn allows the skilled user to exploit the "Gilbreth principle" (1911, see also Card *et al.,* 1983) where unnecessary movements and unsuccessful attempts at interaction are minimized. What is the body referenced scale of a virtual environment? When are objects within reach and what does the user need to do in order to engage or move beyond an object? A large portion of VE displays are not stereoscopic and do not present different views to each eye, but are bi-ocular, where the user's eyes view a single common image. In this case then the visual size of an object does NOT indicate its spatial scale. If an object fills a large proportion of a cinema screen, the object could be a large building 50 m from the camera, or it could be a coffee cup 50 cm from the camera (Figure 3). In a natural setting we are able to judge the spatial scale of visual features through the optic angle that they sub-tend and their apparent distance from the point of observation. The latter factor, apparent depth, is crucial. If body-scaled information, such as that arising from a binocular viewpoint (Figure 4) is removed, then the task of recovering apparent depth in a bi-ocular setting, where only "monocular" depth cues are present is more difficult. In the case of a virtual building at 50 m vs. a virtual coffee cup at 50 cm, it

† The precise nature of the mapping is more subtle than this simple account. Users who prefer tracker balls sometimes "flip" the cursor to an initial position and then implement some fine adjustments. This suggests some mapping between cursor motion and the force or momentum that the user should apply.

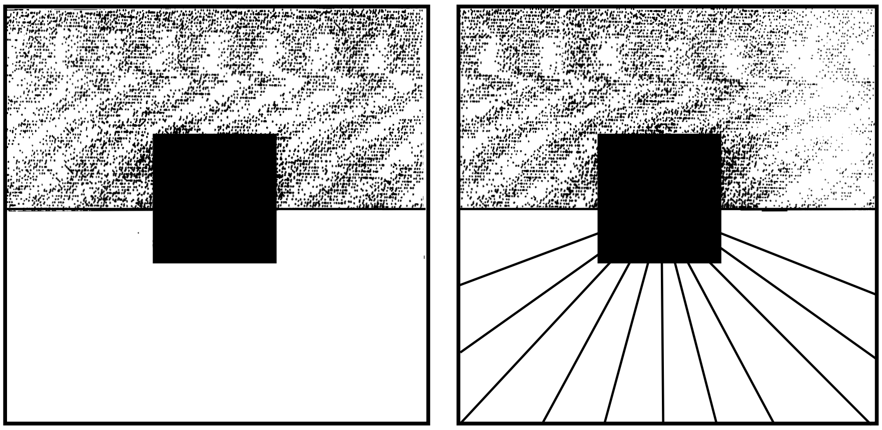


FIGURE 3. A building 50 m away or a coffee cup at 50 cm? The picture on the left is ambiguous, it could be a proximal object in front of a vertical surface that changed in colour or texture. The picture on the right includes some linear perspective which, based on an assumption that the lines are actually parallel, suggests the object is some distance from the observer.

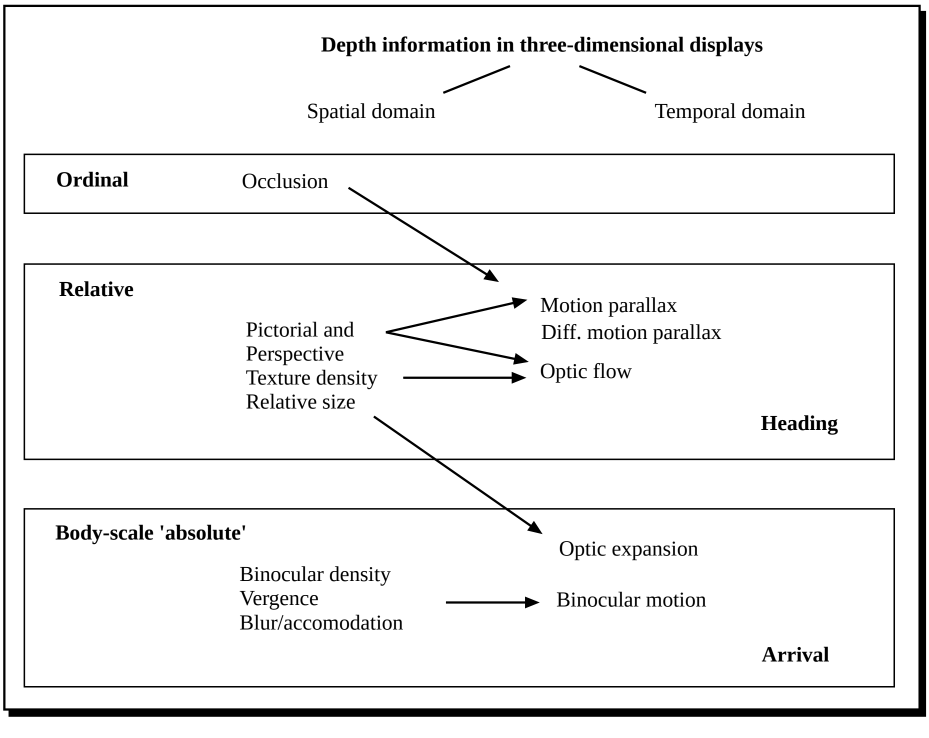


FIGURE 4. Sources of information supporting the judgements of depth and motion in depth. The arrows indicate the way in which the nature and quality of the information changes when transformations are observed over time. Hence a change, over time, in relative size provides optic expansion which can provide a direct viewer centred estimate of time of arrival.

might be assumed that one could glance down and see a ground surface receding towards the building. But in a VE where ground surfaces are flat-shaded, there is no perceptual distinction between a smooth horizontal surface receding towards the building and a surface of equivalent size that is suspended vertically behind the coffee cup (Figure 3). The perception of apparent (i.e. intended) depth is essential to providing the user with information about the spatial scale of a simulated environment.

4.2. SEEING IN THE THIRD DIMENSION

A fundamental requirement of interaction within the world is the perception of three-dimensional depth from the two-dimensional surfaces of the eyes (or eye) of an observer. In the classical account of depth perception (e.g. Woodworth, 1938), information that provides depth may be afforded both from the environment (perspective information) and from the ocular-motor system. There are implications arising from both the presentation of perspective information and the normal functioning of ocular-motor control for the design of VE displays. Within this section we will briefly consider the role of environmental perspective and the ocular-motor system in the perception of depth in both temporal and spatial domains, before looking at possible interactions between these information sources. We will then consider the implications of VR systems for the ocular-motor system.

The naturally occuring cues to depth have been extensively documented, but their salience in a VE may often be different from in the natural world and therefore their potential role requires careful consideration. Depth information within the world may be considered as either ordinal, relative or body-scaled (Figure 4).

Ordinal cues provide information regarding the order of objects in depth. The primary ordinal source of information is *occlusion.* An object that occludes another may be assumed to be in front of the surface it masks. Although this information may suggest the order of objects in the environment, it provides no impression of the relative depth of fixated objects.

Relative information, on the other hand, provides evidence that allows an observer to judge the relative depth that separates one object from another. One such relative source of information within a visual scene is pictorial. The term *pictorial* will be used to refer to information sources such as shadowing, height in the scene, linear perspective and aerial perspective. Shadowing refers to the perception of depth afforded by a shadow of one object falling upon another surface when the source of light is either known or assumed. Objects at infinity are optically projected at eye-height, but not all objects at eye-height are at infinity. Because the human visual apparatus is supported at some distance above the ground surface, objects that move in depth along that surface will increase in their vertical eccentricity and reciprocal motion will occurs for objects that lie above eye-height. In simple terms, *height in the scene* dictates that an observer must gaze up or down to track objects as they approach, unless they lie on a direct collision course with the eye. Hence, given the assumption that unsupported objects gravitate to a ground surface, then height in the scene provides an estimate of their relative depth. Linear perspective arises from the convergence of parallel lines as they recede in depth (e.g. Figure 3) and aerial perspective refers to the relative increase in light of the blue wavelength that arises from distant objects due to the refractive properties of particles in air.

Two sources of relative information that have been separately listed in Figure 4 are *texture gradients* and *relative size.* Their exclusion from the pictorial category is not because they are not present in a pictorial representation, but because of their particular salience. Many materials within the natural world have a degree of irregularity in surface colour or reflectance that may be generally called visual texture. These texture elements subtend smaller visual angles as they recede in depth, resulting in a comprehension of detail and a change in relative density. This in turn allows distance judgements to be made both about the relative position of segments within the surface and of objects placed upon that surface. Relative optical size may operate in three ways: a larger object may be assumed to be closer than a smaller object; an object that changes in optical size may be assumed to be moving in depth rather than physically expanding or contracting; an object of familiar size, such as a human, may provide information for a relative judgement of its depth. An important difference in the use of relative information sources in the natural and VE setting is that in the latter case the underlying assumptions that support their use may not be implemented (see below).

Body-scale information, provides a direct indication of the distance of surfaces from the body. Such information is occasionally referred to as ‘‘absolute’’, but this is misleading because the scaling is dictated by individual factors. The information is scaled in relation to the respective perceptual system of the observer. Hence, *binocular disparity* and *vergence angle* are directly related to the user’s inter-pupillary distance and if this is changed by viewing the world through a tele-stereoscope errors in depth perception occur. Similarly *blur* is depth information relative to the accommodative state of the eye and *accommodation* is relative to the state of the ciliary muscle. We will therefore use the term **body-scale** to describe these information sources because they are referenced to the geometry or scale of the user’s perceptual machanisms. This should not be confused with the previous use of the term ‘‘body-scaled’’ to describe size judgements relative to eye-height (Warren, 1984). The latter refers to the general catagorization of relative judgements and is an approximation to body scale, not a direct estimate of depth scaled to the viewer’s perceptual geometry.

4.3. PERCEPTUAL ILLUSIONS AND ENVIRONMENT DIMENSIONS

Relative depth information is based upon an underlying set of assumptions that may render it open to misinterpretation. Linear perspective requires the assumption that environmental features such as roads and buildings often have parallel edges (e.g. Figure 3), whereas height in the scene relies upon the natural tendency for unsupported objects to rest on the ground plane. Texture gradient also relies upon regularity of surface detail. Hence, environments where these assumptions do not hold, will promote errors in the judgement of depth. Such errors can be observed in some naturally occurring settings, or in contrived environments such as in the Ames room illusion. The latter is produced by an irregular room with systematic distortions of its surfaces such that it appears to be a conventional cube. The profound effect is that, to a stationary observer, an object of a familiar size, such as human standing in the room, is perceived disproportionately. Relative or familiar size may also contribute to errors in depth perception. It has been proposed that some road accidents involving children are due to car drivers assuming that a child pedestrian is actually a larger adult pedestrian at a greater distance from the driver (Stewart, Cudworth & Lishman, 1993). The basic principles of perspective geometry are intrinsic to most three-dimensional graphics systems, such that motion of the viewpoint will result in appropriate transformations of size and perspective. These geometric principles, however, do not guard against perceptual errors. Virtual environments do not need to conform to the regularities of the natural world: there is no reason why objects within a VE display should gravitate to the ground plane, this is actually a rather inefficient use of the display space, so it may be logical to have icons at a range of heights within the scene. A related argument might be that objects or icons should be recognizable whatever their distance, so that relative size scaling might also be inappropriate. Light sources in the natural world are often unidirectional (daylight) or diffuse (interiors). Multiple light sources, such as streetlights produce complex and transient shadows. There have been few concrete proposals as to how light should be exploited in virtual environments. Are multiple spotlights acceptable as a device to direct attention or should the sun always shine? Some virtual environments may not include a specific ground surface or rectangular geometric features. A further question then arises as to whether in VEs, where some of relative distance cues are ambiguous, there may be the need to introduce information sources that support more accurate depth judgements. A promising candidate for providing information support is "smart" texturing. Providing some global texture detail provides a consistent depth reference and incurs less computational demand than an equivalent increase in geometric detail. The introduction of texture does not have to be through the creation of a ground plane. Transparent textures can be used to provide a relative depth estimate without restricting the view volume, provided the texture conforms to a stable laminar structure. An alternative is to supplement relative depth cues with information that provides a stable body referenced metric such as binocular disparity and degree of accommodation.

4.4. BODY-SCALE INFORMATION

As an object approaches from optical infinity (e.g. 6 m) towards an observer, so its retinal image becomes de-focused. In order to bring clarity to the retinal image the eye must focus through contraction of the ciliary muscle. This process is known as accommodation. When two eyes are used to fixate, then an approaching object not only causes blur but also creates an error between the target and the angle of ocular vergence (or vergence-disparity). In order to overcome the vergence-disparity, the eyes must converge to maintain fixation in corresponding retinal areas. In order to ensure that accommodation and convergence responses are accurate, the two systems are cross-linked with one another so that accommodation produces vergence eye movements (accommodative vergence) whilst vergence causes accommodation (vergence accommodation). Under normal viewing conditions, accommodation and convergence thus vary synkinetically and are dependent on object distance.

4.4.1. Binocular issues

As previously argued, when creating a virtual environment display designers must decide what information they wish to present to the user. In creating a visual display designers may wish to present vergence-disparity, in common with that occurring in a natural environment, when fixating objects at different depths. The creation of appropriate vergence-disparity may be relatively easily created through the presentation of separate images to the right and left eyes using appropriate geometrical representations of an object as it would normally appear to the visual system. The creation of appropriate blur information is problematic, because the blur should respond to the user's accommodative effort, which requires introducing blur through focal depth adjustment, rather than Gaussian filtering. Current VE systems require fixation on a planar image surface and all virtual objects have equivalent focal depth and blur irrespective of their virtual depth (as specified through pictorial cues or disparity). Such an arrangement causes problems for the visual system because of the normal cross-linkage between vergence and accommodation so that the normal relationship between accommodation and convergence is disrupted, and blur and disparity cues are mismatched (Wann, Rushton & Mon-Williams, 1995). Early VR systems have been shown to cause problems to the visual system (Mon-Williams, Wann & Rushton, 1993). Removal of vergence-disparity information in a VR system has been shown to avoid such problems of binocular vision disruption (Rushton, Mon-Williams & Wann, 1994) and evidence has been forwarded that the presence of conflicting vergence-disparity and blur information forces the visual system to make adaptive changes (Mon-Williams, Rushton & Wann, in press).

In view of the demands placed upon the visual system by vergence-disparity information, one may ask whether this information is necessary in depth perception? The classical view has been that depth information is available from the ocular-motor system. When an eye accommodates on an object, body referenced information is theoretically available from the ciliary muscle about relative exertion and therefore relative depth. Similarly, depth information is potentially available about convergence angle, through the extra-ocular musculature, when two eyes fixate an object that may allow the estimation of depth through triangulation. The empirical evidence supporting the role of the ocular-motor system is equivocal. Accommodation has been demonstrated to be sufficient to allow coarse grading of depth in some individuals (25% of a research population) but that judgements were somewhat inaccurate and compressed in scale (Fisher & Cuiffreda, 1988). It has also been shown that it is possible to drive vergence eye movements without inducing a percept of motion in depth and, hence, vergence eye movements do not provide a sufficient cue for depth perception (Erkelens & Collewijn, 1985).

4.4.2. Is binocular information desirable?

As the ocular-motor system, *per se,* apparently provides little depth information, one may ask whether it is necessary to support vergence-disparity information in a VE? One advantage in providing the user with binocular viewpoints is that it provides a metric for environmental scaling. Depth information is available to the binocular visual system by virtue of the inter-ocular separation of the eyes in the forehead. Stereo-perspective (or stereopsis) arises because of small disparities between retinal images located away from the fixation point. Such stereo-perspective may be differentiated from vergence-disparity as the differences between retinal images are small enough to allow fusion without the need for ocular vergence. Although it is commonly proposed that the enhanced visual information available from stereo-perspective is necessary for fine manipulative tasks, there is a paucity of robust empirical evidence that this is the case. This does not devalue its role in spatial scaling. Even if the binocular information supplied does not, in itself, allow precise depth judgement its supplementary role to pictorial cues may be vital. Binocular information in the form of vergence disparity also clearly distinguishes between objects which are proximal (e.g. 30cm), distal (6m) or mid-way between these extremes (e.g. 1m). Hence in the example of the building vs. coffee cup (Figure 3) there would be no ambiguity in a stereoscopic presentation. Binocular presentations therefore provide a useful tool for scaling and re-scaling the virtual environment. Compressing or expanding the spatial detail relative to the binocular perspective is equivalent to changing the user's inter-ocular distance and rescaling the world relative to the perceiver. There are subtle adaptation effects that follow the introduction of such a manipulation in the natural world, through the use of tele-stereoscopes. Over large scale changes in binocular perspective, however, the effect should be robust and it should be possible for the observer to scale themselves between an apparent micro or macro size relative to the objects on display.

4.5. SUMMARY: NEED-1

A number of the depth cues that are generally reliable in the natural world may be deliberately manipulated in a virtual environment. An important issue in the design and construction of such environments is how stable and accurate perception of the virtual environment can be supported. Natural perception of the environmental dimensions cannot be considered to be a "natural" consequence of presenting the observer with a three-dimensional computer generated scene. In constructing a three-dimensional environment there needs to be a principled consideration of the following.

(i) What spatial features will be present and what depth information might they supply?
(ii) Are there potential sources of depth information that may be ambiguous or misleading in the environment?
(iii) What sources of information provide a robust, unambiguous estimate of both relative depth and environmental scale?
(iv) Is it necessary to include any body-scale information, such as binocular perspective to disambiguate proximal and distal surfaces?

5. Supporting navigation, exploration and engagement

Navigation

Up until this point we have been considering a passive observer presented with a static computer generated scene. Interaction between a user and the environment, however, involves the use of information that reflects both spatial and temporal changes of the relative environment. It is important that an observer is able to determine where he or she is heading when moving through the world and also to estimate how contact with an object can be made or avoided. Sources of information available within the spatial domain can also support temporal judgements, but the nature and quality of the information is changed (Figure 4). The relative position of edges that furnish occlusion information in a static scene, provides powerful information through *motion parallax* when relative motion occurs. As an observer moves to the right, objects move to the left and increase or decrease their relative degree of occlusion. Hence an ordinal source can provide relative motion information, that may in some cases support a direct estimate of the heading through differential motion parallax (DMP: Cutting, 1986). Another important source of heading information arises through the global change in features and texture detail as the observer moves. Relative motion of the viewpoint produces a temporal transformation of visual detail that has been called *optic flow* (Gibson, 1950). Typical examples are that motion forwards or backwards produces a radial pattern arising from either expansion or contraction of detail, respectively. Rotary head-eye movement produces a solenoidal flow around the point of observation. The structure of the optic flow pattern appears to be sufficient to judge the direction of heading (Warren, Morris & Kalish, 1988) and to accurately adjust heading in simulated environments (Wann *et al.,* 1995*b*). There is a current debate as to the accuracy of control that can be achieved purely on the basis of the optic flow pattern, without recourse to non-visual, body-referenced information about the direction of gaze, but this does not reflect directly upon the VE design issue. Current research suggests that the fine-grain optic flow that arises from motion relative to a textured surface provides valuable information to the observer about heading and velocity of motion. In this respect it should be considered as an *essential* feature in the design process if effective database traversal is to be supported. The research debate is that in some circumstances, such as when the observer turns to gaze at other proximal objects, the optic velocity field may not be *sufficient* and other information may be required to disambiguate the direct of motion. It has already been discussed that large scale features can also provide heading information through differential motion parallax. But given the geometric overheads incurred by populating a three-dimensional computer environment with non-essential objects, it is important to reconsider the issue of whether additional edge/object information is *necessary* or whether the existing essential-features, supplemented by some texture detail is *sufficient.*

Arrival

In order to engage a data structure, whether this entails detailed inspection or direct manipulation, it is essential that the observer places themselves in a suitable action-space. Drawing an analogue back to the word-processor on which this is being written, to inspect the introductory section on "**who neads virtual reality?**" a slider on the document scroll bar must be dragged upwards. It is efficient if one can judge how far it needs to be moved otherwise scrolling is laborious. To manipulate the text I need to position the line cursor within a suitable action space. So to correct "**neads**" to "**needs**" in the bold title above, I must click just before the "**d**". The precision of the action required depends upon the spatial scale of the text I am operating upon. In a virtual environment equivalent requirements are presented. Fundamental to most VE settings is the requirement that the users will want to change viewpoint, traverse the three-dimensional structure and position themselves at areas of interest. Where they should stop is dictated by the spatial scale and user intention. The proximity needs to be such that the user can discern relevant visual detail, or activate/manipulate the structure with a three-dimensional mouse input.

The information for judging arrival may be supplied through changes in relative size. As an object approaches, so does its retinal size increase and the change in size provides *optic expansion* information that allows for judgements of time of arrival (Figure 4). The classic treatment of this problem was provided by Lee (1976) who proposed that the relative rate of optic expansion (τ) can specify time to arrival information irrespective of the spatial scale of the environment being traversed. The empirical evidence supporting Lee's τ proposal is equivocal (Wann, in press), but the salience of optic expansion in providing general information about approach and potential collision is well documented. The issue for VE design is that optic expansion is dictated by the optical size of the object being approached, its distance from the observer and observer velocity. Even if τ is not affected by global spatial scaling, such scaling may reduce optic expansion below human perceptual thresholds. This may also be confounded by display resolution and spatio-temporal aliasing effects. From the previous section on body-scaled information it is obvious that binocular information may provide an essential metric for the user to judge when they are within a proximal action-space. There is also some evidence that it may aid judgements of impending collision or arrival (Wann & Rushton, 1995).

5.1. SUMMARY: NEED-2

We have discussed that an essential option afforded by a virtual environment is the ability to change the three-dimensional viewpoint and to move to areas that allow user engagement with sub-components of the environment. A first stage in making VE systems easy to use is to satisfy the "Gilbreth principle" and avoid unnecessary, poorly judged and executed interactions. Considerable attention has been directed towards such tasks in the design of two-dimensional mouse-window interfaces. Extending the interface into three-dimensional does not simplify the issue but amplifies its complexity. How can we support the user's judgements of self-motion and object motion? How can we avoid users veering off course and crashing into virtual obstacles? As for NEED-1 we can posit questions that should be addressed in VE construction as follow.

(i) What information about relative motion will arise from the spatial features that are present in the environment?
(ii) Are there delimits, such as update rate, display resolution and velocity of motion that may produce misleading relative motion information?
(iii) What sources of information can be supported that provide a robust, unambiguous estimates of relative motion?
(iv) Is it necessary to include any body-scale information, such as binocular perspective to supplement the motion information?

6. Summary and conclusions

The intended goal of this paper was two-fold. First, given the proliferation of information about "virtual reality" systems we attempted to define what may constitute a virtual environment, and the circumstances in which such displays may be useful occupational tools. It is inevitable that others may disagree with our terms of reference and the delineations we propose in Figure 2, but there is a growing need for some consensus on what the academic and manufacturing communities accept is "virtual reality" and the transmission of this information to the media and public. A suitable example is the "*Tree of Knowledge*" a lay reference guide published in weekly installments and widely advertized as a substitute for the traditional encyclopaedia. Issue 1, Volume 1 (January 1996) included a tutorial on virtual reality, which was quite accurate on both technical criteria and some areas of application. Unfortunately, the notion of recreating "reality" drags the reader into areas of speculation, when it is stated that: "*you will be able to do your shopping without leaving home by visiting a virtual supermarket. You can walk up and down the aisles and fill your basket or trolley as you would in a real shop*". Although the application is technically feasible, why would anyone want to locomote up and down virtual aisles, vainly searching for the virtual spaghetti hoops, before standing in a virtual queue and avoiding eye contact with the other virtual people. Home-based computer shopping has already been piloted and, in its more feasible form, involves using list-based selections to access pictures and descriptions of products before assembling a list of required purchases. In this form, it is a variant of conventional database interaction with multi-media support. Extending the use of virtual reality to encompase any pictorial computer-based interaction is pointless. For this reason we propose specific perceptual criteria that not only set limits on what constitutes a virtual environment, but also serve to define the areas in which such environments may be useful.

The second goal was to stress that virtual environments are not constrained by the rules that govern the behaviour of objects in the physical world. This is one of the attractions of the virtual setting. A consequence of the relaxed constraints, however, is that it cannot be assumed that "natural" perception of the environment will occur. The design of environments needs to consider how the percepts of three-dimensional space and motion in space, that are fundamental to our everyday behaviour, can be supported in a virtual environment. This consideration needs to take a principled approach that considers the ordinal, relative and body-scaled information that can be provided, in terms of what is *essential,* what is *desirable* and what is *optimal.* It is also important to consider any additional consequences that may arise from the way in which the environment is structured. A primary example of such a design consideration is whether or not vergence-disparity is required within a display. Vergence-disparity information may be utilized when making temporal and spatial judgements, but the presence of vergence eye movements, with a display that does not promote a normal accommodative response, will then have consequences for the visual system. It is therefore imperative that designers consider what tasks will be undertaken in the virtual environment and what are the performance requirements for the user. The goal is to build environments that minimize the learning required to operate within them, but maximize the information yield. How can we build systems that allow a neurologist to rapidly acquire control and deftly move through a three-dimensional reconstruction of scan data, such that it either enhances his/her knowledge or reduces the time required to access the information. Or as Brooks (1988) suggests, to make systems "*so simple* [*that*] *full professors can use them, and so fruitful that they will*".

Some of the research reported in this paper was supported by JCI/MRC grant G911369: *Principles for perception and action in Virtual Reality.* Research into the perceptual design of VE displays is being supported by EPSRC grant GR/L18693: *Level of detail in interactive three-dimensional environments to enable efficient database traversal.*

References

Brooks, F. P. (1988). Grasping reality through illusion. In E. Soloway, D. Frye, & S. Sheppard, Eds. *Proceedings of Fifth Conference on Computers and Human Interaction,* pp. 1–11. Reading, MA: Addison Wesley.

Card, S. K., Moran, T. P. & Newell, A. (1983). *The Psychology of Human–Computer Interaction.* Hillsdale, NJ: Lawrence Erlbaum.

Cutting, J. E. (1986). *Perception with an Eye for Motion.* Cambridge, MA: MIT.

Erkellens, C. J. & Collewijn, H. (1985). Eye movements and stereopsis during dicoptic viewing of moving random-dot stereograms. *Vision Research* **25,** 1689–1700.

Fisher, S. K. & Cuiffreda, K. J. (1988). Accommodation and apparent distance. *Perception,* **17,** 609–621.

Gibson, J. J. (1988). (1980). *The Perception of the Visual World.* Boston, MA: Houghton Mifflin.

Gilbreth, F. B. (1911). *Motion Study.* New York, NY: D. van Nostrand.

Grillner, S. (1996). Neural networks for vertebrate locomotion. *Scientific American,* **274,** 48–53.

Hamming, R. W. (1962). *Numerical Methods for Scientists and Engineers.* New York, NY: McGraw-Hill.

Hutchins, E. L., Hollan, J. D. & Norman, D. A. (1986). Direct manipulation interfaces. In D. A. Norman, S. W. Draper, Eds. *User Centred System Design.* pp. 87–124. Hillsdale, NJ: Lawrence Erlbaum.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception* **5,** 437–459.

Loomis, J. M. (1992). Distal attribution and presence. *Presence* **1,** 113–119.

Ming, O. Y., Pique, M., Hughes, J. & Brooks, F. P. (1988). Using a manipulator for force display in molecular docking. *Proceedings of the IEEE Robotics & Automation.*

Mon-Williams, M., Wann, J. & Rushton, S. (1993). Binocular vision in a virtual world: visual deficits following the wearing of a head-mounted display. *Opthalmic and Physiological Optics,* **13,** 387–391.

Mon-Williams, M., Rushton, S. K. & Wann, J. P. (in press) Investigating the reciprocal cross-links between accommodation and vergence: implications for virtual reality displays. *Ophthalmic and Physiological Optics.*

Rushton, S., Mon-Williams, M. & Wann, J. (1994). Binocular vision in a bi-ocular world: new generation head-mounted displays avoid causing visual deficit. *Displays,* **15,** 255–260.

Rushton, S. K. & Wann, J. P. (1993). Problems in perception and action in virtual worlds. In T. Feldman, Ed. *Virtual Reality 1993: Proceedings of the 3rd Annual Conference on Virtual Reality.* London: Antony Rowe, pp. 43–55.

Stewart, D., Cudworth, C. & Lishman, J. R. (1993). Misperception of time-to-collision by drivers in pedestrian accidents. *Perception,* **22,** 1227–1244.

The Tree of Knowledge Vol. 1. London: Marshall Cavendish.

Tromp, J. G. (1995). The cognitive factors of embodiment and interaction in virtual environments. In M. Slater, Ed. *Proceedings of FIVE Working Group Conference,* December, London.

Tsernoglou, D., Petsko, G. A., McQueen, J. E. & Hermans, J. (1977). Molecular graphics application to the structure determination of a snake venom neurotoxin. *Science,* **197,** 1378–1381.

Wann, J. P. (in press) Anticipating Arrival: is the tau-margin a specious theory? *Journal of Experimental Psychology*: *Human Perception & Performance.*

Wann, J. P. & Rushton, S. K. (1994). The illusion of self-motion in virtual reality environments. *The Behavioral and Brain Sciences,* **17,** 338–340.

WANN, J. P. & RUSHTON, S. K. (1995). Grasping the impossible: stereoscopically presented virtual balls. In B. BARDY, R. BOOTSMA & Y. GUIARD, Eds. *Proceedings of the 8th International Conference on Event Perception and Action.* Marseilles, pp. 207–210.

WANN, J. P., RUSHTON, S. K. & MON-WILLIAMS, M. (1995*a*) Natural problems for stereoscopic depth perception in Virtual Environments. *Vision Research* **19,** 2731–2736.

WANN, J. P. & RUSHTON, S. K. & LEE, D. N. (1995*b*) Can you control where you are heading when you are looking at where you want to do? In B. BARDY, R. BOOTSMA & Y. GUIARD, Eds. *Proceedings of the 8th International Conference on Event Perception and Action.* Marseilles, pp. 171–174.

WARREN, W. H. (1984). Perceiving affordances: visual guidance of stair climbing. *Journal of Experimental Psychology*: *Human Perception & Performance,* **10,** 683–703.

WARREN, W. H., MORRIS, M. W. & KALISH, M. (1988). Perception of translational heading from optic flow. *Journal of Experimental Psychology*: *Human Perception & Performance,* **14,** 646–660.

WOODWORTH, R. S. (1938). *Experimental Psychology.* New York, NY: Holt.