Natural Problems for Stereoscopic Depth Perception in Virtual Environments

JOHN P. WANN,*† SIMON RUSHTON,* MARK MON-WILLIAMS*

The use of virtual reality (VR) display systems has escalated over the last 5 yr and may have consequences for those working within vision research. This paper provides a brief review of the literature pertaining to the representation of depth in stereoscopic VR displays. Specific attention is paid to the response of the accommodation system with its cross-links to vergence eye movements, and to the spatial errors that arise when portraying three-dimensional space on a two-dimensional window. It is suggested that these factors prevent large depth intervals of three-dimensional visual space being rendered with integrity through dual two-dimensional arrays.

Stereoscopic display   Vergence   Binocular fusion

INTRODUCTION

The increasing sophistication and availability of virtual reality (VR) systems warrants consideration by the vision research community for the following reasons. Firstly, VR systems may prove an attractive alternative to more conventional computer and stereoscopic displays in experimental research. Secondly, those working in applied areas of vision may become involved in human factors research on the use and application of VR technology in a number of occupational settings (e.g. Bajura, Fuchs & Ohbuchi, 1992). Finally, the increasing availability of this technology, for both vocational and recreational use, will raise ophthalmic issues of safety. Consequently it would seem timely to examine the limitations and potential problems of such display systems.

VR systems engender the percept of a visual environment via computer generated, structured optic arrays. Stereoscopic depth may be introduced through the presentation of disparate images. Stereoscopic displays may be produced on conventional desktop computer screens by using polarising filters or anaglyphs and overlaying two disparate images, or by using shutter spectacles to time-multiplex the generated arrays. An alternative, and more innovative, method of displaying computer generated images is to use a head-mounted display (HMD). An HMD typically uses two liquid crystal displays (LCDs), one in front of each eye, viewed through a simple lens system (Fig. 1). An HMD can be used as a bi-ocular display, where both eyes are presented with identical images that are tilted or displaced to allow the images to be fused as a quasi-planar surface at a specific depth. The most common use of HMDs, however, is for binocular presentation of large depth intervals, extending from pericorporal space to a virtual horizon. Whilst the creation of binocular images, equivalent to a wide range of object depths, is a simple task we wish to highlight a number of fundamental ophthalmic and physiological issues that are pertinent to the use of large depth stereoscopic displays for virtual environments. As the basic principles that underlie stereoscopic presentation in an HMD, a desktop system and a full-room stereoscopic volume (Cruz-Neira, Sandin & Defanti, 1993) are identical, these problems are common across any system that seems to present large spatial intervals from a dual two-dimensional (2-D) image source.

PROBLEMS FOR ACCOMMODATION AND VERGENCE

Under normal viewing conditions, accommodation and vergence eye movements vary synchenetically and are dependent on object distance. In contrast, within a VR system the eyes must maintain accommodation on the fixed LCD screens, despite the presence of disparity cues that necessitate vergence eye movements in the virtual scene. The computer generated optic array lacks the blur cues that arise in natural vision, and the focal characteristics of the display are closer to those of a pictorial representation. The possibility of producing a relative prismatic effect through decentration of the viewing optics is common to both bi-ocular and stereoscopic HMDs. A greater problem is posed to the visual system, however, by displays that attempt to generate a wide range of stereoscopic depths.

*VR Project, Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, Scotland.
†Department of Human Movement, University of Queensland, St Lucia, Australia 6072.

Current HMDs use high powered lenses so that even very small displacements of the screens relative to the lenses are liable to produce significant amounts of induced prism (Mon-Williams, Wann & Rushton, 1993). The visual system may respond to such prismatic demands by changing the resting position of vergence (e.g. by shifting the heterophoria). Rapid adaptation to induced prism is well within the visual system's capabilities but may produce problems within the fusional system after prolonged adaptation (North & Henson, 1981; Henson & Dharamshi, 1982). If too large a prism is placed in front of the eyes then the ability to cope with the prism will be reduced and this will result in abnormal binocular vision with related symptomatic complaints (Sethi & North, 1987). What constitutes too large a prism varies from individual to individual (North & Henson, 1981) and between age groups (Winn, Gilmartin, Sculfor & Bamford, 1994). Importantly, it has been shown that individuals with binocular vision anomalies lack the ability to adapt to prisms (Henson & Dharamshi, 1982) and this has been hypothesized as being one of the causes of binocular vision problems (Schor, 1979). Despite the problems that occur through the creation of prismatic errors, this source of visual stress is resolvable through the reduction of any induced prism. It is possible to remove prismatic error by either physically changing the screen separation or, in a fixed screen system, by varying the inter-camera distance as a software parameter.

The demands placed upon the visual system by stereoscopic displays with large depth intervals, are somewhat different to those created by prisms. When presented with binocular disparities the eyes are required to make vergence movements (Fig. 2). Such vergence movements would normally be accompanied by vergence driven accommodation through the cross-coupling of accommodative vergence (conventionally expressed as the gain of vergence accommodation to accommodation or the CA/C ratio) and accommodative vergence (AC/A ratio). When large stereoscopic disparities are presented to the eyes, shifting the vergence bias (heterophoria) will only reduce the prismatic demands for one viewing distance. If a range of disparities is presented in a display, no individual heterophoria calibration will be the same for distant and near space. A reduction in the vergence demand across the whole range of binocular disparities therefore necessitates a change in the relationship between accommodation and vergence that must be mediated through the reciprocal cross-links (e.g. the AC/A and CA/C ratio must be altered). Although it has been claimed that the visual system can alter the AC/A ratio, through a process of adaptation, in response to an unusual visual environment (Miles & Judge, 1981; Judge & Miles, 1985), the possibility (Fisher & Ciuffreda, 1990)
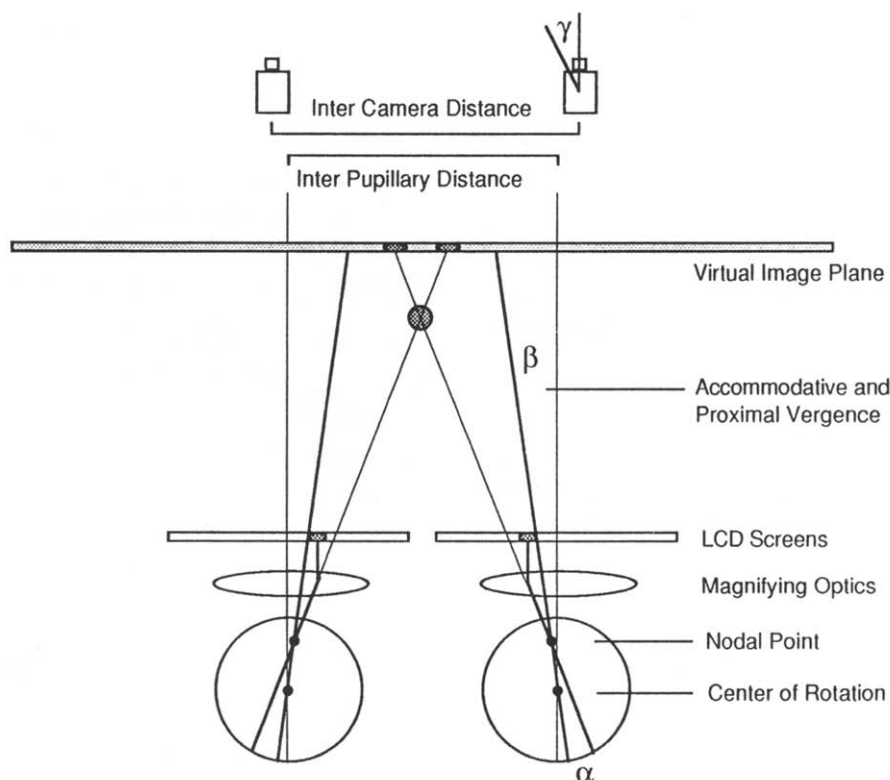

FIGURE 1. Stereoscopic design, and factors affecting spatial perception, with a head-mounted display. Images presented on LCD screens are viewed through magnifying optics, which project two half images at a visual angle (α) and a fixed focal depth (virtual image plane). Fusion of these images will result in the percept of an object in front of, or behind the virtual image plane. The screen inter-camera distance, the inter-screen separation and the viewer's inter-pupillary distance need to correspond for comfortable fusion. A degree of accommodative vergence (β) will be induced by the focal depth of the screen images, but this will be independent of, and may conflict with, disparity driven vergence. Note the angle of the virtual cameras (γ) will normally be zero, but some vergence may occur in tele-operated systems with remote cameras. Schematic layout not to scale.
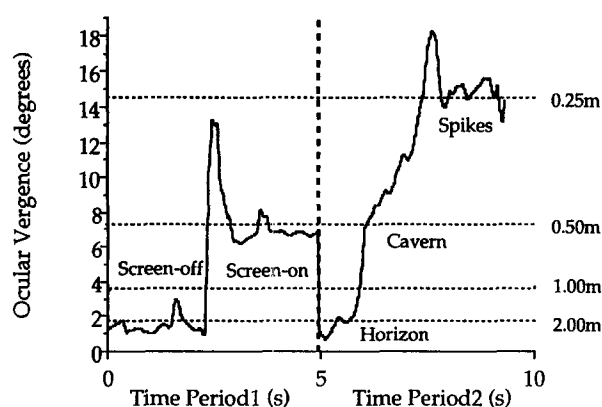
FIGURE 2. An illustration of vergence eye movements in response to a virtual display recorded with infra-red limbus tracking during the use of a HMD. During Time Period 1 a participant wore the HMD, but the screens were extinguished and the participant attempted to relax accommodation and convergence. When the screens were illuminated the participant's eyes converged to match accommodation. Time Period 2 displays vergence records for the participant travelling on a virtual roller-coaster ride. The participant rose to the top of a switchback hump and could only see distant objects ahead (Horizon). The display then "accelerated" the participant down a slope toward the mouth of a cavern, then into the dark and through a set of virtual spikes. The two peaks in Time Periods 1 and 2 are artefacts of eye-blinks. Horizontal dotted lines indicate estimated depth for the point of convergence at different ocular angles. It can be seen that the participant does not passively sample the information, but responds to disparity cues in the VR display by converging appropriately, despite conflicting accommodative stimuli.

or desirability of an adaptation response remains open to question.

Our own research used a calibration unit (Fig. 3) that could be set within the visual scene and adjusted to remove any induced prism in a bi-ocular display. In stereoscopic displays, however, the problem of a depth-range was clearly demonstrated. A setting that reduced binocular stress for a user working predominantly within a desktop area of virtual space produced diplopic images of distant objects, and a reciprocal effect occurred if the display was calibrated for distance viewing. Allowing a change in inter-camera distance caused subjects to move the image position until it was equivalent to the point of ocular vergence, regardless of the apparent location within the virtual world (Fig. 4). When presenting a large interval stereoscopic display there is no single setting of inter-camera distance that will reduce fusional effort for a user across the full depth range, hence a compromise must be reached by defining the depth range that the observer wishes to sample.

Interestingly, some forced reduction in spatial intervals occurs with desktop systems that reduces the incidence of reported problems. For desktop systems that use field sequential (time-sliced) presentations, phosphor persistence produces some degree of inter-ocular cross talk. This in turn has been shown to reduce the tolerance of diplopia in the presence of vergence movements (Yeh & Silverstein, 1990). Observers therefore find the presentation of large depth intervals on field sequential displays unsatisfactory (diplopic), and thus limit the range of

disparity presented and hence the conflict with focal cues.

THE PROBLEM OF ALBERTI'S WINDOW

The principles of perspective representation, where a view of three-dimensional (3-D) space is mapped onto a 2-D picture window, has been attributed to Alberti in the fifteenth century (Cutting, 1986). This simple procedure is an additional source of spatial error when translated to stereoscopic displays. The correct presentation of disparity cues is critically dependent on the half-image separation, image plane distance and the user's inter-pupillary distance (IPD). In traditional desk mounted stereoscopes these factors are assumed to be stable and the principles for correct half-image placement have been well documented. Vergence eye-movements, however, introduce small changes in the inter-ocular distance (IOD) because the centre of ocular rotation lies some 5–6 mm behind the nodal point of the optical system (Bennett & Francis, 1962). In a system where disparity is used to specify a wide range of depths, and vergence eye movements are to be expected (Fig. 2), a compromise must be made. Using an estimate of IOD is equivalent to assuming that the user will be fixating a single stable point. If, for example, the IOD is set for infinity but the user actually converges to a distance of 30 cm this will effectively reduce the IOD by 1 mm (this calculation is based only upon consideration of the ocular rotation and ignores the translational effects of ocular vergence). Although this is a small percentage change it can introduce spatial bias into the small disparities necessary to specify depth increments. Table 1 presents a simulation of the bias introduced when a user fixates a proximal object but the placement of half-images does not change in response to ocular position following rotation. The effective disparity is increased and results in an increase in the projected horizontal depth and eccentricity of more distant objects. It should be noted that when the user fixates at 0.5 m, virtual objects beyond 3 m fall outside the normal range of patent stereopsis and hence would normally be seen as diplopic. It is also the case that as gaze moves toward more distal locations the bias is reduced (Table 1). The introduced spatial bias may limit VR's utility in applications requiring the precise perception of the location of an object, or effector, in space. Spatial bias may be compounded in some systems that attempt to "exaggerate" stereo-depth by deliberately setting the inter-camera distance to be greater than the observer's IPD. A similar conflict, created with a telestereoscope, has previously been shown to produce errors in distance judgements (Fisher & Ciuffreda, 1990).

Spatial bias may also affect ocular-motor control. Adaptation to spatial bias is possible within the ocular-motor system. Adaptation has been demonstrated within the vergence (Maxwell & Schor, 1994) and pursuit system (Schor, Gleason & Horner, 1990) and it has been shown that a graded adaptive response occurs when two conflicting disparities are presented alternately at two
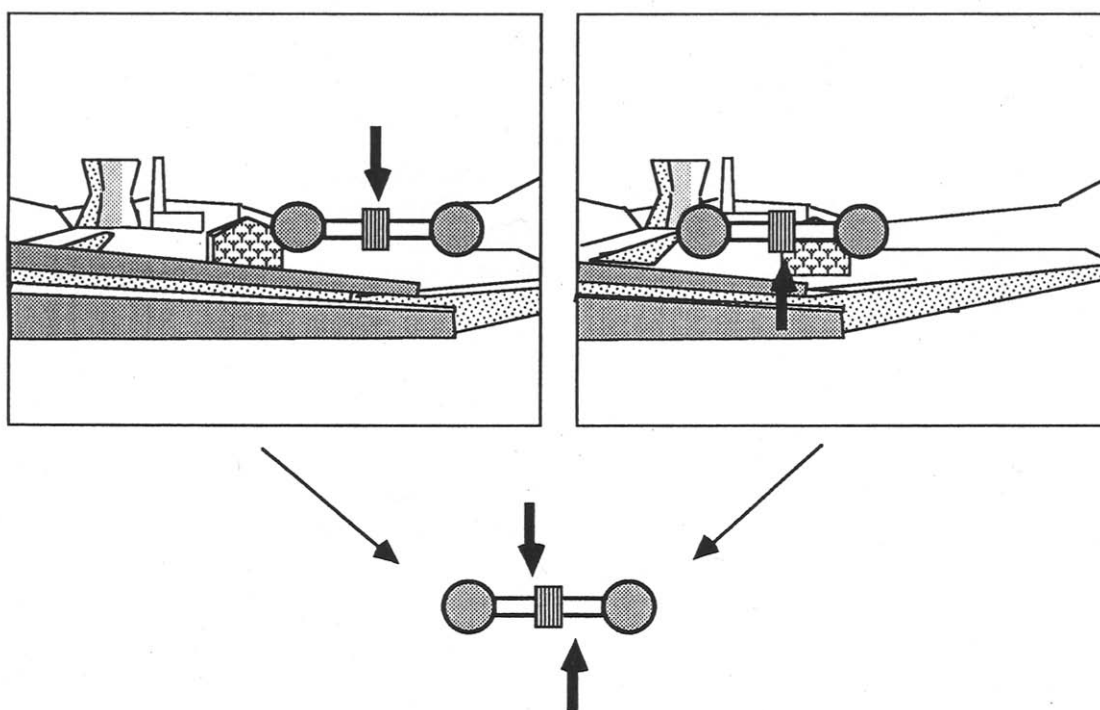
FIGURE 3. Sketch of the left and right screen views of a VR scene with equivalent portions of a calibration unit overlaid. This can be set at any apparent depth while subtending the same visual angle. The screen inter-camera distance is adjusted by the participant until the two monocularly seen arrows are aligned and centred above the binocularly fused grid. Original scene in 24-bit colour with textured overlays.

different eye positions (Schor, Gleason, Maxwell & Lunn, 1993). Non-conjugate adaptation is also possible with human saccades (Lemij & Collewijn, 1991a, b) and furthermore the adaptation has been shown to be meridian-specific (Lemij & Collewijn, 1992). The adaptation of eye movements has been shown to follow a two-part process, with an initial global change in the position of the eyes followed by a slower local change that occurs in response to the specific position demands of the stimuli (Maxwell & Schor, 1994). The time–course of

adaptation may be important when considering the consequences of immersion within a virtual environment. It should be noted that the adaptation of the ocular–motor system has been explored in relation to a fixed depth rather than over a range of disparity depths. Non-conjugate adaptation over a range of disparity depths may affect the accuracy of rapid vergence, saccadic and pursuit eye movements within a VR application and therefore have consequences for the visual system on returning to the natural world, although at this time no study has examined the importance of these factors within a virtual environment.

There are also settings where Albertian errors may be amplified. A growth area for the use of stereo displays is tele-operation, where an operator guides a remote vehicle through a real, but distant environment. A recent implementation of this approach, developed at the U.K. National Advanced Robotics Research Centre, incorporated a system which servo linked both the camera focus and camera vergence to a surface distance detector (Stone, 1992). Allowing the remote cameras to converge,
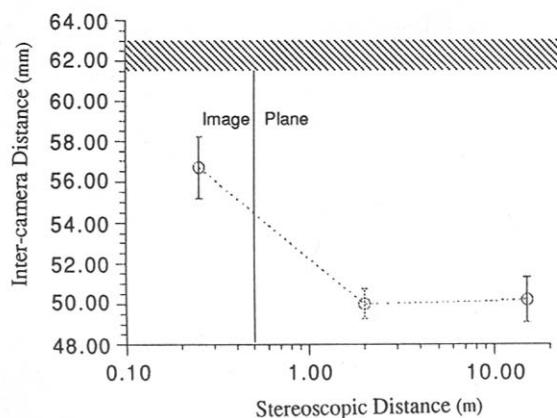

FIGURE 4. Inter-camera separation settings (O) resulting from the calibration illustrated in Fig. 3. Vertical bars indicate the SE across six subjects. The hatched area indicates the range of subjects' measured IPD and the vertical line the depth of the image plane (focal depth). At a close working distances the inter-camera separation that reduces fusional effort is significantly less than subjects' IPD due to prismatic errors within the lens–screen configuration, illustrating that the use of a measured IPD may be inappropriate for stereoscopic projections.

TABLE 1. Simulated changes in the virtual depth of elements of a stereoscopic image as a result of motion of the fixation point, resulting in a change in effective IOD

	Depth cued by retinal disparity (m)					
Intended depth						
(IOD = 65.0 mm)	1.00	2.00	3.00	4.00	6.00	8.00
Fixation at 0.5 m						
(IPD = 64.3 mm)	1.01	2.06	3.16	4.30	6.75	9.43
Fixation at 2.0 m						
(IPD = 64.8 mm)	1.00	2.01	3.02	4.07	6.17	8.31

however, is equivalent to painting the perspective scene on a tilted image window. This results in an expansion of the medial portions of the scene and compression of the lateral portions. When two such images are presented binocularly the result is a complete collapse of spatial linearity. The capture and re-presentation of true 3-D space by way of 2-D image planes can introduce greater conflicts than the generation of two equivalent perspective drawings. In this respect the monitoring of a real environment through a stereoscopic display can actually create more problems than the recreation of a virtual environment.

A SOURCE OF VISUAL STRESS?

It has been demonstrated that use of an HMD can cause deficits of binocular function after a period as short as 10 min (Mon-Williams *et al.*, 1993). Following immersion within a stereoscopic HMD environment a shift in heterophoria has been observed and, more importantly, the visual system has been shown to be less able to compensate for the presence of heterophoria as evidenced by an increase in associated heterophoria. Contributory effects may have been poor illumination, poor contrast and an unusually close working distance. It would also seem possible, however, that the discordant cues provided for accommodation and vergence may have contributed to the observed changes in binocular stability. Two possible explanations exist to explain the change in heterophoria: the visual system may experience fusional problems adapting to the unusual visual environment, or alternatively may adapt successfully, resulting in a heterophoric shift. Either process may stress the visual system and account for the observation of associated heterophoria correlated with symptomatic complaints. Such effects are not evident following the use of a bi-ocular display for extended periods (Rushton, Mon-Williams & Wann, 1994). Unfortunately however, there does not appear to be a simple solution to avoid producing visual stress when a large stereoscopic depth range is required for conventional binocular designs.

It is not being suggested that the dissociation of accommodation and convergence is the only possible (or even primary) cause of visual stress in stereoscopic VR systems, but rather that the dissociation of accommodation and vergence presents problems that are not easily resolvable by increasing display quality. It should also be noted that it is the variable, rather than the fixed, disassociation of accommodation and vergence that is problematic in the design of these displays. Hypothetically it is possible that a fixed conflict between the accommodation and vergence system within a headset might eventually produce a long-term change in the AC/A and CA/C ratio (a local effect) but this would be a later effect and secondary to the more global prismatic adaptation. It is also prudent to note that improvements in image quality (e.g. screen resolution) will produce a better stimulus for accommodation and hence exacerbate, rather than reduce, such problems.

CONCLUSIONS

Developments in screen technology and computational graphics will improve the image quality of VR displays. Unfortunately, however, it is not possible to render a large interval of veridical 3-D space with any system that used a fixed accommodation stimulus with binocular disparities. In order to accurately simulate 3-D space, from dual 2-D image sources, the system must be able to respond to where the user is regarding in visual space. One technical solution might seem to be the monitoring of eye movements and the use of this data to adjust image plane depth via a servo–lens system. The problem, however, of monitoring vergence eye movements with precision in real-time is considerable due to the response frequency of the ocular–motor system, and the lags inherent in digital transmission and servo-control. No current eye-monitoring system appears to meet the requirements of high resolution, high stability and real-time data transmission with no restriction of the field of view or head movement, and be simple enough to allow calibration by the VR user. The probable result of using conventional eye tracking techniques would be a VR display with flickering focal depth and a lagged response to eye movements. This would undoubtedly be an undesirable modification. A lack of intentional sensitivity, however, is likely to lead to the eventual breakdown of the illusion or, in some cases, the user's visual system.

REFERENCES

Bajura, M., Fuchs, H. & Ohbuchi, R. (1992). Merging virtual objects with the real world: Seeing ultrasound imagery within the patient. *Computer Graphics, 26,* 203–210.

Bennett, A. G. & Francis, J. L. (1962). In Davson, H. (Ed.), *The eye* (Vol. 4). New York: Academic Press.

Cruz-Neira, C., Sandin, D. J. & Defanti, T. A. (1993). Surround-screen projection-based virtual reality: The design and implementation of the CAVE. *Computer Graphics, 27,* 135–142.

Cutting, J. E. (1986). *Perception with an eye for motion.* Cambridge, Mass: MIT Press.

Fisher, S. K. & Ciuffreda, K. J. (1990). Adaptation to optically-increased interocular separation under naturalistic viewing conditions. *Perception, 19,* 171–180.

Henson, D. B. & Dharamshi, B. G. (1982). Oculomotor adaptation to induced heterophoria and anisometropia. *Investigative Ophthalmology and Visual Science, 22,* 234–240.

Judge, S. J. & Miles, F. A. (1985). Changes in the coupling between accommodation and vergence eye movements (the stimulus AC/A ratio) induced in human subjects by altering the effective interocular separation. *Perception, 14,* 617–629.

Lemij, H. G. & Collewijn, H. (1991a). Long-term non-conjugate adaptation of human saccades to anisometropic spectacles. *Vision Research, 31,* 1939–1954.

Lemij, H. G. & Collewijn, H. (1991b). Short-term non-conjugate adaptation of human saccades to anisometropic spectacles. *Vision Research, 31,* 1955–1966.

Lemij, H. G. & Collewijn, H. (1992). Non-conjugate adaptation of human saccades to anisometropic spectacles: Meridian-specificity. *Vision Research, 32,* 453–464.

Maxwell, J. S. & Schor, C. M. (1994). Mechanisms of vertical phoria adaptation revealed by time-course and two-dimensional spatiotopic maps. *Vision Research, 34,* 241–251.

Mon-Williams, M., Wann, J. & Rushton, S. (1993). Binocular vision in a virtual world: Visual deficits following the wearing of a

head-mounted display. *Ophthalmic and Physiological Optics, 13*, 387–391.

Miles, F. A. & Judge, S. J. (1981). Optically-induced changes in the neural coupling between vergence and accommodation in human subjects. In Lennerstrand, G., Zee, D. S. & Keller, K. S. (Eds), *Functional basis of ocular motility disorders* (pp. 93–96). Oxford: Pergamon Press.

North, R. V. & Henson, D. B. (1981). Adaptation to prism-induced heterophoria in subjects with abnormal binocular vision or aesthenopia. *American Journal of Optometry and Physiological Optics, 59*, 746–752.

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.

Schor, C. M. (1979). The influence of rapid rapid prism adaptation upon fixation disparity. *Vision Research, 19*, 757–765.

Schor, C. M., Gleason, G. A. & Horner, D. (1990). Selective non-conjugate binocular adaptation of vertical saccades and pursuit. *Vision Research, 30*, 1827–1844.

Schor, C. M., Gleason, G. A., Maxwell, J. S. & Lunn, R. (1993). Spatial aspects of vertical phoria adaptation. *Vision Research, 33*, 73–84.

Sethi, B. & North, R. V. (1987). Vergence adaptive changes with varying magnitudes of prism-induced disparities and fusional amplitudes. *American Journal of Optometry and Physiological Optics, 64*, 263–268.

Stone, R. (1992). Virtual reality and telepresence. *Robotica, 10*, 461–467.

Tyrrell, R. A. & Liebowitz, H. W. (1990). The relation of vergence effort to reports of visual fatigue following prolonged near work. *Human Factors, 32*, 341–357.

Yeh, Y.-Y. & Silverstein, L. D. (1990). Limits of fusion and depth judgement in stereoscopic color displays. *Human Factors, 32*, 45–60.

Winn, B., Gilmartin, B., Sculfor, D. L. & Bamford, J. C. (1994). Vergence adaptation and senescence. *Optometry and Vision Science, 71*, 1–4.