Perceptual stability during head movement in virtual reality

P.M. Jaekl [2], R.S. Allison [3], L.R. Harris [1,2], U.T. Jasiobedzka, H.L. Jenkin[2], M. R. Jenkin [3], J.E. Zacher, D.C Zikovitz [1]

Centre for Vision Research and
[1]Departments of Biology, [2] Psychology and [3] Computer Science
York University, 4700 Keele St., Toronto, Ontario, Canada, M3J 1P3

jaekl@cs.yorku.ca, allison@cs.yorku.ca, harris@yorku.ca, ursula@cs.yorku.ca
hjenkin@yorku.ca, jenkin@cs.yorku.ca, zacher@hpl.crestech.ca, danz@yorku.ca

Abstract

Virtual reality displays introduce spatial distortions that are very hard to correct because of the difficulty of precisely modelling the camera from the nodal point of each eye. How significant are these distortions for spatial perception in virtual reality? In this study we used a helmet mounted display and a mechanical head tracker to investigate the tolerance to errors between head motions and the resulting visual display. The relationship between the head movement and the associated updating of the visual display was adjusted by subjects until the image was judged as stable relative to the world. Both rotational and translational movements were tested and the relationship between the movements and the direction of gravity was varied systematically. Typically, for the display to be judged as stable, subjects needed the visual world to be moved in the opposite direction of the head movement by an amount greater than the head movement itself, during both rotational and translational head movements, although a large range of movement was tolerated and judged as appearing stable. These results suggest that it not necessary to model the visual geometry accurately and suggest circumstances when tracker drift can be corrected by jumps in the display which will pass unnoticed by the user.

1. Introduction

In a perfect virtual world, each eye would be presented with exactly the visual geometry that it would be exposed to when viewing a real scene. Eye, head and body movements would lead to perfect and instantaneous updates of each eye's view in a way that depended on the spatial relationships between its nodal point and the elements in the scene. To render these changes exactly in virtual reality is a demanding task that requires accounting for the exact spatial relationship of each eyes' nodal points with the equipment and with the virtual world. Virtual reality (VR) head tracking systems are typically referenced to the head and the actual position and orientation of the eyes can only be approximated from the head tracking data. Head mounted display (HMD) systems are typically not calibrated to the individual user, and although eye tracking systems have found application in VR as an input device [1], eye tracking data is typically not used to update the viewpoint with precisely the correct geometry for display generation.

Since it is not possible to model all of the components that contribute to the visual geometry of viewing a real environment, it becomes useful to establish the effect of approximations of the geometry on the user's perception. Which features of the normal relationship between the user and the environment can be excluded or approximated before the user's perception, performance or sense of immersion and comfort are affected?

Answering these questions fully is hampered by the fact that there is no standard that can be presented in VR against which performance can be measured. Rather we set out to measure people's tolerance to inaccuracy in one aspect of the geometry involved in VR simulation: the scale of the motion of the visual display that is intended to compensate for head movement when looking around a static virtual environment. Head movement – typically full rotation and limited travel translation – is a common feature of HMD-based systems, and even larger

Figure 1. Experimental setup. A Puppetworks mechanical head tracker provided tracking data which is displayed on a V8 HMD using a SGI O2.

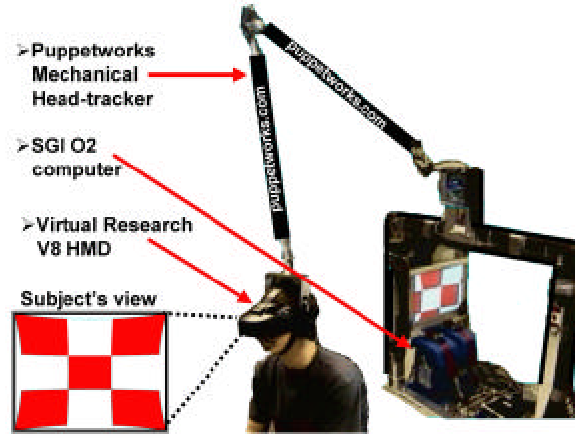

Figure 2: Presenting the visual world. The subject's head was centred inside a virtual sphere with a 2m radius. The sphere is textured with a red and white checkerboard pattern.

range translations are becoming typical in CAVE™-like systems.

2. Methods

A HMD-based VR system was constructed in which a variable gain was introduced between the measured head motion and the resulting visual motion displayed in the HMD. Subjects made controlled head rotations and translations and adjusted the gain until the world appeared stable. Head movements were either parallel or orthogonal to gravity.

2.1 Subjects

Eleven subjects participated in these experiments. Subjects had normal or corrected-to-normal visual acuity and had no history of vestibular or balance problems. Experiments were approved by the York Ethics Approval Committee. Subjects were paid above the standard York subject rates.

2.2 Visual Simulation and Head Tracking

An immersive visual display was presented using a Virtual Research V8 stereoscopic head mounted display operating in monoscopic mode. The displays, one for each eye, present full-colour, 640 by 480 pixel images at 60 Hz. The displays subtended a diagonal field of view of 60 degrees. Stereo headphones presented stereophonic sound to the subject.

The position and orientation of the head was sensed by a Puppetworks six degree of freedom mechanical head tracker which was equipped with a counterweight to reduce the load on the user (Figures 1 and 2). One end of the mechanical tracker was earth-fixed and the other end was fixed rigidly to a custom mount on the helmet. The head tracker sensed the orientation of 7 joints between the rigid links that made up the head tracker and transmitted these data via a serial link to the SGI O2 display computer. Head position and orientation were calculated from the known kinematics of the tracker and this information was used to drive the simulation.

The virtual environment was created using custom code and Open-GL graphics. The modelled virtual world (see Figure 2) was deliberately kept simple for both computational and scientific reasons. A simple environment allowed an update rate of 30 Hz. The world used was a textured sphere similar to that used earlier in a study of display lag [2]. The visual environment consisted of a sphere 2 meters in diameter and the subject's head was initially placed in the centre of the sphere at the start of each trial. One advantage of the use of a sphere is that all imagery is equidistant and complications of parallax are minimised. The sphere was patterned with a grid lattice similar to lines of latitude and longitude (and hence the lines of longitude converged to a point above and below the subject). Before each trial the sphere was rotated so that the same portion of the sphere – that section away from the poles where the texture patterns converged -- was used for each condition, Alternate squares making up the visual texture of the sphere were coloured red or white to form the pattern. The sphere was illuminated by a single light source located at its centre. The visual display was generated using a projection whose nodal point was located at the centre of the head for the rotation experiments and between the eyes for the translation experiments.

2.3 Experimental Conditions

The tolerance of visual movement during rotational and translational head movements was assessed on separate occasions. In the rotation experiment, subjects rotated their heads under voluntary control $\pm 45^o$ around either the roll (x), pitch (y) or yaw (z) axes with the axis of rotation either orthogonal or parallel to the direction of gravity (see inserts to Figure 3) resulting in six conditions. As the direction of gravity may provide cues as to actual head motion, experimental conditions were carried out either orthogonally to the direction of gravity or parallel to it. Rotations about different axes were run in separate counterbalanced blocks. Only rotation of the head was used in generating the image motion. Subjects were instructed to synchronize 0.5 Hz movements with an electronic metronome played through the HMD's headphones. Subjects pressed the left and right buttons of a standard three-button computer mouse to increase or decrease the ratio between the amount of visual motion in the display and their head motion by increments of 0.05 or 0.1. When subjects judged the display to be earth stable they indicated this by pressing the centre button. Each condition was repeated eight times.

In the translation experiment, oscillatory movements of about +/-15 cm were made in either the naso-occipital (x), interaural (y) or up/down directions (z) directions. Subjects were arranged so that the movements were either along earth-vertical and earth-horizontal axes (see inserts to Figure 4). Subjects controlled the pace of their movements using the metronome in the same manner as for the rotation experiments. All movements were carried out by the subjects. To allow movements orthogonal to gravity, a garage creeper that rolled on a track was used. Subjects adjusted the ratio between visual and head motion in steps of 0.04 until the display appeared earth stable

For both experiments initial ratios were varied pseudorandomly, with half the trials beginning at 0.5 times the head motion and the other half initiating with twice the head motion.

2.4 Data analysis

The values reported as stable were plotted as frequency histograms (Figures 3 and 4) and fitted with a Gaussian.

$$\text{Frequency} = a * \exp(-0.5 * ((x-x_0)/b)^\wedge 2)$$

Where:

x_0 is the peak of the Gaussian
b is an estimate of the width
a is an arbitrary scaling factor

A log transformation of the independent variable (the ratio) improved the fit significantly and was used in the analyses. A repeated-measures ANOVA was conducted to determine any significant differences between the x_o's for each condition. Tukey's HSD test was then used to determine where those differences were.

3. Results

3.1 Rotation

Figure 3 shows the frequency at which each ratio of visual to head movement was regarded as stable during self-generated approximately sinusoidal head rotations. The axis about which subjects moved their heads and the orientation of that axis with gravity is shown in the inserts to each histogram in the figure. When subjects rotated their heads and adjusted the visual gain of the display to appear stable, they each accepted a range of ratios between visual and head

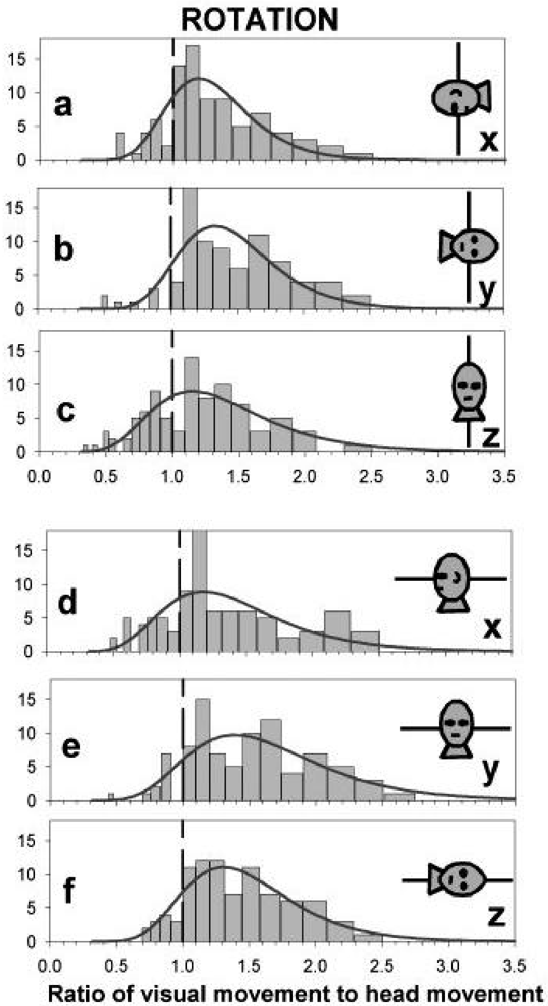

Figure 3: Frequency at which a particular ratio of visual to head movement is regarded as 'world stable' during rotation around various axes. The vertical dashed line at a ratio of 1 corresponds to visual motion which agrees, within the limits of calibration of the hardware to the subject's physical motion. Inserts show the axes of rotation and the orientation of the axis relative to gravity. a-c rotation is around earth-vertical axes, d-f rotation is around an earth-horizontal axis.

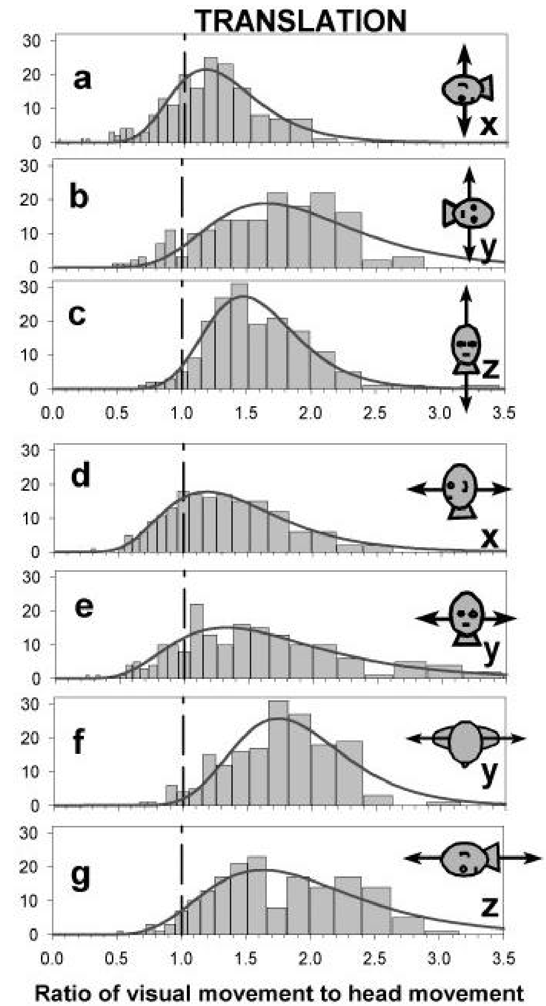

Figure 4: Frequency at which a particular ratio of visual to head movement is regarded as 'world stable' during translation in various directions and with the motions either parallel or orthogonal to gravity (see inserts). The vertical dashed line at a ratio of 1 corresponds to visual motion which agrees, within the limits of calibration of the hardware to the subject's physical motion. a-c motion is along the line of gravity, d-g motion is orthogonal to gravity.

movement. A ratio of 0 corresponds to a head-stable visual display, a ratio of 1 corresponds to a world-stable visual display, while a ratio above 1 corresponds to visual motion which exceeds that which would be generated by the physical motion. Gaussian fits to the frequency responses were constructed based on the log of the ratio of visual movement to physical movement. The most likely value to be selected by subjects as appearing stable was 1.26 as given by the average of the peaks of the Gaussian fits to the frequency histograms shown in Figure 3. That is, subjects required 26% more visual motion to match their head rotation: if the head was moving at 20 degs/sec to the right, they would judge visual movement of 25.2 degs/sec to the left as corresponding to the most stable visual display. Because the Gaussian was fitted to the log of the ratio, when plotted as a function of the ratio itself, as in Fig 3, the best fit appears skewed. That is, there was a larger range of ratios likely to be regarded as stable above the mean (more visual motion than physical motion) than below (less visual motion) it.

The histograms of Figure 3 allow us to compare the responses about different axes and for rotation about each axis when it was parallel or orthogonal to the direction of gravity. Thus Figure 3a plots the results for rotation about an axis passing through the nose (roll rotation) with the subject looking down towards the ground (axis parallel to gravity). This can be compared to Figure 3d in which the subject rotates about the same axis but with the axis orthogonal to gravity. There were no significant differences between any of the histograms in Figure 3 showing that there was no difference between any of the axes of rotation and no effect of whether this axis was parallel or orthogonal to gravity.

3.2 Translation

Figure 4 shows the frequency at which each ratio of visual to head movement was regarded as stable during self-generated approximately sinusoidal head translations. The direction in which subjects moved their heads and the orientation of that movement relative to gravity is shown in the inserts to each histogram in the figure. As for the rotation data each subject accepted a range of ratios between visual and head movement. A ratio of 0 corresponds to a head-stable visual display, a ratio of 1 corresponds to a world-stable display, while a ratio above 1 corresponds to visual motion which exceeds that which would be generated by the physical motion. Gaussian fits to the frequency response were constructed based on the log of the ratio between visual movement to head movement. The most likely ratio to be selected by subjects as appearing stable was 1.45. This is the average of the peaks of the best-fit Gaussians. That is, if subjects were moving at 20cm/s they would judge visual movement of 29 cm/s as corresponding to the most earth stable visual display. As was the case for rotation, values higher than that (faster movement) were more likely to be accepted as stable than slower movement. Indeed a ratio of two times was often judged as stable.

	ROTATION			TRANSLATION		
axis	x_0	x_0+b	x_0-b	x_0	x_0+b	x_0-b
Aligned with gravity						
x	1.20	1.55	0.93	1.16	1.54	0.88
y	1.32	1.70	1.02	1.65	2.30	1.19
z	1.15	1.62	0.81	1.48	1.86	1.17
Orthogonal to gravity						
x	1.17	1.66	0.83	1.17	1.70	0.81
y	1.38	1.95	0.98	1.53	2.13	1.12
z	1.32	1.78	0.98	1.63	2.31	1.15
Av.	1.26	1.71	0.93	1.45	2.00	1.06

Table 1: The peaks (x_0 in the Gaussian equation) and widths (b) of the best fit Gaussians to the data. The arrangements of the rotation axes and translation directions can be seen in the inserts to Figures 3 & 4.

The histograms of Figure 4 allow us to compare the responses in different directions and whether there was a difference depending on whether the movement was parallel or orthogonal to the direction of gravity as indicated by the inserts by each histogram. Thus Figure 4a plots the results for forward/backward movement of the head with the subject facing down towards the ground (movement parallel to gravity). This can be compared to Figure 4d in which the subject moves in the same forward/backward direction but orthogonal to gravity.

The ratio most likely to be judged as stable (the peak of the Gaussian fit: x_0) during translation in the x direction (naso-occipital Figures 4a & d) were significantly lower then the rest of the motions tested (1.17: $F(1,63)$, $p<0.05$, d=0.46). Whether the direction of motion was parallel to or orthogonal to gravity had no effect on the variability or the means.

4. Discussion

This study has quantified the human tolerance to visual motion in a virtual environment during rotation around yaw, pitch and roll axes and during translation along the naso-occipital, interaural and up/down directions.

4.1 Means

Surprisingly more visual movement was required than was geometrically necessary for the display to be judged earth stable, and very large amounts of movement in the opposite direction to the head movement were tolerated and judged as earth stable. When movement was slower than geometrically required - corresponding to a ratio of less than one – subjects rarely accepted the display as stable. Thus there are very few settings in any of the histograms on the left side of the dashed vertical line corresponding to 'geometrically correct'.

Why do subjects find that more visual motion than is geometrically correct provides a better sense of visual stability? Since the centre of rotation of the head is some distance behind the eyes, almost all naturally generated head rotations cause some translation of the eyes. For example, during a yaw rotation of +/- 45degs, the eyes are translated by between 10 and 15 cms. This incidental translation is associated with parallax such that retinal images are displaced depending on the distance of objects in the field. These retinal movements due to translation adds linearly with retinal movement due to the rotation. Our display moved the retinal image in response to only the rotation of the head and ignored the forward displacement of the eyes and the resulting translation. Perhaps the larger visual movement demanded by our subjects represented adding not only movement due to the rotation (ratio of 1) but also movement expected as a result of the incidental translation of the eyes (making a ratio of greater than 1). Given that our display was simulated at 1m, the extra movement needed was not as much as 26%. However the amount of translational movement expected depends on the perceived distance of the display which is likely to be underestimated [3]. The combination of allowing for incidental eye translation and judging the image to be closer than 1m, might account for the increased amount of visual motion required to be judged as stable.

An explanation based on underestimated perceived distances might also explain the excessive visual movement judged as stable during translation. For these movements, the fact that the eyes are in front of the head's centre of movement makes a negligible contribution to the geometry. But the perceived distance is critical to calculating the expected amount of visual movement.

4.2 Width of Functions

For each condition in the experiment, Gaussian fits were best for a log transformation of the data. A Gaussian fit to the logarithm of the ratio means that, when plotted against the untransformed ratio (as in Figures 3 and 4), there is an asymmetry between the frequency at which a ratio of less than one (to the left of the vertical dashed lines in Figures 3 and 4) is perceived as stable compared to ratios of more than one. That is, errors of matching visual and physical motions are more likely to be noticed by subjects for both translational and rotational head movements, if they involve too little visual movement rather than if they involve too much.

4.3 Rotation, variations between axes

For the rotational movements, there was no difference between the results obtained from rotations around different axes. Furthermore, there was no difference between judgements of visual stability associated with a head rotation that incurred a change relative to gravity and those that didn't (compare Figures 3a-c with 3d-f). Head movements that change the head's position relative to gravity (such as a pitching motion while in an upright body posture, Figure 3e) activated the otoliths as well as the semicircular canals whereas rotations around an earth vertical axis activate only the canals. That there was no difference between these movements was surprising to us. We expected the additional contribution from the otolith organs to narrow the width of the function by improving knowledge about the head movement as had been predicted by Wallach [4]. No difference was found however, suggesting that knowledge of head position was already optimal or that otoliths were not used during these active head movements [5; 6].

4.4 Translation, variations between directions

The repeated measures ANOVA for the translation experiment revealed that movement in the forwards/backwards (x) direction (Figures 4a and d) was matched with the least visual movement (lowest ratios) and was the closest of all the movements studied to requiring the geometrically correct amount of visual movement. With our display, forward/backwards movement creates radiating as opposed to laminar optic flow which is a regular feature of navigation through normal environments [7]. Subjects overestimate their self motion using this visual cue [8]. This suggests that perhaps all head movements were overestimated and matched with excessive visual movement but in the case of forwards/backwards movement, the optic flow might also have been overestimated leading to a match with a ratio closer to unity.

Again the expected difference between movements in the direction of gravity and those orthogonal to it was not found (compare Figures 4a-c with 4d-g). We expected the otolith system to be compromised when detecting movements added to or subtracted from gravity.

4.5 Consequences for VR systems

Many VR systems generate their head position signals from a system that measures movement rather than position. For example, systems that measure the acceleration of the head and then double integrate to obtain position. Such systems are vulnerable to cumulative tracking errors (drift). Such systems rely on some external reference to correct the drift, but often these drift corrections are available at a considerably slower update rate than the inertial tracker. When this correction takes place within the rendering loop it can involve a distracting jump in the display. Our insensitivity to visual movement during head movements opens the possibility of hiding the recalibration jumps by executing them during head movements. The tuning curves given in Figures 3 and 4 and summarized in Table 1 provide guidance as to the amount of correction which can be introduced before the display looses its stability. One general observation is that corrections to the visual display due to tracker drift which cause the visual display to move slower than is expected due to the normal head movement should be avoided, but that considerable latitude can be taken in introducing tracker corrections which increase the relative motion of the display.

References

[1] Jacob, R.J.K. Eye tracking in advanced interface design., In: Barfield, W. and Furness, T.A. (ed) *Virtual environments and Advanced Interface Design*. Oxford University Press, New York pp. 258-288, 1995

[2] Allison, R.S., Harris, L.R., Jenkin, M., Jasiobedzka, U. and Zacher, J.E. Tolerance of temporal delay in virtual environments, *IEEE Int. Conference on Virtual Reality* 3: 2001

[3] Viguier, A., Clement, G. and Trotter, Y. Distance perception within near visual space., *Perception* 30: 115-124, 2001

[4] Wallach, H. Perceiving a stable environment when one moves, *Annual Review of Psychology* 38: 1-27, 1987

[5] Gdowski, G.T., Boyle, R. and Mccrea, R.A. Sensory Processing In The Vestibular Nuclei During Active Head Movements, *Archives Italiennes De Biologie* 138: 15-28, 2000

[6] Roy, J.E. and Cullen, K.E. Selective processing of vestibular reafference during self-generated head motion., *J Neurosci.* 21: 2131-2142, 2001

[7] Gibson, J.J. *The Ecological Approach to Visual Perception.* Houghton Mifflin Company, Boston, 1979

[8] Redlick, F.P., Harris, L.R. and Jenkin, M. Humans can use optic flow to estimate distance of travel., *Vision Research* 41: 213-219, 2001

Acknowledgements

We would like to gratefully acknowledge generous financial support from NSERC (to MJ and LRH) and CRESTECH (to MJ, LRH). Also thanks to Jeff Laurence for building the equipment.