Target Acquisition in Fish Tank VR: The Effects of Lag and Frame Rate

Colin Ware and Ravin Balakrishnan
Faculty of Computer Science
University of New Brunswick
P.O. Box 4400
Fredericton, N.B.
Canada E3B 5A3
Email: cware@unb.ca

Abstract

A study is reported of target acquisition in 3D using a head coupled stereo display and a hand tracking device. The effects of lag and frame rate were studied by introducing lag in three different ways: by queuing the hand input and delivering it to the cursor an integer number of frames later, by increasing the frame interval and sampling the input device at the start of the display interval, by increasing the frame interval and sampling the device close to the end of the display interval. The data suggests that a simple modification of Fitts' Law can account for system lag data, also that the main degradation of performance from low frame rate is due to the lag introduced and not to the disruption of perceptual processes.

Keywords: Fitts Law, Virtual Reality.

Introduction

A number of researchers have demonstrated that a form of virtual reality (VR) display using shutter glasses to provide stereo viewing and continuous head tracking to determine the correct perspective view can result in a high resolution virtual image localized to the region in the vicinity of the monitor of a high performance graphics workstation [1,2]. Evidence to date suggests that the feeling of three dimensionality is much enhanced by head coupling, and that tree structures and networks of information can be better understood using this kind of display, which has been called Fish Tank VR [12,13].

One of the critical issues concerning the usability of this kind of environment is the influence of system lag and low update rates on performance of reaching tasks. In the 2D WIMP environment high screen update rates, typically of 60 Hz or better have become the norm, likewise hardware support for the cursor and mouse ensure minimal lag between the motion of the mouse and the cursor. This is not the case in VR environments where a 10 Hz update rate is more typical and there is significant lag between hand motion and the update of the screen [6]. This situation is likely to continue for some time because designers of VR environments have an almost insatiable need to be able to render large numbers of polygons and this necessarily increases the rendering time and consequently reduces the screen update rate.

There is a direct link between screen update rates and system lag in reaching tasks as a result of the standard double buffering technique used to achieve smooth animation. In a double buffered graphics system that responds to real-time inputs the input devices are polled and the resulting value is used to compute the graphical image for the next frame of the display. After this frame is constructed the buffers are switched at the next available vertical blanking interval. Thus image construction time contributes directly to lag. If we assume that perception occurs in the middle of the frame interval then the total lag becomes:

$$\text{MachineLag} = \text{DeviceLag} + \text{FrameInterval}*1.5 \qquad (1)$$

Given the current state of technology, a display with a 10 Hz update rate and a device lag of 60 msec (including communication delays) is fairly typical; this will yield a total lag of at least 210 msec.

Fitts' Law Incorporating a Model of Lag

The obvious experimental paradigm to use in studies of reaching behavior is the classical Fitts' Law study, although this traditionally has been used only for one dimensional reaching [3]. Of the many variants on Fitts' law the one we chose to use is:

$$\text{Mean Time} = C1 + C2\log_2(D/W + 1.0) \qquad (2)$$

which MacKenzie has argued is the most satisfying from the perspective of information theory [7]. The value of the logarithmic part of the expression or $\log_2(D/W + 1.0)$ is called the index of difficulty (ID). The quantity 1/C2 is called the index of performance, the units are bits per second.

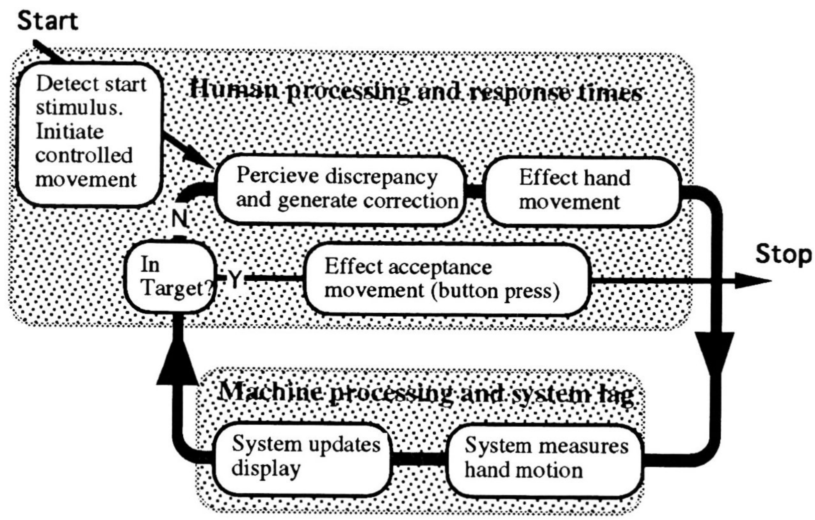

Figure 1. The control loop involved in reaching for a target in a computer graphics environment. Both human and machine processing components are represented.

There is some evidence that the process modeled by Fitts' Law is a series of movements each of which gets the hand guided probe closer to the target, until the probe actually falls within the target area [5,11]. In reality, the hand does not come to a complete stop, instead a series of corrective movements are applied in a dynamic feedback loop. This loop is illustrated in Figure 1, where it can be seen that both human and machine components are performed iteratively in series. According to this model the ID portion of Fitts' Law can be interpreted as a measure of the average number of movements (or movement corrections) required to acquire the target, or in other words the number of times the main human-machine processing loop is executed. Most Fitts' Law studies to date have assumed the machine processing lag to be zero. However, this is clearly not the case for computer graphics or telerobotics applications. We therefore modify Equation 2 so that it becomes:

$$\text{Mean Time} = C1 + C2(C3 + \text{MachineLag})\text{ID} \tag{3}$$

where C3 represents the human processing time required to make a corrective movement, MachineLag represents the machine processing time, C2 ID represents the average number of iterations of the control loop and C1 represents the sum of the initial response time and the time required to confirm the acquisition of the target. Others have found similar three parameter models to be an excellent description of the data obtained from one dimensional Fitts' Law experiments with lag [4, 9].

While the classical Fitts' Law is a model of one dimensional movement, MacKenzie and Buxton have proposed and tested a number of two dimensional variations on Fitts' Law using rectangular targets and found two of these to be successful [8]. In one variant:

$$\text{ID} = \log_2(D/\min(W1,W2) + 1.0) \tag{4}$$

where W1 and W2 are the target widths in the X and Y directions respectively, and D is the distance from the cursor to the center of the target. This rule effectively states that performance is determined by the smaller of the two target dimensions and it can be trivially extended to three dimensions:

$$\text{ID} = \log_2(D/\min(W1,W2,W3) + 1.0) \tag{5}$$

where W3 is the width in the Z direction. In the other successful variant:

$$\text{ID} = \log_2(D/W' + 1.0) \tag{6}$$

where W' represents the width of the target in the direction of hand motion - this can also be applied to 3D targets.

With large targets the subject may always group the position of the target hits well inside the target boundaries, whereas with a small target the distribution of hits may overlap the target boundaries. There is a variant on Fitts' Law which is based on the idea of an "effective target width". In calculating the index of difficulty, the actual target width is replaced by 4.13 times the standard deviation of the distribution of hits (representing a 5% error rate) [7,15]:

$$ID_S = \log_2(D/4.13\sigma + 1.0) \quad (7)$$

where σ represents the standard deviation of hits in the direction of movement. This metric may provide a more accurate measure of the rate of information processing achieved in the performance of controlled movement tasks. Nevertheless, if the goal is to predict performance in some particular situation, models of performance which include the actual target dimensions may be preferable, therefore we did not use this variation in analyzing our data.

Experiment: The Effects of Lag and Frame Rates on Target Acquisition

This study addresses three questions of theoretical interest in the context of a standard Fitts' Law target acquisition task [3] extended to three dimensions: 1) whether performance in the Z direction (in and out of the screen) is different to that in the X direction (in the plane of the screen), 2) whether the lag model given by Equation 3 provides a good account of the data, and 3) whether the performance decrement that occurs with low frame rates can be attributed solely to the lag caused by double buffering. With regards to the last question, it is plausible that low frame rates disrupt the perceptual processes and therefore reduce performance or it might be that lag is the only factor.

Experimental Platform

The experiment was conducted entirely in stereo using a Silicon Graphics IRIS Crimson/VGX graphics workstation and a 19-inch stereo capable monitor (120 Hz, 60 Hz to each eye), with a resolution of 1280 by 1024 pixels (approximately 37 pixels per cm). The subject's head position was continually tracked in order to provide a correct perspective view. Stereoscopy and tracking of head position were achieved using the StereoGraphics CrystalEyes™ shutter glasses with integral Logitech™ head tracker. To measure hand position, we used the Bat [14] (a Polhemus Isotrak™ sensor with a button wired into the mouse). Figure 2 shows the important components. This system was capable of maintaining an update rate of 60 Hz (for each eye) under all experimental conditions, although this was sometimes reduced as an experimental manipulation. Device lags were measured using a modified version of the technique presented by Liang et. al. [6] and were found to be 70 msec for the Bat and 97 msec for the head tracker.

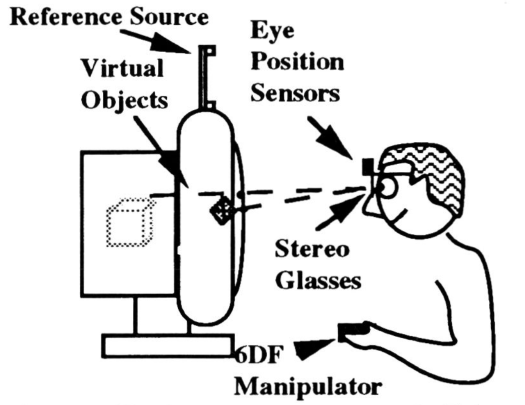

Figure 2. The important components in Fish Tank VR are the head position sensors (which allow eye position to be computed), the stereo glasses and a monitor capable of running at 120 Hx, 60 Hz to each eye. The component o the virtual scene are arranged in the vicinity of the monitor screen and the correct perspective view *from the subject 's actual viewpoint* is computed and rendered on the monitor.

Stimuli

The screen background was set to a dark gray color and two light gray wire mesh grids were drawn in the horizontal plane at the top and bottom of the screen. The purpose of the grids was to enhance the perception of depth in our VR display. A blue diamond shaped cursor, 0.43 cm wide (measured from two opposing points of the diamond) was coupled to the user's hand via the Bat. The target consisted of a reddish-purple cube drawn with antialiased wire frame edges and translucent faces. The choice of colors was primarily determined by an attempt to avoid bleeding of the image from one eye to the other which is mainly caused by the relatively slow green phosphor of the monitor. The size of the cubic target varied and represents the width for index of difficulty calculations.

Design and Procedure

The method we adopted was to decouple frame rate and lag. To achieve this we introduced lag into the hand tracking device in three different ways:

1. High Frame Rate: In this condition the frame rate was maintained at 60 Hz and lag was introduced by queuing the hand tracking device input so that they took effect an integer number of frames later.

High Frame Rate (5 conditions)			Early Sampling (5 conditions)			Late Sampling (7 conditions)		
frame rate (Hz)	frame interval (msec)	hand lag (msec)	frame rate (Hz)	frame interval (msec)	hand lag (msec)	frame rate (Hz)	frame interval (msec)	hand lag (msec)
		137	15	67	145	15	67	95
		187	10	100	195	10	100	112
60	16.7	337	5	200	345	5	200	162
		537	3	333	545	3	333	228
		787	2	500	795	2	500	312
						1	1000	562
						0.667	1500	812

Table 1. Lag Conditions

2. Early Sampling: In this condition lag was manipulated by varying the frame rate. The device was always sampled at the start of the frame intervale prior to the buffer swap whereby the data was displayed.
3. Late Sampling: In this condition lag was manipulated by varying the frame rate. The device was sampled $1/60^{th}$ of a second prior to a buffer swap. The graphical image of the cursor and the target was constructed in the ensuing $1/60^{th}$ second interval.

Twelve paid volunteers served as experimental subjects, eight of whom had prior experience with the apparatus used in the experiment. The experiment was conducted over two one hour sessions on separate days and each subject was presented with trials for a base condition with minimal lag (70 msec hand lag, frame rate = 60 Hz, frame interval = 16.7 msec) and 17 other conditions where varying amounts of lag were introduced in one of the three ways as shown in Table 1.

Lag in the head tracking device was 97 msec throughout. Each lag condition was evaluated for both the X and Z directions. This resulted in 18*2 = 36 different lag-direction combinations. Two distances (4 and 8 cm, measured from center of cursor to center of target) and one target size (1 cm) resulted in two distance-size combinations and a total of 36*2 = 72 trial conditions. At the start of each session, the subject received a practice run of 1 trial for each of the 72 conditions. Following this, they were presented with 36 blocks of trials, one for each lag-direction combination. A block consisted of 12 trials – 5 trials for each of the two distance-size combinations together with 2 practice trials given at the start of each block to familiarize the subject with that particular lag and direction. Ignoring practice trials, the result is 10 trials per block, 10*36 = 360 trials per session and 2*360 = 720 trials per subject.

The blocks were presented in random order, and the trials within each block were also randomized.

At the start of a trial in the X direction, the cursor appeared 8 cm to the left of the center of the screen while the target appeared 0.33 sec later to the right of the cursor by the appropriate distance for that trial. In the Z direction the cursor appeared in the center and in the plane of the screen and the target appeared behind the cursor (i.e., going into the screen) by the appropriate distance. The subject completed a trial by pressing the button on the Bat, moving the cursor into the target and releasing the button when she was satisfied that the *center of the cursor* was inside the target. Timing started the moment the target appeared and stopped when the Bat's button was pressed and then released. The next trial began approximately 1.0 sec later. The subject was allowed to take breaks between each block of trials, but not within a block.

Results

Overall, the data showed that performance in the Z direction was 10% slower than that in the X direction ($F(1,11) = 10.7$, $p < 0.01$).

Linear regression was used to fit the lag model given in Equation 3 to the data for the three methods of introducing lag, yielding the equations in Table 2.

In addition to the separate analyses given in Table 2 we also ran a regression on all the data combined to see how well a single equation could account for the data:

$$\text{Mean Time} = 0.739 + 1.95(0.209 + \text{lag})\text{ID} \qquad r^2 = 0.89$$

Note that this ignores the 10% difference between the X and Z directions.

Type of Data	Direction	Equation 3 fitted to experimental data
High Frame Rate	X	Mean Time = 0.78 + 1.66(0.189 + lag)ID $r^2 = 0.90$
	Z	Mean Time = 1.25 + 1.80(0.120 + lag)ID $r^2 = 0.97$
Early Sampling	X	Mean Time = 0.98 + 1.80(0.130 + lag)ID $r^2 = 0.99$
	Z	Mean Time = 0.63 + 2.01(0.211 + lag)ID $r^2 = 0.98$
Late Sampling	X	Mean Time = 0.48 + 2.29(0.204 + lag)ID $r^2 = 0.97$
	Z	Mean Time = 0.24 + 2.32(0.292 + lag)ID $r^2 = 0.96$

Table 2. Lag model in Equation 3 fitted to experimental data

The plots shown in Figure 3 illustrate the mean response times plotted against index of difficulty for the three methods of introducing lag (X and Z data combined). The overall index of performance for the above data is 1/(1.95*0.209) = 2.4 bits per second which is considerably lower than that reported in previous Fitts' Law studies which typically have been 6-10 bits per second.

Average target acquisition times for both early and late sampling of the hand tracking device, as illustrated in Figure 4, clearly shows an overall advantage for late sampling as should be expected.

Considering the very different ways in which the lag was introduced we feel that the 0.89 r^2 value obtained from all the data is remarkably good. Nevertheless we decided to reevaluate one of our assumptions to see if we could do better. This is the assumption (Equation 1) that an image is perceived at the middle of the frame interval. In the introduction, we also alluded to the possibility that lag could also be effectively introduced because of low device sampling rates. Consider the case of a very low sampling rate and a long frame interval. A subject sees the frame change and a new relative position of the cursor and the target and makes a movement towards the target based on this observation. However, this movement is only sampled at the beginning of the next frame. Thus the feedback loop can, in effect, contain an additional lag representing the lag between the time the movement is made and the time at which it is sampled. In our experiment this additional lag value cannot be separated from the perception-occurring-in-the-middle-of-the-scene lag. But the combined lags might easily be greater than the 0.5 times the frame interval that we assumed.

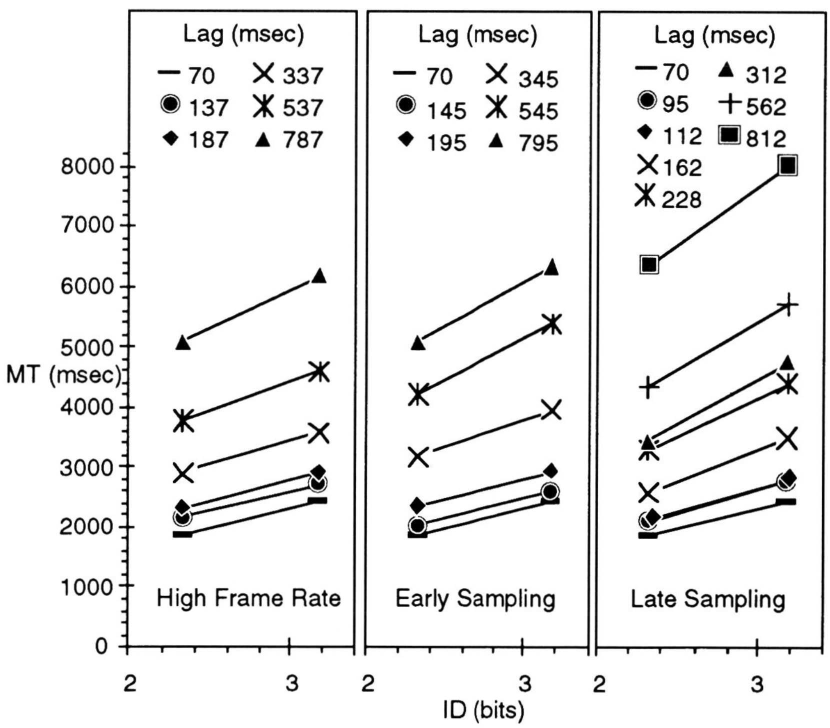

Figure 3. Averaged mean response times in both directions plotted against index of difficulty for the three methods of introducing lag.

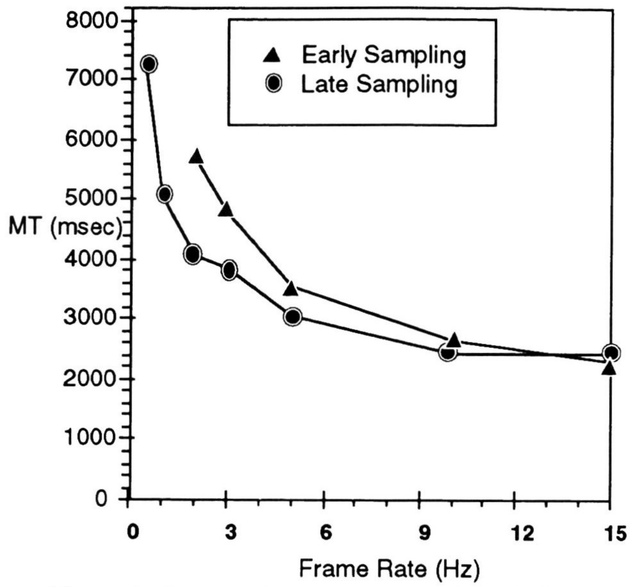

Figure 4. Averaged mean response times in both directions plotted against frame rate for both early and late device sampling methods of introducing lag

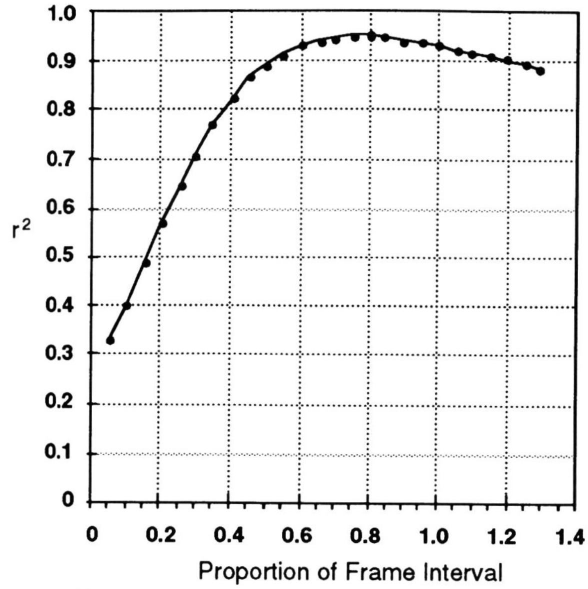

Figure 5. Regressions were computed for the entire set of experimental data with adjustments in estimation of machine lag.

To determine if some value other than 0.5 is more appropriate we ran a regression of all the data combined with different values for this lag component from 0.1 to 1.3 in steps of 0.05. The results from this exercise are plotted in Figure 5 and they show that the r^2 value peaks at 0.95 with a perception plus sampling lag value of approximately 0.75 times the frame interval, giving the following equation:

Mean Time = 0.739 + 1.59(0.266+ lag)ID

$r^2 = 0.95$

assuming

lag = DeviceLag + FrameInterval*1.75

We wish to state clearly that although this ad hoc analysis suggests that lag primarily accounts for the reduction in performance it by no means rules out some smaller effect due to the disruption in perception that could be caused by low frame rates.

Discussion and Conclusions

Our best estimate of the detrimental effect of lag is 1.59 multiplied by the index of difficulty multiplied by the lag. It is worth noting that there is at least some system lag in all Fitts' Law experiments. Those that have used a 30 Hz update rate on the monitor should probably consider a machine lag of at least 50 msec (1.5*1/30) even if the device lag is negligible. This factor has undoubtedly affected previous estimates of the human component of the processing loop.

We can derive a number of practical recommendations from these results:

- Acquire input devices which have low lag, ideally less than 50 msec.
- If double buffering is used, keep the frame rate up.
- If possible, separate head lag from hand lag. In a head coupled stereo environment, the target to be selected and the 3D cursor may be relatively small parts of the 3D graphics environment. Thus it should be possible to sample the head tracking device, draw most of the scene, and then sample the hand tracking device and draw the target and the 3D cursor. This will introduce lower lags in the task critical parts of the scene, namely the target and the cursor.
- If possible create higher update rates for the target and the 3D cursor (and hence lower lags). Pauch et. al. recently described a software architecture that supports this kind of decoupling [10].

With respect to the issue of whether 3D target acquisition is essentially different than 1D target acquisition, our data suggests that the traditional version of Fitts' Law accounts for the data well but there is a difference in the coefficients. The index of performance values we obtained are considerably lower than the values typically obtained by previous one dimensional studies, and although we did not do a direct comparison between 1D and 3D tasks this suggests that there may be problems with the simple extensions to Fitts' Law given in Equations 4, 5 and 6. However, this interpretation relies on comparisons made across experiments and more research is clearly needed.

It is also worth noting that while the index of performance concept satisfactorily describes the information content for a one dimensional task, if we wish to talk about information processing in three dimensions then the information content of task performance should presumably relate to the ratios of the target volume to the workspace volume, not to the linear distances (which is implicit in MacKenzie and Buxton [8]).

Acknowledgements

This work was funded though a grant from the National Sciences and Engineering Research Council of Canada.

References

1. Arthur, K., Booth, K.S. and Ware, C., (1993) Evaluating 3D Task Performance for Fish Tank Virtual Worlds. ACM Transactions on Information Systems. 11(3) 239-265.

2. Deering, M. (1992) High Resolution Virtual Reality. Proceedings of SIGGRAPH '92. Computer Graphics, 26, 2 (July 1992), pp. 195-202.

3. Fitts, P.M. (1954) The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. Journal of Experimental Psychology, 47, pp. 381-381.

4. Hoffman, E.R. (1992) Fitt's Law with transmission delay. Ergonomics, 35, 1, 37-48.

5. Keele, S.W. and Posner, M.I. (1968) Processing Visual Feedback in Rapid Movements. Journal of Experimental Psychology, 77, pp. 155-158.

6. Liang, J., Shaw, C., and Green, M. (1991) On Temporal-Spatial Realism in the Virtual Reality Environment. In Proceedings of ACM UIST '91 Conference, pp. 19-25.

7. MacKenzie, I.S. (1992) Fitts' Law as a Research and Design Tool in Human-Computer Interaction. Human-Computer Interaction, 7, pp. 91-139.

8. MacKenzie, I.S. and Buxton, W. (1992) Extending Fitts' Law to Two-Dimensional Tasks. In Proceedings of ACM CHI'92 Conference, May 1992, pp. 219-226.

9. MacKenzie, I.S. and Ware, C. (1993) Lag as a Determinant of Human Performance in Interactive Systems. In Proceedings of INTERCHI '93 Conference, pp. 488-493.

10. Pausch, R., Conway, M., DeLine, R., Gossweiler, R., and Miale, S. (1993) ALICE and DIVER: A Software Architect for Building Virtual Environments, In Adjunct Proceedings of INTERCHI '93 Conference, pp. 13-14.

11. Sheridan, T.B. and Ferrell, W.R. (1992) Remote Manipulative Control with Transmission Delay. IEEE Transactions on Human Factors in Electronics, 4, pp. 25-29.

12. Ware, C. Hui, D., and Franck, G. (1993) Visualizing Object Oriented Software in Three Dimensions. In Proceedings of CASCON '93 Conference, IBM Center for Advanced Studies.

13. Ware, C., Arthur, K., and Booth, K.S. (1992) Fish Tank Virtual Reality. In Proceedings of INTERCHI '93 Conference, pp 37-42.

14. Ware, C., and Jessome, D. (1988) Using the Bat: A Six Dimensional Mouse for Object Placement. IEEE Computer Graphics and Applications, 8(5), pp. 41-49.

15. Welford, A.T. (1960) Fundamentals of Skill. London Methuen.