Reaching for Objects in VR Displays: Lag and Frame Rate

COLIN WARE and RAVIN BALAKRISHNAN
University of New Brunswick

This article reports the results from three experimental studies of reaching behavior in a head-coupled stereo display system with a hand-tracking subsystem for object selection. It is found that lag in the head-tracking system is relatively unimportant in predicting performance, whereas lag in the hand-tracking system is critical. The effect of hand lag can be modeled by means of a variation on Fitts' Law with the measured system lag introduced as a multiplicative variable to the Fitts' Law index of difficulty. This means that relatively small lags can cause considerable degradation in performance if the targets are small. Another finding is that errors are higher for movement in and out of the screen, as compared to movements in the plane of the screen, and there is a small (10%) time penalty for movement in the Z direction in all three experiments. Low frame rates cause a degradation in performance; however, this can be attributed to the lag which is caused by low frame rates, particularly if double buffering is used combined with early sampling of the hand-tracking device.

Categories and Subject Descriptors: 1.3.6 [**Computer Graphics**]: Methodology and Techniques —*interaction techniques*

General Terms: Human Factors

Additional Key Words and Phrases: Fitts' Law, Haptics, virtual reality

1. INTRODUCTION

Virtual reality (VR) display systems induce the illusion of a truly three-dimensional graphical scene by coupling the user's eye positions to the graphical image in such a way that the correct perspective view of a three-dimensional object is always maintained. This coupling is achieved by means of a head-tracking system such as the Polhemus Isotrak™, the Bird™, or the Logitech™ tracking device. The position of the user's two eyes are computed from offsets with respect to the measured head position. If the user wishes to manipulate an object in the graphical scene then an image of a hand or a 3D cursor can be coupled to the user's own hand given an input device such as

Funding for this project was provided to the first author under grants from the National Science and Research Council of Canada.
Authors' address: Faculty of Computer Science, University of New Brunswick, P.O. Box 4400, Fredericton, NB, Canada E3B 5A3; email: cware@unb.ca.

the Data Glove™, the Bat [Ware et al. 1993], the Bird™, or the Logitech 3D Mouse™.

In order to create the illusion of "virtuality," or "presence" [Sheridan 1992], it is important that the screen update rate be fast and there be minimal lag in the position-sensing and display systems. Conventional wisdom for computer graphics holds that ten updates per second are required for the perception of smooth motion. However, researchers in the field often state informally that this is not really enough. The purpose of the present study is to obtain some empirical data concerning the effects of lag and frame rate on performance in 3D target selection, and to model them. In order to address this topic there are a number of areas of prior research which must be reviewed: stereo displays with head-coupled perspective, the Fitts' Law paradigm for reaching studies, the effects of lag on performance, and the use of stereo displays in computer graphics. These topics are the basis for the following introduction.

The kind of display we chose to study is one in which a conventional monitor is used to create the VR image which is localized to the region in the vicinity of the screen. Shuttered glasses are used to create field sequential stereopsis, and the user's head position is tracked in real time to ensure that a correct perspective view is obtained. This particular configuration has been called Fish Tank VR [Ware et al. 1993], and it results in a high-resolution virtual image [Deering 1992]. It has been previously shown that for the task of determining connectivity in a data network, head tracking appears to be more important than stereopsis in enhancing the comprehension of 3D information [Arthur et al. 1993; Ware et al. 1993]. However, a much more fundamental task, common to many applications is that of reaching for a target using visually guided hand motion. Target acquisition has been extensively studied in one- and two-dimensional reaching tasks, and many studies have shown that average times can be accurately accounted for using Fitts' Law [Fitts 1954; Keele and Posner 1986; Liang et al. 1991; MacKenzie 1992; MacKenzie and Buxton 1992; MacKenzie and Ware 1993]. It is likely that if this kind of task is carried out in an environment with three-dimensional head-coupled stereo viewing, factors such as the lag in head tracking or hand tracking may influence performance. A recent study by McKenna [1992] showed differences in errors for a reaching task with and without head tracking. But these were not large, and no statistical tests were applied.

2. FITTS' LAW WITH A MODEL OF LAG

Fitts' Law is one of the most successful formulas in human factors research. This law describes the time taken to acquire a visual target using some kind of manual input device. Although there are many variants on Fitts' Law the most commonly used is

$$\text{Mean Time} = C_1 + C_2 \log_2(D/W + 0.5) \qquad (1)$$

where D is the distance to the center of the target, W is the target width; C_1 and C_2 are experimentally determined constants. Fitts' Law was originally derived from information theory, and recently MacKenzie has argued from this perspective that a slight variation on this formula is more satisfying [Liang et al. 1991]. He replaced the 0.5 constant with a 1.0 constant so that the formula becomes:

$$\text{Mean Time} = C_1 + C_2 \log_2(D/W + 1.0). \tag{2}$$

Whichever variant on Fitts' Law is chosen, the value of the logarithmic part of the expression, $\log_2(D/W + 0.5)$ or $\log_2(D/W + 1.0)$, is called the index of difficulty (ID). Thus Fitts' Law can be expressed as

$$\text{Mean Time} = C_1 + C_2 ID. \tag{3}$$

The quantity $1/C_2$ is called the index of performance; the units are bits per second.

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 [Sheridan and Ferrell 1963]. In reality, the hand will not come to a complete stop; instead a series of corrections will be 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 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 Eq. (3) so that it becomes:

$$\text{Mean Time} = C_1 + C_2(C_3 + \text{MachineLag}) ID \tag{4}$$

where C_3 represents the human processing time required to make a corrective movement; MachineLag represents the machine processing time; $C_2 ID$ represents the average number of iterations of the control loop; and C_1 represents the sum of the initial response time and the time required to confirm the acquisition of the target. If an additional sensory or motor processing load is introduced because the human operator is highly stressed (or tired) then any of the human processing components C_1, C_2, or C_3 may be increased. MacKenzie and Ware [1993] found a three-parameter model of this kind to be an excellent description of the data from a one-dimensional Fitts' Law experiment with lag, although they did not interpret it in terms of a control loop. In a much earlier study Sheridan and Ferrell [1963] proposed a similar open loop control model to account for data derived from a task with machine lags of between zero and three seconds.

2.1 2D and 3D Fitts' Law

The classical Fitts Law is a model of one-dimensional movement. MacKenzie and Buxton [1992] tested a number of two-dimensional variations on Fitts' Law on rectangular targets. They found two of these to be successful. In the

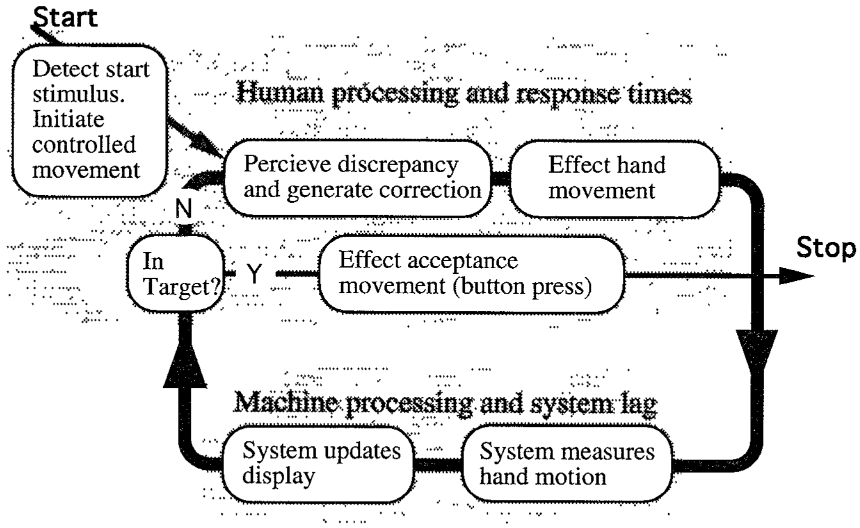


Fig. 1. This diagram shows the control loop assumed to govern guided reaching in a computer graphics environment. It contains components representing machine and human processing operations.

first the index of difficulty was modified by taking target width in the two dimensions into account:

$$ID = \log_2(D/\min(W_1, W_2) + 1.0) \tag{5}$$

where W_1 and W_2 are the target sizes in the X direction and the Y direction respectively, and D is the distance to the center of the target. Essentially this rule states that performance is determined by the smaller of the two target dimensions. This variation on Fitts' Law can be trivially extended to three-dimensional data.

$$ID = \log_2(D/\min(W_1, W_2, W_3) + 1.0) \tag{6}$$

MacKenzie and Buxton's second model also modified the index of difficulty:

$$ID = \log_2(D/W' + 1.0) \tag{7}$$

where W' represents the thickness of the target in the direction of hand motion.

2.2 Effective Target Width

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 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, we see that the actual target width is replaced by 4.13 times the standard deviation of the distribution of hits (representing a 5%

error rate) [MacKenzie 1992; Welford 1960]:

$$ID_\sigma = \log_2(D/4.13\sigma + 1.0) \quad (8)$$

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; however, if the goal is to predict performance in some particular situation, models of performance which include the actual target dimensions may be preferable.

2.3 Lag and the Display Cycle

The basic display cycle used in interactive 3D graphics is as follows. An input device is sampled immediately following the buffer swap. This value is then used to construct the graphical image for the next frame of the display, and after this frame is constructed the buffers are switched at the next available vertical blanking interval. If the image construction time is 100 msec then a minimum of a 100 msec lag occurs before the effects of that input are made visible. That image remains on the screen for another 100 msec. If we assume that perception occurs in the middle of the frame interval then the total lag becomes:

$$\text{MachineLag} = \text{DeviceLag} + \text{FrameInterval}^* 1.5. \quad (9)$$

At 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 210 msec.

While the assumptions in the above estimate are probably reasonable for rapid frame rates they become questionable when the frame rate is low. In this case it is probable that perception of the effect of a movement occurs at some time *before* the middle of the frame interval, and in addition the low rate of sampling the hand position may have adverse effects. For example, at a 1 Hz frame rate an entire corrective movement may be missed. Evidence suggests that the maximum rate of controlled forearm movement is approximately 3 Hz, and the Nyquist theorem requires that to sample this we need at least a 6 Hz sampling rate, preferably more. We will return to these issues in the discussion of Experiment 3 where it is shown that low frame rates can have particularly pernicious effects on performance.

3. STEREOPSIS IN COMPUTER GRAPHICS

A stereo display takes advantage of the ability of the visual system to resolve the differences between the images presented to the two eyes as information about the layout of objects in space. Figure 2(a) shows the simplest possible stereo display. Two lines are spaced differently for the two eyes (the difference in angles α and β subtended at the eyes is called the stereo disparity). Figure 2(b) shows the geometric solution for a layout of the lines in three-dimensional space. Note that a unique solution supposes that the brain also

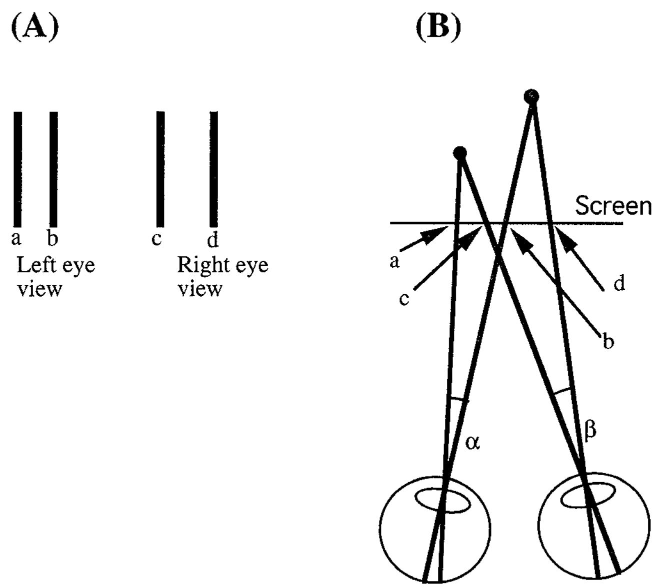


Fig. 2. If the patterns in (A) are shown to the left and right eyes respectively then the result is a perceived layout in space as shown in (B). The points a, b, c, and d represent the projections onto the screen of the vertical lines shown in plan view.

knows the relative orientation of the eyes in their sockets, with special reference to the extent to which they are crossed (vergence). This is important because vergence is coupled to accommodation (depth of focus) in the human visual system, and it poses a problem for VR displays because the only place where the image is actually in focus is at the monitor screen. Objects that are closer or further away than the point of fixation should be out of focus. What this means is that correct vergence and focus information can be provided only for objects in the plane of the screen. (An excellent introduction to human stereo vision is Patterson and Martin [1992].)

3.1 Panum's Fusion Area

If disparities become too large then a single (fused) image is no longer perceived; instead diplopia occurs—the appearance of a double image. However, depth judgments can still be made from a diplopic image, although they will be less accurate [Ogle 1964; Patterson and Martin 1992; Yeh and Silverstein 1990]. The area in which fusion occurs is called Panum's fusion area and is illustrated in Figure 3. As shown, larger disparities can be fused, as distance from the point of fixation increases. At the fovea the maximum disparity before fusion breaks down is only one tenth of a degree, whereas at 6 degrees eccentricity the limit is one third of a degree [Patterson and Martin 1992]. Unless a stereo image is kept in the fusion area diplopia occurs. However, these are worse-case figures, and depending on various spatial and temporal factors the fusion volume will be larger; also depth judgments can

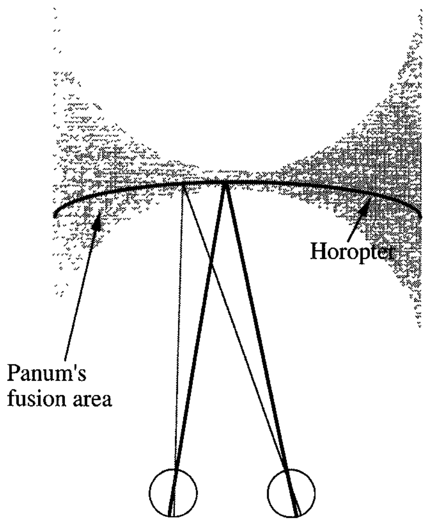


Fig. 3. A smaller disparity can be fused closer to the point of fixation than away from the point of fixation. This area over which fusion takes place is called Panum's fusion area. The horoptor is the locus of constant zero disparity given a particular fixation point.

still be made from a diplopic image, although they will be less accurate [Ogle 1964; Patterson and Martin 1992; Yeh and Silverstein 1990]. Nevertheless, in Experiment 1 we took considerable care to try to minimize diplopia. In Experiment 2 we examined the problem of small target selection under conditions where diplopia did exist.

3.2 Display Resolution in Depth

Display resolution for conventional flat screen displays is computed by the number of pixels per centimeter, typically about 30 for a high-resolution system. Given a viewing distance of 65 cm and an interpupilary distance of about 6.5 cm we can compute the resolution in depth available in a stereo display. Figure 2 illustrates the geometry. The smallest possible horizontal disparity is one pixel which results in a 10-pixel depth difference. Thus, a typical display of this type can be considered as having 30 pixels/cm in the plane of the screen but only 3 pixels/cm in and out of the screen. Antialiasing techniques can increase the effective resolution, but the ten-to-one ratio between horizontal resolution and depth resolution remains in effect at this viewing distance.

This concludes the introduction to Fitts' Law, lag, and stereopsis. The remainder of this article is devoted to a description of three experiments designed to gain an understanding of the important parameters affecting performance in three-dimensional placement tasks. In VR systems some measure of lag in the head-tracking and hand-tracking systems is inevitable; also relatively low image update rates must often be endured. We investigated the following: direction of movement, the effects of lag in the hand-tracking system, the effects of lag in the head-tracking system, target acquisition with flat pizza box targets and with cube targets, the effects of diplopia, and finally the effects of frame rate on performance.

4. EXPERIMENT 1: FITTS' LAW IN 3D (ONE-DIMENSIONAL TASK)

The first experiment had the following two goals.

(1) *Test extended Fitts' Law*: If the lag model described in the introduction is correct then it should account for most of the variance in a variable lag target acquisition experiment.

(2) *Test to see if motion into the screen obeys Fitts' Law*: It is reasonable to presume that there is no significant difference between vertical and horizontal motion in the plane of the screen, and the available evidence supports this. But motion in and out of the screen has to rely on stereopsis and on the lower resolution in depth that is available in a stereo display. It is plausible that when the critical dimension of motion is in and out of the screen target acquisition will be significantly harder. The present study compares horizontal motion (X direction) and motion in and out of the screen (Z direction) to find out if they can be accounted for by the same model.

4.1 Method for Experiment 1

Apparatus (All Three Experiments). The apparatus is illustrated in Figure 4. For all three experiments the visual stimuli were generated using a Silicon Graphics IRIS Crimson with VGX graphics 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). To measure hand position, we used the Bat [Ware and Jessome 1988] (a Polhemus Isotrak™ sensor with a button wired into the mouse). Stereoscopy and tracking of head position was achieved using the StereoGraphics CrystalEyes™ shutter glasses with integral Logitech™ head tracker. All three experiments were conducted entirely in stereo, and the subject's head position was continually tracked in order to provide a correct perspective view. Lag in the hand- and head-tracking devices was introduced by buffering the appropriate device's samples and delaying processing by multiples of the frame rate. 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.

Stimuli. The screen background was set to a dark grey color, and two light grey wire mesh grids were drawn in the horizontal plane at the top and bottom of the screen. The purpose of these grids was to enhance the perception of depth in our VR display. A blue diamond-shaped cursor, 60 pixels wide (measured from two opposing points of the diamond), was coupled to the user's hand via the Bat. The target consisted of two purplish-red, 5 cm square tiles with solid borders (1-pixel wide antialiased lines) 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

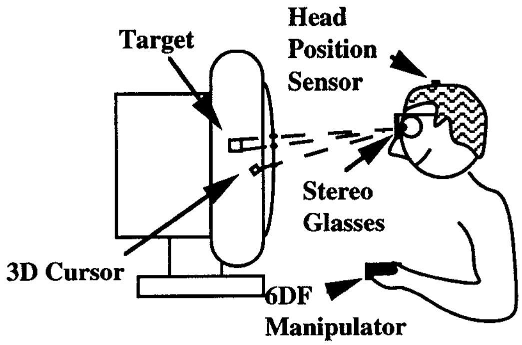


Fig. 4. The apparatus. This photograph shows a subject using the system. All the major components are represented: Head tracking and stereo using CrystalEyes™ VR shutter glasses, Bat input device, the cursor and the target. The subject is closer to the monitor than he would normally be.

relatively slow green phosphor of the monitor. The separation between the tiles varied and represents the width of the target for index-of-difficulty calculations. The targets are shown in Figure 5.

Procedure. There were a total of five different lag conditions which included three levels of head lag and three levels of hand lag as shown below.

Base condition		
Head lag (msec):	114	
Hand lag (msec):	87	
Head Lag conditions		
Head lag (msec):	214	364
Hand lag (msec):	87	87
Hand Lag condition		
Head lag (msec):	114	114
Hand lag (msec):	187	337

The actual lag was measured using the method described in the Appendix. Performance was evaluated for both horizontal motion (X direction) and motion into the screen (Z direction). This results in $5*2 = 10$ different direction-lag combinations. Since we wished to carry out a Fitts' Law analysis for each, subjects were tested using three target distances (4, 8, and 16 cm) and two target widths (2 and 4 cm). This yields a total $5*2*3*2 = 60$ conditions. There were 10 trials per condition structured in the manner described below.

The experiment was conducted over two one-hour sessions on separate days. At the start of each session, the subject received a practice set of blocks consisting of all possible lag, direction, and distance-width combinations but with no repetitions. Following this, subjects were presented with ten blocks of trials, one for each direction-lag combination. A block consisted of 32 trials, five trials for each of the six distance-width combinations, together with two practice trials given at the start of each block to familiarize the subject with

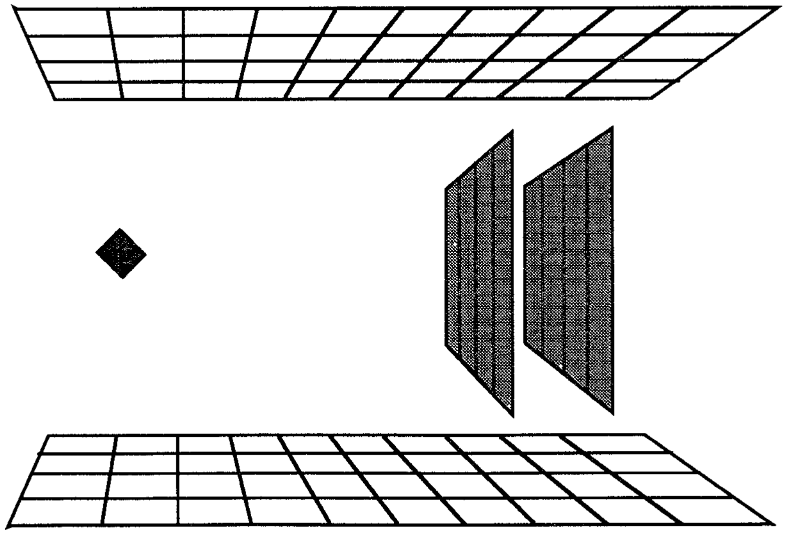


Fig. 5. The target and the cursor used for Experiment 1.

that particular lag and direction. Ignoring the practice trials, we see that the result is 30 trials per block, $10*30 = 300$ trials per session, and $2*300 = 600$ 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 and in the plane of the screen. The target then appeared 0.33 sec later to the right of the cursor by the appropriate distance for that trial (measured from the center of the cursor to the center of the target). In the Z direction, the cursor appeared 8 cm in front of the center of the screen, and the target appeared behind the cursor (i.e., going into the screen) by the appropriate distance. In both directions, the front face of the target was perpendicular to the cursor in the X and Z directions respectively. Therefore, although the user moves in three-dimensional space the task is essentially one dimensional because of the flattened nature of the target.

The subject completed a trial by pressing the button on the Bat, which had the effect of binding the xyz position of the hand to the start position of the cursor, moving the cursor into the box bounded by the target's two tiles 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.

Subjects. Twelve computer-literate subjects from the authors' university served as paid volunteers. Three of the subjects had prior experience with the apparatus used in the experiment.

4.2 Results for Experiment 1

We found no significant effects of head lag by an analysis of variance $F(2, 22) = 1.58$. Performance in the Z direction was 9% slower than in the X direction overall. However, this effect just failed to reach significance at the 5% level $F(1, 11) = 4.47$. To understand the effects of task difficulty and lag on performance, we ran a set of regressions using the three-coefficient model given by Eq. (4) (this assumes that lag will have a multiplicative effect on the index of difficulty).[1]

The regression results for the hand lag conditions were as follows:

In the X direction:

$$\text{Mean Time} = 1.42 + 1.67(0.106 + \text{lag})ID \qquad r^2 = 0.90$$

In the Z direction:

$$\text{Mean Time} = 1.57 + 1.16(0.253 + \text{lag})ID \qquad r^2 = 0.90$$

X and Z combined:

$$\text{Mean Time} = 1.49 + 1.41(0.166 + \text{lag})ID \qquad r^2 = 0.86$$

The plot shown in Figure 6 shows the mean response times plotted against index of difficulty for the three hand lag conditions (X and Z values combined). The overall index of performance for the above data is $1/(1.41^* 0.166) = 4.3$ bits per second which is in the range cited by MacKenzie [1992].

Although the difference between the estimated human processing times (0.106 for X direction and 0.253 for Z) are markedly different we note that these are highly sensitive to noise in the data, a point which is confirmed by the fact that a high-regression coefficient is obtained from the combined X and Z data. The major difference in performance between the two directions is that there is a broader distribution of hits in the Z direction which caused the error rates for Z direction performance to more than double. This data is given in Table I which also shows that error rates increase with lag.

4.3 Discussion of Experiment 1

In general these data are reasonably consistent with previous Fitts' Law studies that have used a similar task (albeit in only one direction). The estimated human processing time of 166 msec is consistent with previous estimates of between 100 and 200 msec [Carleton 1981; Keele and Posner 1968]. If the lag is set to zero then the information processing rate becomes 4.27 bits per second which is fairly typical for Fitts' Law studies. The

[1] We also analyzed the data for all three experiments both with and without the modified index of difficulty (Eq. (8)). We decided in the end to present only the data analyzed using the unmodified index of difficulty for two reasons: (1) the unmodified ID accounts for more of the variance and (2) the unmodified ID can be used to predict actual performance. As mentioned in the introduction the modified index of difficulty is only arrived at after a post hoc analysis of the distribution of hits.

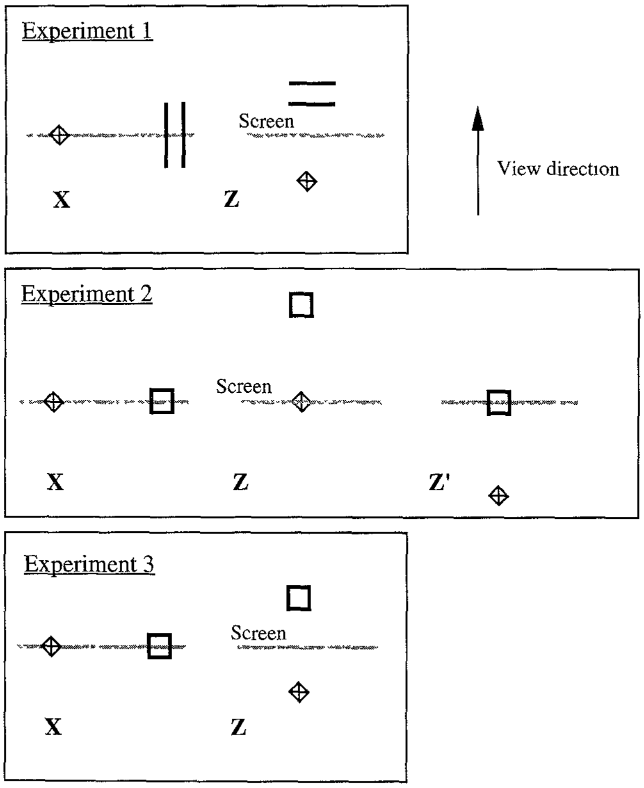


Fig. 6. This diagram shows a schematic plan view diagram summarizing the conditions for all three experiments.

Table I. Percentage Errors for the Different Hand Lag Conditions in the X and Z Directions

	87 msec lag	187 msec lag	337 msec lag
X direction	0.28	1.1	2.50
Z direction	2.64	4.03	4.58

estimated lag multiplier is about 40% larger than that found previously by MacKenzie and Ware [1993].

We believe the task constraints were largely responsible for the lack of any performance degradation due to head lag. In the current placement task subjects tended not to move their heads much; presumably the stereo depth cues were sufficient to give an adequate perception of depth information.

The finding that errors were much larger in the Z direction shows that movement in and out of the screen is not isomorphic with movement in a horizontal direction. This could be due to the lower (stereo) resolution in and out of the screen, described in the introduction.

The most significant overall finding is that the performance decrement due to lag is given by multiplying the system lag by 1.4 times the index of difficulty. Thus for selection of a small target (ID = 5.0) a lag of 200 msec will cause a simple selection to take 1.5 seconds longer than it would without lag. In many highly interactive systems target selection is a fundamental building block of the interface, and this kind of performance degradation may easily make the difference between a system that is perceived as useful and one that is not.

5. EXPERIMENT 2

The second experiment had the following two goals:

(1) *Test extended Fitts' model for 3D cube targets*: Whereas Experiment 1 was designed to be a task for which only one dimension of movement was critical (either X or Z), Experiment 2 was designed to investigate the problem of the capture of three-dimensional targets which are small in all three dimensions. According to both of MacKenzie and Buxton's preferred models (Eqs. (5) and (7)) there should be no difference between the capture of a 3D cube and the capture of a box-shaped object flattened in the direction of movement, so long as the sizes in the direction of motion are the same [MacKenzie and Buxton 1992]. Our initial pilot work suggested to us that this was not in fact the case, and so we undertook to investigate the matter in a formal experiment in which the targets were cubes of different sizes.

(2) *Measure performance under conditions of diplopia*: The first experiment was designed to minimize the occurrence of double images (diplopia). However, in many situations diplopia will occur because the binocular disparity is too great, and it is important to determine if this is a significant factor in target acquisition times.

5.1 Method for Experiment 2

Stimuli. The target was changed to a cube with solid borders (1-pixel wide antialiased lines) and translucent faces. The back face of the cube, respective to the direction of movement, was made more opaque than the other five faces. This served as an aid in determining when the cursor had penetrated the back face and was no longer inside the target. The cursor width was reduced to 0.43 cm because the smallest target was a 0.5 cm (approximately 18 pixels) cube.

Procedure. The target acquisition task was performed in the X direction and in two variations in the Z direction (see Figure 7). As in Experiment 1, at the start of a trial in the X direction, the cursor appeared 8 cm to the left of the center of the screen and in the plane of the screen while the target appeared to the right of the cursor by the appropriate distance for that trial. In the first variation in the Z direction, henceforth referred to simply as 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. This did not cause diplopia. In the second variation, henceforth referred to as the Z′ direction, the target appeared in the center

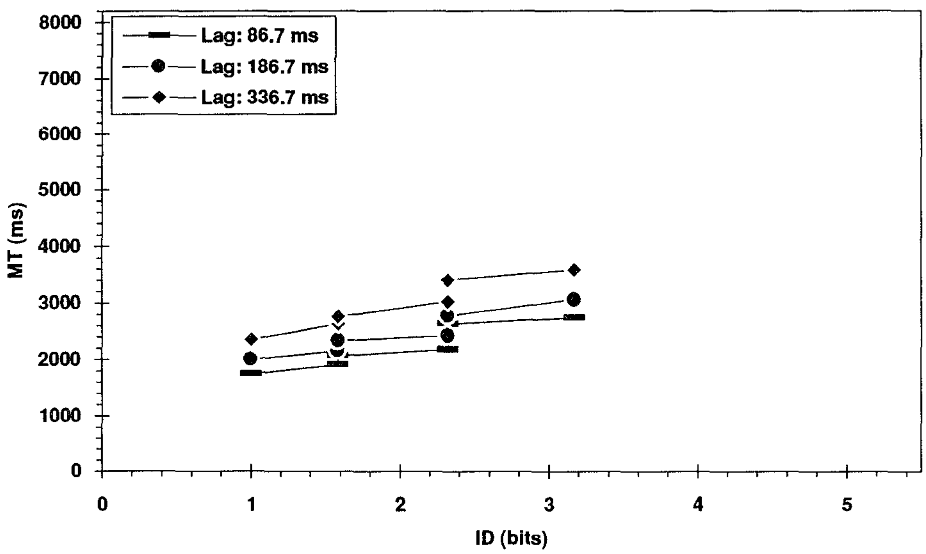


Fig. 7. The averaged results from Experiment 1. Mean time to respond is plotted against index of difficulty for all three lag conditions.

and in the plane of the screen and the cursor appeared in front of the target (i.e., coming out of the screen) by the appropriate distance. When the distance was large, the cursor appeared diplopic.

Three levels of hand lag (87, 187, and 337 msec) were investigated in all three directions. Head lag was the lowest possible: 114 msec. This resulted in $3^*3 = 9$ different lag-direction combinations. For each lag-direction subjects were tested with two target distances (4 and 16 cm) and three cube sizes (0.5, 1, and 2 cm) resulting in six distance-size combinations. The experiment was conducted in a similar manner to Experiment 1 with eight trials per experimental condition. Since there were only nine different lag-direction conditions, subjects were presented with nine blocks of trials per session, for a total of $9^*24 = 216$ trials per session and $2^*216 = 432$ trials per subject.

Target selection and timing was performed in an identical manner to Experiment 1. The experiment was carried out over two one-hour sessions with practice sessions and blocks of trials randomized in a manner similar to that used for Experiment 1.

Subjects. Twelve computer-literate subjects from the authors' university served as paid volunteers. Seven of the subjects had prior experience with the apparatus used in the experiment.

5.2 Results for Experiment 2

On our initial analysis the data from Experiment 2 showed large departures from the classical Fitts' Law relationship and anomalous regression coefficients. However, closer examination of the data revealed that the anomalies could be traced to the data obtained with the 0.5 cm cubic target. These conditions contained very high error rates (17% on average), and our experience observing the subjects suggested an extreme difficulty in task performance. In retrospect this is not entirely surprising given that the depth disparities for a half centimeter are less than two pixels (see Introduction), and that our input device had an inherent noise of approximately 0.25 cm in the region where we used it. We therefore excluded these data from subsequent analysis.

We performed an analysis of variance between the X, Z, and Z′ conditions which showed a significant main effect for the X, Z, and Z′ directions, $F(2, 22) = 4.9$. However an analysis of variance comparing the diplopia conditions (Z and Z′) revealed no significant effect $F(1, 11) = 1.58$. Overall, performance in the Z and Z′ directions was 9% slower than performance in the X direction, as was found for Experiment 1. Overall these results are consistent with a degradation in performance due to direction but none due to diplopia. As in Experiment 1 we ran regressions using the model given by Eq. (4).

In the X direction:

$$\text{Mean Time} = 1.48 + 1.52(0.221 + \text{lag})ID \qquad r^2 = 0.95$$

In the Z direction:

$$\text{Mean Time} = 1.65 + 1.54(0.237 + \text{lag})ID \qquad r^2 = 0.96$$

In the Z′ direction:

$$\text{Mean Time} = 1.32 + 1.44(0.277 + \text{lag})ID \qquad r^2 = 0.95$$

All three combined:

$$\text{Mean Time} = 1.48 + 1.50(0.276 + \text{lag})ID \qquad r^2 = 0.95$$

The surprising result here is that the combined r^2 value is nearly as high as the individual values. The overall index of performance for the above data is $1/(1.50 * 0.276) = 2.4$ bits per second which is considerably lower than that found for the first experiment.

Figure 8 shows the mean response times plotted against index of difficulty for three lag conditions (X, Z, and Z′ values combined). In this plot the excluded 0.5 cm target points are shown but not connected to the other points. The error data (excluding 0.5 cm targets) is given in Table II which shows no consistent effect for direction.

5.3 Discussion of Experiment 2

The use of targets that were symmetric in the X and Z conditions can account for the finding that errors did not vary in the X and Z conditions as they did in Experiment 1.

The fact that diplopia had no effect is good news for users of this kind of display because diplopia cannot be avoided given a reasonable depth to the image space.

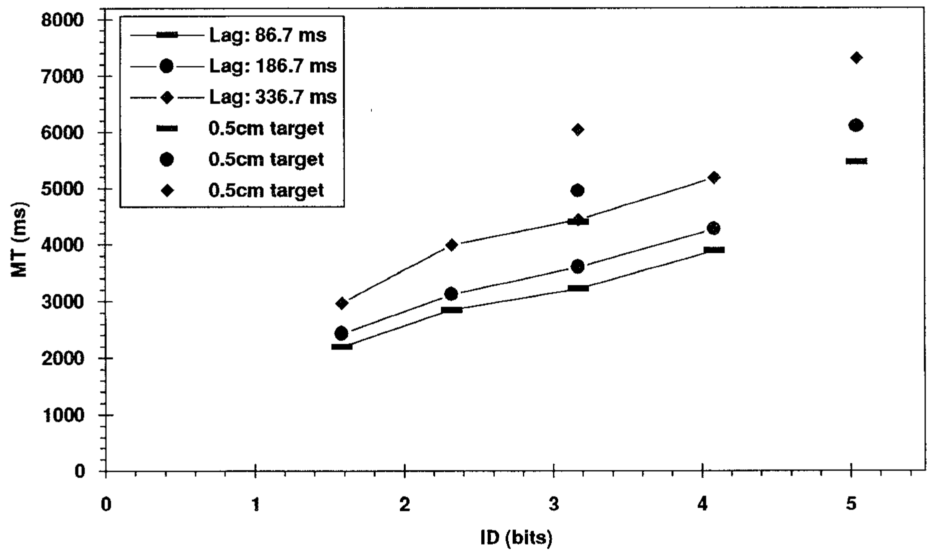


Fig. 8. The averaged results from Experiment 2. Mean time to respond is plotted against index of difficulty for all three lag conditions. The points obtained with the 0.5 cm targets are shown not connected to the other points. Due to high error rates these values were excluded from the data analysis.

Table II. Percentage Errors are Given for the Different Hand Lag Conditions in the X, Z, and Z′ Directions

	87 msec lag	187 msec lag	337 msec lag
X direction	2.86	0.26	4.69
Z direction	4.43	2.86	5.73
Z′ direction	3.65	3.65	2.65

While we cannot be clear about the causes of the problems with the 0.5 cm targets, it appears likely that the difficulty of holding the unsupported hand steady, noise in the device, and the problems of stereo resolution of the front and back target surfaces all contributed. The four to seven seconds required to make a selection is inordinately long for such a simple task, suggesting that such targets should be avoided.

The reduced bit rate as compared to Experiment 1 suggests that the simple generalization from one-dimensional selection to three-dimensional selection given by Eqs. (5) or (6) are not adequate. However, not much weight should be given to comparisons made across experiments.

6. EXPERIMENT 3: THE EFFECTS OF LOW FRAME RATE

The third experiment had the following goal:

Test effects of frame rate and lag on performance.

One of the major causes of lag in interactive animation systems is the practice of double buffering. As explained in the Introduction, a lag is introduced which is one and a half times the frame interval under reasonable assumptions.

It seems likely that low frame rates will disrupt task performance; the question of theoretical interest which the present study addresses is whether the performance decrement can be attributed to the lag caused by double buffering or whether there is some additional performance decrement which can be attributed simply to the low frame rate.

6.1 Method for Experiment 3

Stimuli. The background stimulus was identical to that of Experiments 1 and 2. The target and cursor were identical to that of Experiment 2.

Procedure. The base condition with minimal hand lag was combined with 17 other conditions in which hand lag was introduced in three different ways. Head lag was 97 msec throughout.

In this experiment lag was introduced 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.
(2) *Early sampling*: In this condition lag was manipulated by varying the frame rate. The device was always sampled immediately after the buffers were swapped.
(3) *Late sampling*: In this condition lag was manipulated by varying the frame rate. The device was always sampled 1/60th of a second prior to a buffer swap. The graphical image of the cursor and the target was constructed in the ensuing 1/60th sec interval.

Note: Between Experiments 2 and 3 we removed a source of delay in the device driver, resulting in a shorter lag in the best case.

Base Condition: 70 msec. (frame interval = 16.7 msec)

High frame rate: 5 conditions
frame rate = 60 Hz
frame interval = 16.7 msec

hand lag (msec):	137	187	337	537	787		

Early sampling (normal double buffering): 5 conditions

frame rate (Hz):	15	10	5	3	2		
frame interval (msec):	67	100	200	333	500		
lag (msec):	145	195	345	545	795		

Late sampling (double buffering with late sampling): 7 conditions

frame rate (Hz):	15	10	5	3	2	1	0.666
frame interval (msec):	67	100	200	333	500	1000	1500
lag (msec):	95	112	162	228	312	562	812

Each condition was evaluated for both the X and the Z directions. This resulted in $18*2 = 36$ different lag-direction combinations. There were only two distances (4 and 8 cm) and one size (1 cm) resulting in two distance-size combinations and a total of $36*2 = 72$ conditions. The experiment was conducted in a similar manner to Experiment 1 with ten trials per experimental condition resulting in 720 trials per subject. Practice sessions were given as in Experiments 1 and 2.

The target acquisition task was performed in the X and Z directions. As in Experiments 1 and 2, at the start of a trial in the X direction, the cursor appeared 8 cm to the left of the center of the screen and in the plane of the screen while the target appeared 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 front of the screen, and the target appeared behind the cursor (i.e., going into the screen) by the appropriate distance.

Target selection and timing was performed in an identical manner to Experiments 1 and 2.

Subjects. Twelve computer-literate subjects from the authors' university served as paid volunteers. Eight of the subjects had prior experience with the apparatus used in the experiment.

6.2 Results for Experiment 3

Figure 9 shows averaged target acquisition times with both early and late sampling of the hand-tracking device. This shows clearly an overall advantage for late sampling as should be expected. Overall, the data showed that performance in the Z direction was 10% slower than that in the X direction $F(1, 11) = 10.7$.

The following regression values were obtained for the various conditions applying the model given in Eq. (4):

High frame rate data

In the X direction:

$$\text{Mean Time} = 0.78 + 1.66(0.189 + \text{lag})ID \qquad r^2 = 0.90$$

In the Z direction:

$$\text{Mean Time} = 1.25 + 1.80(0.120 + \text{lag})ID \qquad r^2 = 0.97$$

Early sampling data

In the X direction:

$$\text{Mean Time} = 0.98 + 1.80(0.130 + \text{lag})ID \qquad r^2 = 0.99$$

In the Z direction:

$$\text{Mean Time} = 0.630 + 2.01(0.211 + \text{lag})ID \qquad r^2 = 0.98$$

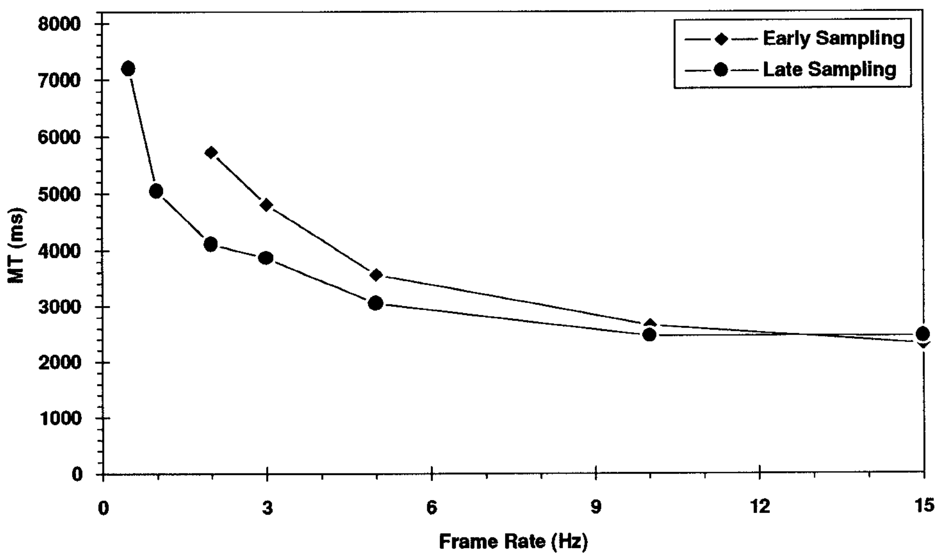


Fig. 9. Data from Experiment 3. The mean response times are plotted against frame rate for both early and late device sampling conditions.

Late sampling data

In the X direction:

$$\text{Mean Time} = 0.480 + 2.29(0.204 + \text{lag})ID \qquad r^2 = 0.97$$

In the Z direction:

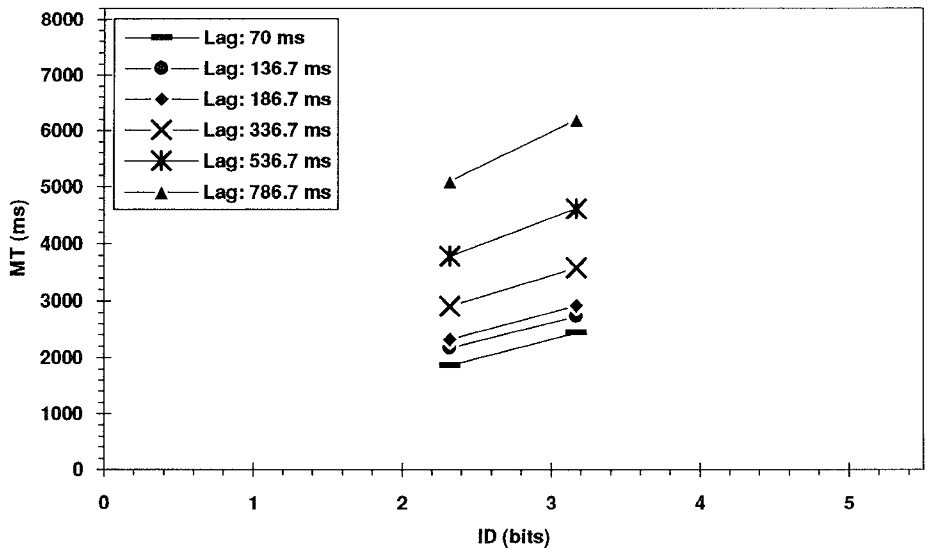
$$\text{Mean Time} = 0.241 + 2.32(0.292 + \text{lag})ID \qquad r^2 = 0.96$$

All data combined

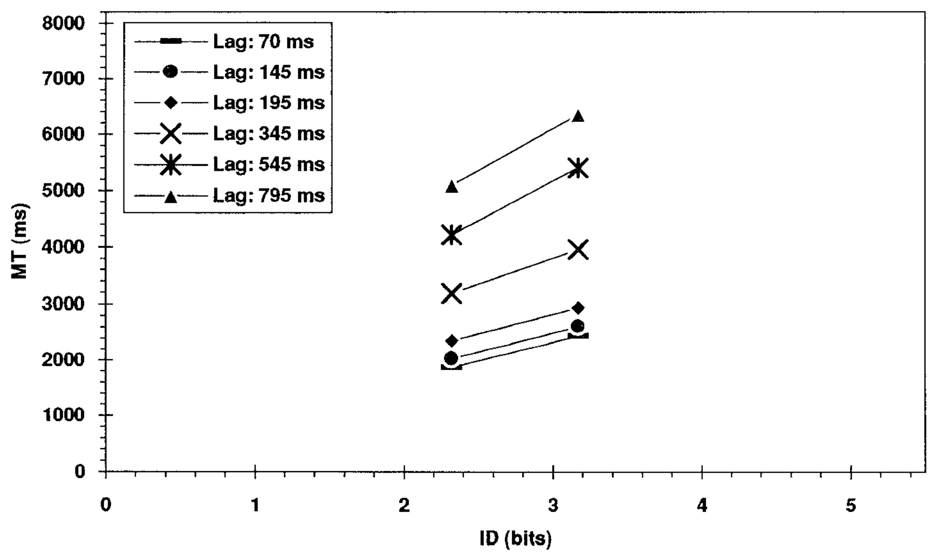
$$\text{Mean Time} = 0.739 + 1.95(0.209 + \text{lag})ID \qquad r^2 = 0.89$$

The plots shown in Figure 10 illustrate the mean response times plotted against index of difficulty for 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 the same as that found for Experiment 2 and again considerably lower than that found for the first experiment.

The real test of the model from Eq. (4) is how well a single regression equation accounts for the data from all three sets of conditions. As can be seen above when we combined three sets of conditions the overall value for r^2 dropped to 0.89. This is still a respectable value, but we decided to reevaluate one of our assumptions to see if we could do better. This is the assumption (Eq. (9)) that an image is perceived at the middle of the frame of 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

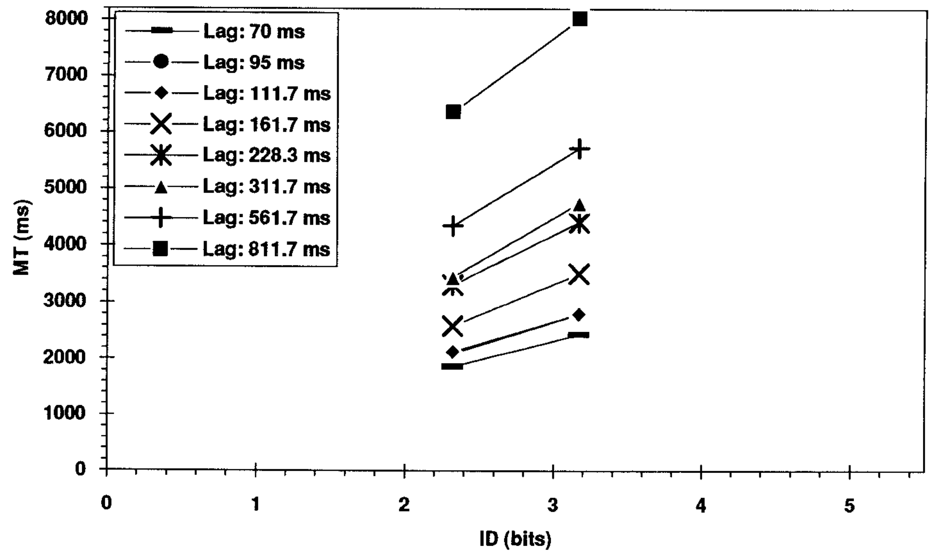


(a)



(b)

Fig. 10. (a) The averaged results from Experiment 3 in the hand lag conditions. In these conditions lag was introduced by queuing device values. (b) In these conditions lag was introduced by reducing the frame rate and sampling the device immediately after a buffer swap. (c) In these conditions lag was introduced by reducing the frame rate and sampling the device 1/60th of a second before a buffer swap.



(c)

Fig. 10. Continued.

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. Based on this observation she makes a movement toward the target. However the movement is only sampled at the beginning of the next frame. Thus the feedback loop can, in effect, have an additional lag to take into account 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.

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 11, 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:

All data combined

$$\text{Mean Time} = 0.739 + 1.59(0.266 + \text{lag})ID \qquad r^2 = 0.95$$

6.3 Discussion of Experiment 3

This last experiment contained more levels of lag and collected more data than the other two. Therefore our best estimate of the detrimental effect of lag is 1.59 multiplied by the index of difficulty. It is worth noting that there is

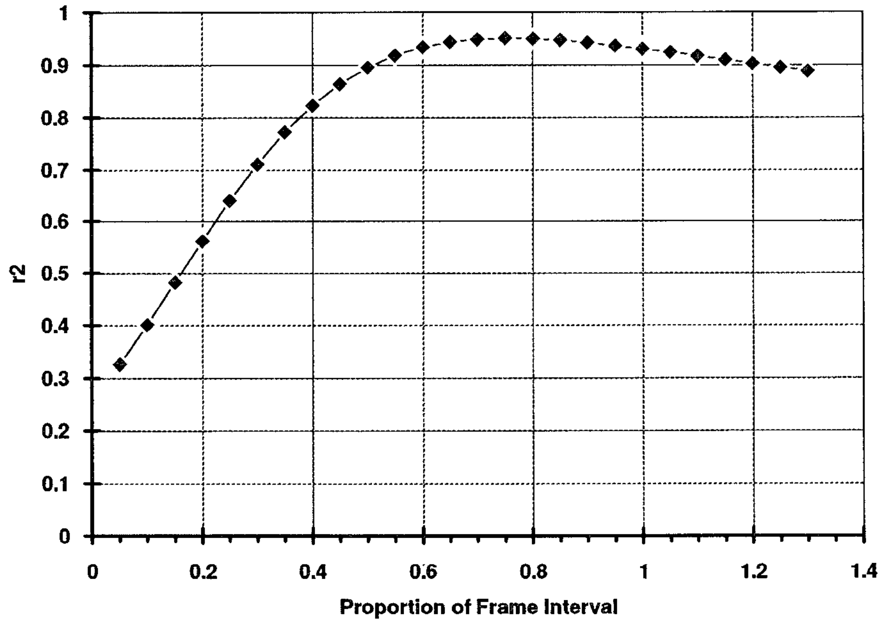


Fig. 11. Regressions were computed for the entire set of data from Experiment 3 with adjustments in the estimation of machine lag.

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 could have used our revised estimate of the machine lag to reanalyze the results from the first two experiments, but we felt that this would be taking post hoc analysis too far. Also, since the frame rates were always high for the first two studies the change would have only resulted in a change of 4 msec (0.25/60) in the estimated machine lag.

7. CONCLUSION

We have discovered that system lag introduced between the movement of an input device and visual feedback is a major factor in reducing the speed of target selection.

To a first crude approximation the simple formula

$$\text{Mean Time} = C_1 + 1.59(\text{HumanProcessing} + \text{MachineLag})ID$$

accounts for most of our data. Experiment 3 suggests that the best method for estimating MachineLag is

MachineLag = DeviceLag + FrameInterval* 0.75
+ time between sampling of the device and the buffer swap if double buffering is used in the main rendering loop.

The HumanProcessing constant in the above formulation represents the time to initiate a visually guided movement correction in the control loop illustrated in Figure 1. The results from our study are consistent with previous studies in suggesting that this value is between 0.1 and 0.25 seconds. C_1 will depend on the particular task since it represents a combination of initial reaction time to start the task and the time taken to terminate the task, for example, by means of a button press. *ID* represents an index of task difficulty as defined according to Fitts and modified by MacKenzie and Buxton [1992].

The other factors we investigated, namely, lag in the head-coupling system, the effect of low frame rates (independent to the lag introduced), and the direction of hand motion had relatively minor effects on performance. The most significant of these, movement in the Z direction caused a consistent 9–10% performance decrement in all three experiments compared to movement in the X direction. We also found evidence for higher error rates for motion in the Z direction.

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

(1) Acquire input devices which have low lag, ideally less than 50 msec. Note that even this small lag can cause an 8% or more performance cost when selecting small targets.
(2) If double buffering is used, keep the frame rate up. For example, at a frame rate of 10 Hz an effective lag of 175 msec is introduced, and this could add 1.2 sec to target selection times when selecting small targets.
(3) 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 at this point 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.
(4) If possible create higher update rates for the target and the cursor (and hence lower lags). Pausch et al. [1993] recently described a software architecture that supports this kind of decoupling.
(5) Avoid designing systems that require the acquisition of small targets with the unsupported hand.

With respect to the issue of whether 3D target acquisition is essentially different than 2D (or 1D) target acquisition, our data suggests that there is a difference. The index of performance values were considerably lower for the cube target than they were for the pizza box target which means that neither of the simple extensions to Fitts' Law given by MacKenzie and Buxton (and described in the Introduction) can be valid. However, this interpretation relies on comparisons made across experiments; more substantial evidence would come from a single experiment that combined the conditions. Nevertheless, the low bit rates and the very substantial acquisition times suggests that reducing a three-dimensional task to a one-dimensional task is not satisfactory for the purposes of modeling. It is also worth noting that while the index of performance describes the information content for a one-dimensional task satisfactorily, 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 (this is implicit in MacKenzie and Buxton [1992]).

With respect to the issue of lag in the head-position sampling affecting performance, we found no effect of this variable. However, we feel that this result only applies to the Fish Tank VR situation that we used for these studies. In full-immersion VR with head-mounted monitors, changes in head orientation, would, for example, result in dramatic changes in the scene that do not occur in Fish Tank VR. These changes, coupled with lag, would be likely to handicap performance. However, we are not equipped to evaluate this possibility.

Last, one of the reviewers of this article commented that the use of predictive filters on both hand and head sampling is widespread, and that the effects of these filters on task performance is unknown. This is clearly an important topic for further research as there is a distinct possibility that in some circumstances (e.g., where the sampling rate is low) these filters may cause a degradation in task performance.

APPENDIX

Measurement of Lag

In studies of this type, it is essential to measure accurately the actual system lag. We used a modified version of the method developed by Liang et al. [1991] to measure the lag for both the Polhemus Isotrak which we used for hand tracking and the Logitech ultrasonic sensor which we used for head tracking. We designed a stepper motor-driven pulley assembly (Figure 12) which sat on top of the computer monitor. The sensor (the Polhemus and Logitech in turn) was attached to the belt driven by the stepper motor and was moved back and forth across the monitor screen at a constant speed. The monitor displayed a graphic ruler and a cursor which reflected the position reported by the sensor (we only used one dimension of the 3D position information). A video camera recorded both the movement of the sensor across the monitor and the graphic image displayed on the screen. The video tape was later played back frame by frame, and we recorded the difference in position between the physical sensor and the reported position as displayed by the graphic cursor. Since we knew the amplitude and velocity of the sensor, we could calculate the lag from this displacement. The use of a computer-controlled stepper motor to move the sensor, instead of a pendulum

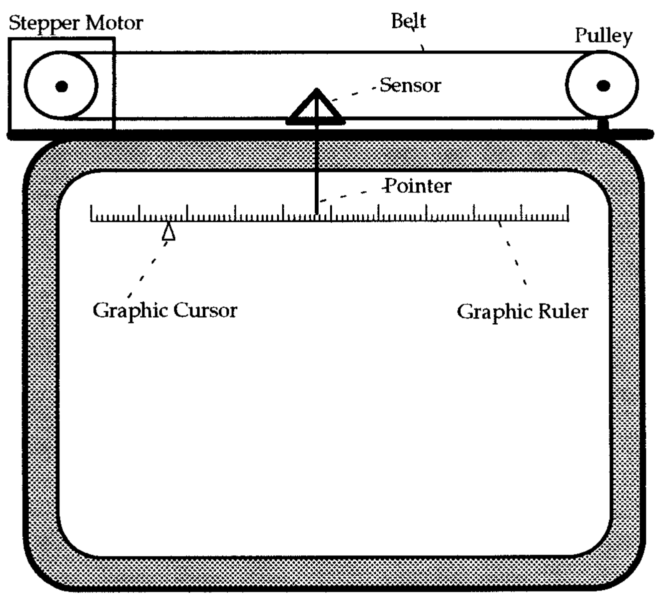


Fig. 12. The apparatus used to measure lag in the system.

as used by Liang et al., ensured a constant predetermined linear velocity which reduced the possibility of errors in our calculations.

In order to ensure that the lags measured using this technique reflected accurately the lags in our three experiments, the program used for calibration closely resembled the software used in those experiments: the device drivers were implemented using the same shared-memory client-server architecture; double buffering was used throughout, and a screen update rate of 60 Hz was maintained. The Polhemus was used in continuous binary mode with default filter parameters, and a baud rate of 19.2 K. The Logitech was used in demand-reporting mode also at 19.2 K baud. Not filtering was done with the Logitech.

We found the device lags to be

—45 msec for the Polhemus Isotrak
—72 msec for the Logitech

exclusive of lags introduced by double buffering, etc. The lags that actually occurred in the context of the experiments are given in the method sections to the three experiments.

We are grateful to an anonymous reviewer who pointed out that because the gain of the Polhemus device actually depends on the frequency of the movement [Adelstein et al. 1992] our calibration was not complete. Unfortunately, it is not at all clear how this information will affect human performance characteristics for the reaching task, and this is therefore an uncontrolled factor in the experiments.

ACKNOWLEDGMENTS

We are grateful to Mark Paton for help with the device driver code.

REFERENCES

ADELSTEIN, B. D., JOHNSTON, E., AND ELLIS, S. R. 1992. A testbed for characteristic dynamic response of virtual environment spatial sensors. In *Proceedings of UIST'92* (Monterey, Nov.). ACM Press, New York, 15–22.

ARTHUR, K., BOOTH, K. S., AND WARE, C. 1993. Evaluating 3D task performance for Fish Tank virtual worlds. *ACM Trans. Inf. Syst. 11*, 3, 239–265.

CARLETON, L. G. 1981. Processing visual feedback for movement control. *J. Exp. Psychol. Hum. Percep. Perf. 7*, 5, 1019–1030.

DEERING, M. 1992. High resolution virtual reality. In *Proceedings of SIGGRAPH '92. Comput. Graph. 26*, 2, 195–202.

FITTS, P. M. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol. 47*, 6, 381–381.

KEELE, S. W. AND POSNER, M. I. 1968. Processing visual feedback in rapid movements. *J. Exp. Psychol. 77*, 1, 155–158.

LIANG, J., SHAW, C., AND GREEN, M. 1991. On temporal-spatial realism in the virtual reality environment. In *Proceedings of ACM UIST '91*. ACM, New York, 19–25.

MACKENZIE, I. S. 1992. Fitts' Law as a research and design tool in human-computer interaction. *Hum. Comput. Interact. 7*, 1, 91–139.

MACKENZIE, I. S. AND BUXTON, W. 1992. Extending Fitts' Law to two-dimensional tasks. In *ACM CHI'92 Conference Proceedings*. ACM, New York, 219–226.

MACKENZIE, I. S. AND WARE, C. 1993. Lag as a determinant of human performance in interactive systems. In *INTERCHI '93 Conference* (Amsterdam, May). ACM Press, New York, 488–493.

MAYER, D. E., ABRAMS, R. A., KORNBLUM, S., WRIGHT, C. E., AND KEITH SMITH, J. E. 1988. Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychol. Rev. 95*, 3, 340–370.

MCKENNA, M. 1992. Interactive viewpoint control and three-dimensional operations. In *Proceedings 1992 Symposium on 3D Graphics. Comput. Graph. 26*, 1, 53–56.

OGLE, K. N. 1964. *Binocular Vision*. Hafner, New York.

PATTERSON, R. AND MARTIIN, W. L. 1992. Human stereopsis. *Hum. Factors 34*, 6, 669–692.

PAUSCH, R., CONWAY, M., DELINE, R., GOSSWEILER, R., AND MIALE, S. 1993. ALICE and DIVER: A software architect for building virtual environments. In *INTERCHI '93 Adjunct Proceedings*. ACM Press, New York, 13–14.

SHERIDAN, T. B. 1992. Musings on telepresence and virtual presence. *Presence 1*, 1, 120–125.

SHERIDAN, T. B. AND FERRELL, W. R. 1963. Remote manipulative control with transmission delay. *IEEE Trans. Hum. Factors Elec. 4*, 25–29.

WELFORD, A. T. 1960. *Fundamentals of Skill*. Methuen, London.

WARE, C. AND JESSOME, D. 1988. Using the Bat: A six dimensional mouse for object placement. *IEEE Comput. Graph. Appl. 8*, 5, 41–49.

WARE, C., ARTHUR, K., AND BOOTH, K. S. 1993. Fish tank virtual reality. In *Proceedings of INTERCHI '93 Conference on Human Factors in Computing Systems*. ACM Press, New York, 37–42.

YEH, J. J. AND SILVERSTEIN, L. D. 1990. Limits of fusion and depth judgement in stereoscopic color displays. *Hum. Factors 32*, 1, 45–60.

Received September 1993; accepted July 1994