THREE DIMENSIONAL TRACKING IN AUGMENTED ENVIRONMENTS: USER PERFORMANCE TRADE-OFFS BETWEEN SYSTEM LATENCY AND UPDATE RATE

Stephen R. Ellis, Anthony Wolfram, and Bernard D. Adelstein
Ames Research Center, Moffett Field, CA 94035-1000

Three-dimensional tracking performance was measured as a function of system latency (35-335 msec) and update rate (10-30 Hz). Twelve subjects used a custom, see-through head mounted stereo display to control the position of a virtual response cursor with hand and body movements. User performance trade-offs between latency and update rate were measured with objective and subjective measures and a possible performance model was evaluated. The results indicate that earlier findings suggesting that latency influenced tracking performance more than did update rate, could be due to previous studies having tested latency over a larger dynamic range. Iso-performance contours are used to compare objective performance with subjective perception and performance judgments.

INTRODUCTION

Excessive response latency in interactive systems is well known from classical studies with analog CRT and mechanical displays (e.g., Sheridan & Ferrell 1963; Poulton, 1974) to disturb users' target tracking performance. In contemporary virtual environment (VE) systems, even those not including detailed simulations of all the moving elements of the environment, excessive latency and insufficient update rate are the principle dynamic characteristics that can disturb user interaction. (Kim, Ellis, Hannaford, Tyler, & Stark, 1987; Liu, Tharp, French, Lai, & Stark, 1993; Barfield & Hendrix, 1995; Ware & Balakrishnan, 1994) These two types of dynamic disturbances are mathematically distinct and are typically traded off against each other in interactive graphics systems. For example, in Silicon Graphics (SGI) Performer-based multi-processing graphics, high update rates are achieved through parallelizing within the graphics pipeline hardware, which because of staging and buffering introduces delay. Since user interaction is disturbed by long latency and low update rates, there is interest in understanding the cost of such parameter trade-offs.

Due to prior computer system limitations, our previous efforts to study this trade-off were restricted to a maximum update rate of 20Hz and minimum system latency of 80 msec (Ellis, Bréant, Menges, Jacoby, & Adelstein, 1997a; Ellis, Dorighi, Menges, Adelstein, & Jacoby, 1997b; Ellis, Adelstein, Jense, Baumeler, & Jacoby, 1999a). These earlier dynamic difficulties have been resolved in our present computer system, allowing this trade-off to now be studied more fully. Additionally, our current system makes possible an improved search for equivalence classes between objective and subject measures of tracking and simulation fidelity. This search has been proposed previously (Ellis, 1996), but, heretofore, efforts were stymied by the inability to measure smooth objective response surfaces from reported objective and subjective data. We previously suggested that for a subjective performance scale such as presence to be most useful, the response surface formed when simulation parameters are traded off against each other should be geometrically similar for both objective and subjective measures. Comparisons of experimentally measured objective and subjective response surfaces are reported below.

Our previous attempts to measure the relative importance of latency and update were based on comparison of linear correlations between each of these factors and RMS tracking error, an objective measure of tracking performance (Ellis et al., 1997ab, Ellis et al. 1999a). These studies suggested that latency was a more important factor; however, our inability to completely span the desired range of latencies and update rates restricted our observations. Now with better control over these variables, we are able to conduct a previously impossible full factorial design to assess the relative performance impact of the two factors on large amplitude three-dimensional tracking. Distinct from other reports of three-dimensional tracking (e.g., Kim et al., 1987), this study employs unconstrained head, body, and hand movement and uses a see-through display format to minimize pixel fill requirements, thereby maximizing system's temporal performance.

METHODS

Displays

The custom-made see-through HMD used in this study combines the LCD and controls of a Virtual Research V8, custom 50% see-through Virtual Vision optics, and a custom back-light (max luminance = ~40 cd/m^2 in this experiment), and has a Michelson contrast of 0.4-0.7. Virtual objects in the experiment could to be seen easily in a normally lit room. The display allows focus, interpupillary distance, and binocular overlap (15% to 100%) adjustment. Binocular overlap was set at 50%, yielding a total display field of view (FOV) of 40° with pixel resolution ~2.5 arcmin/pixel. The HMD when balanced on a user's head and attached to its cables weighs < 1.3kg.

Simulation

The underlying simulation on an SGI ONYX computer (4 CPU, dual pipe RE-2 graphics) was created by AuSim Inc. (Mountain View, CA) using Sense8, Inc.'s WorldToolKit software. Graphics complexity and system overhead were managed so that the simulation could maintain stable update rates up to 60 Hz. Head and hand positions were each sampled at 120 Hz with dual synchronized Polhemus FasTraks interfaced to the computer via custom software drivers (Jacoby, Adelstein, & Ellis, 1996). Minimum system latency was tuned to 35 ± 5 msec (Mean ± SE). FOV and stereo parameters were subjectively adjusted based on reference targets. These specifications make the simulation system used for this experiment superior to that used in similar previous experiments and lead us to believe the present results are more definitive.

Tracking task

The subject's task was to position a virtual blue tetrahedral cursor (20 cm/side) entirely within a virtual target cube (25 cm/side). The target was programmed to move irregularly in three dimensions within a 1 m cubic space in front of the subjects (Figure 1). When the cursor was entirely within the cube, the cube would change color from red to green. Target position was driven separately in each axis by a sum of 8 randomly phased nonharmonic sinusoids (0.02-0.93 Hz) to provide moderate tracking difficulty—i.e., a mean RMS tracking error spanning ~0.2-1.0 when normalized with respect to the target motion.

Cursor position was controlled by a Polhemus sensor attached to a 50 cm wand held in the subjects' dominant hand. Three frame rates (10, 20, 30 Hz) and four latency conditions (35, 135, 235, 335 msec) were crossed to create 12 different conditions. Controlled latencies were accurate to ±5 msec and frame rates were timed to be stable. Previous work has shown that 16 msec is near the threshold for latency perceptibility (Ellis, et al., 1999ab) during normal head and hand movements. Thus, our selection of latency values ranges from that producing just visible disturbances to that shown to shift operators' tracking from a continuous to a "move and wait" technique (Sheridan, 1992).

All participants were given 20 minutes of familiarization with the equipment as well as four one-minute practice runs before data collection began. Participants were blind to the specific experimental conditions during each trial. The tracking task was presented for 70 sec intervals, with the first 10 sec of data ignored as a warm-up period. One minute breaks were enforced following each pair of data collection periods allowing participants to rest their dominant arm. Every four runs, longer five minute breaks were enforced. There were two 70 sec runs per condition, presented consecutively for a total of 12 randomly sequenced pairs, yielding 24 distinct one minute tracking periods to be analyzed.

After each pair of runs with a given condition, subjects were asked to rate various perceptual and performance attributes of the tracking task according to seven bipolar adjectival scales (Ellis et al., 1999a). Additionally, they rated the controllability of the tracking by an adapted Cooper-Harper scale (Cooper & Harper, 1969; Ellis et al., 1999a). Because no subjects reported, nausea, neck or eye pain, headache, or eye tearing, only the scales concerning the subjective realism of the cube, its subjective stability in space, and user's sense of dizziness were analyzed.

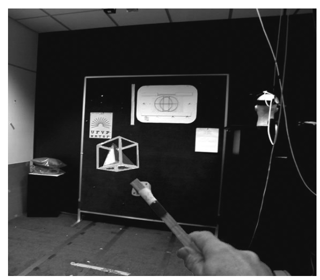

Figure 1. The tracking task is illustrated here as seen from the subjects' viewpoint. The virtual target (cube) is shown with the virtual response cursor (tetrahedron) roughly centered within it by user movement of the Polhemus FasTtrak receiver attached to the end of the wand.

	Update	Latency	Norm_RMS	C+H	Cube reality	Spatial stability	Dizziness instability
Update	1.000						
Latency	0.000 (0.000)	1.000					
Norm. rms	**-0.426** **(-0.447)**	**0.773** **(0.481)**	1.000				
C+H	**-0.479** **(-0.601)**	**0.662** **(0.415)**	**0.810** **(0.568)**	1.000			
Cube reality	-0.131 (-0.171)	0.082 (-0.011)	0.118 (0.036)	0.274 (0.296)	1.000		
Spatial stability	-0.235 (-0.249)	0.334 (0.182)	**0.409** (0.156)	**0.552** **(0.565)**	**0.685** **(0.735)**	1.000	
Dizziness instability	-0.088 (-0.106)	0.064 (0.045)	0.280 (0.120)	0.205 (0.271)	0.196 (0.198)	**0.414** **(0.492)**	1.000

Table 1. Rank order correlations (corrected for ties). Spearman correlation was selected due to the ordinal nature of subjective ratings. Table notes: 10 dof/subject X 12 Subjects = 120 dof; experiment wise type-I error set to 0.05 with 21 independent correlation tests; Critical 2-tailed $\rho = 0.398$. Significant correlations are in bold. Correlations in parentheses were calculated without the 35 msec latency case to balance the ranges of latency and update rate used for the analysis.

RESULTS

A correlation table based on *all* of the individual data was constructed evaluating the linear relations between dependent and independent variables. This table in general replicates earlier studies showing the effect of latency on tracking to be stronger than that of update rate. The adapted Cooper-Harper scale also correlates most strongly with objective performance, i.e., the normalized RMS (nRMS) tracking error.

ANOVAs were conducted for each of the dependent variables using a 3 X 4 repeated measures factorial analysis. No main effects of replication or any of its interactions were statistically significant. The overall average nRMS for the first and second replications were 0.500 and 0.493 respectively (SE = ±0.115). Thus, for the purposes of this experiment, subject's performance were asymptotic.

Of the five dependent variables kept for analysis, four showed statistically significant effects in the ANOVAs. Normalized RMS tracking error (Figure 2) was significantly affected by update, latency and their interaction. These results were verified after a log transform (Figure 3) to correct for deviations from homogenous variances in nRMS across conditions. (Latency: $F(3,33) = 25.72$, $p < 0.0005$; after log transform $F(3,33) = 116.85$, $p < 0.0005$) (Update $F(2,22) = 37.21$, $p < 0.0005$; after log transform $F(2,22) = 230.34$, $p<0.0005$) (Update X Latency $F(6,66) = 7.616$, $p < 0.0005$; after log transform $F(6,66) = 5.29$, $p<0.002$).

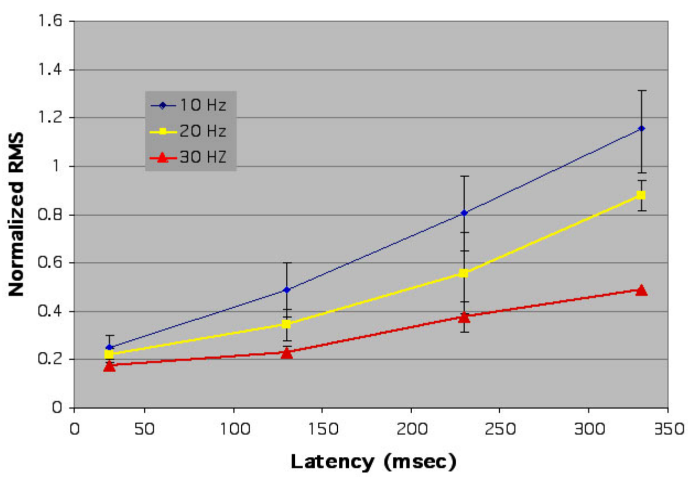

Figure 2. Interaction between latency and update rate based on normalized RMS tracking error: (Error bars = ±1SE)

The Adapted Cooper-Harper (C+H) scale (Figure 4) showed a full complement of main effects and interaction.
Because it was only taken once per subject at the end of the second block, repetition was not a factor in this analysis. (Update: $F(2,22) = 90.241$, p<0005; Latency: $F(3,33) = 63.380$, p < 0005; Update X Latency: $F(6,66) = 4.086$, $p < 0.002$).
make same order as nRMS Latency, Update, interaction

The Subjective Stability scale (Figure 5) also showed a full complement of main effects and interaction. However, because it also was only taken once, repetition was not a factor in this analysis. (Update: $F(2,22) = 8.174$, $p < 0.002$; Latency: $F(3,33) = 7.962$, $p < 0.001$; Update X latency: $F(6, 66) = 4.086$, $p < 0.002$).

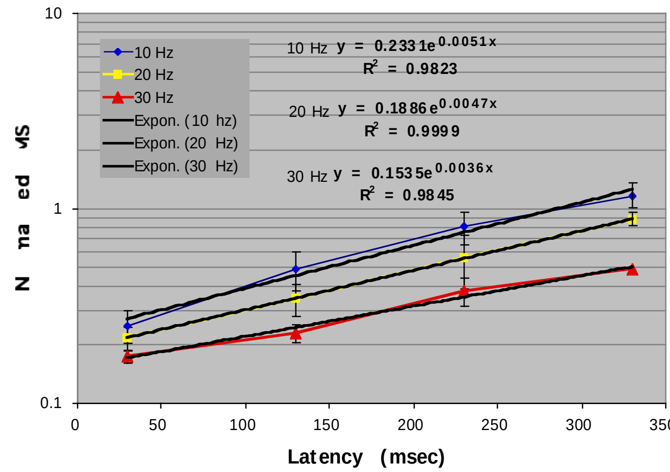

Figure 3. Interaction between latency and update rate based on normalized RMS tracking error: logarithmic transform (Error bars = ±1SE)

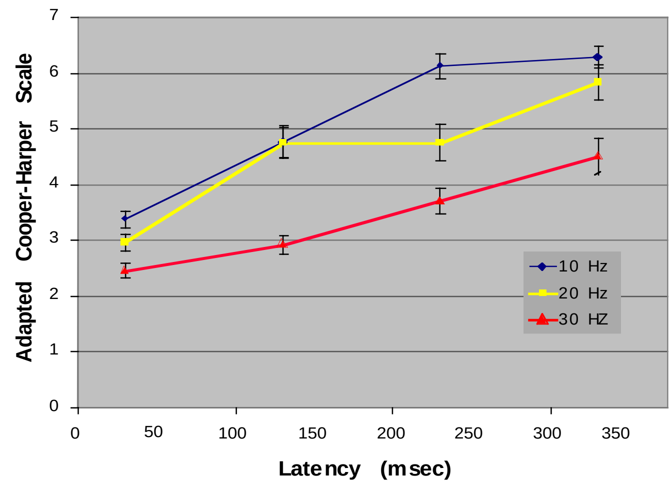

Figure 4. Interaction between latency and update rate based on an adapted Cooper-Harper controlability scale.

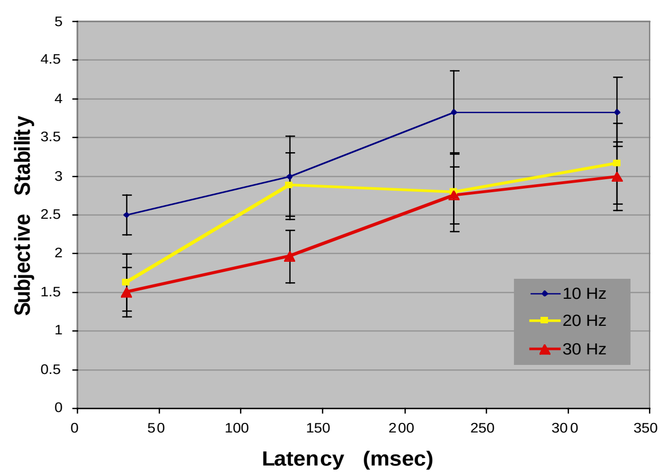

Figure 5. Interaction between latency and update rate based on subjectively judged stability of the tracked target (Error bars = ±1SE)

The Sense of Dizziness or Postural Instability scale had only one weakly significant effect from update rate and consequently will not be discussed below. (Update: $F(2,22) = 3.513$, $p < 0.047$). Subjective Cube Realism showed no statistically significant effects and therefore it too will not be discussed.

DISCUSSION

The intercorrelations between RMS objective and subjective performance measures with latency and update rate strongly resemble previous reports from our laboratory (e.g., Ellis et al, 1999a). By excluding the 35 msec latency from the analysis, both latency and update rate could be balanced, as was not previously possible, to a 3:1 dynamic range for a fair comparison of their correlations with tracking error. This filtering markedly reduced the difference in the magnitude of the nRMS correlations with update rate and latency (Table 1, parenthesized values) though generally supported all other otherwise significant results.

Also, note that now, after dynamic range correction, update rate has relatively higher correlation with the adapted Cooper-Harper scale than latency. In our previous study (Ellis et al., 1999a), the equally high correlation of latency and update rate with the Cooper-Harper scale appeared to be in part due to the greater latency range tested.

The log transform was introduced to correct for inhomogeneity of variance in the normalized RMS tracking data. However, as the approximate parallelism of the semi-log plots in Figure 3 illustrates, the subjects' tracking errors for different update rates are merely proportionate to each other across all latency values, a key feature that will need to be reflected in any tracking model to be fit to these data.

The systematic and smooth effect of both latency and update rate variables on tracking performance makes possible the comparison of objective and subjective response surfaces contemplated earlier (Ellis, 1996). These are plotted in Figure 6 and show the strikingly uniform contours of the nRMS error response surface.

It is interesting to compare the response surfaces for the different dependent measures. The objective RMS tracking error shows a qualitatively uniform performance trade-off for the entire variable range while the subjective measures indicate irregular and variable trade offs. Interestingly, the Cooper-Harper surface (Figure 7) shows two unique regions: one in which changes of latency do not effect performance whereas update rate changes do. In contrast there is also a region in which changes in update rate would not effect performance whereas latency effects would be present. These regions are indicated by text in the figure. It is precisely this kind of conflicting relationship between the subjective rating and objective performance that would make the rating an unsuitable basis for an underlying explanatory construct.

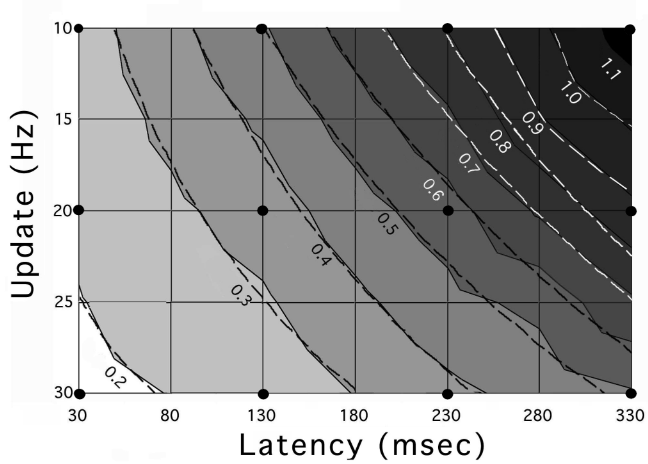

Figure 6. Iso nRMS performance contours based on interpolated surface through measured points (black dots). Dashed lines for respective nRMS levels represent equations fitted to contours.

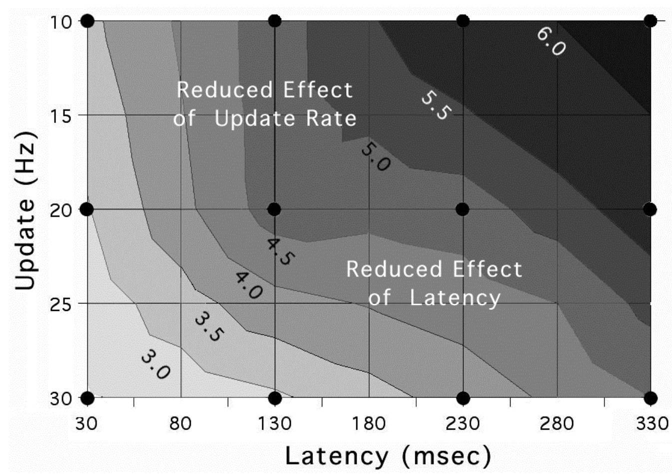

Figure 7. Iso Cooper-Harper scale contours based on interpolated surface passing through measured points.

These conflicting relationships between the object and subjective measures arise from the lack of isomorphism in the response surfaces from the two dependent variables At the very least, the conflicts show that the linear correlations in Table 1 do not capture the full interrelation between the objective and subjective dependent measures. Consequently, neither of the subjective measures reported here provides a basis for defining performance equivalence classes that were argued to be a desirable feature for subjective scales (Ellis, 1996).

Comparisons of the sort illustrated by Figures 6-8 might be interesting if applied to other subjective scales such as presence or workload. Figure 6 provides a convenient summary of the performance trade-off between latency and update rate that motivated the present experiment. For any given level of tracking performance, one can move along the iso-performance contour to find which update rate-latency pairing may be selected to preserve overall performance. For example, if designers want to maintain interactive performance at a level corresponding to 0.5 nRMS tracking for a system with a ~200 msec latency and 20 Hz update rate that in some circumstances might degrade to 10 Hz, they could reduce system latency to ~130 msec in order to maintain comparable performance.

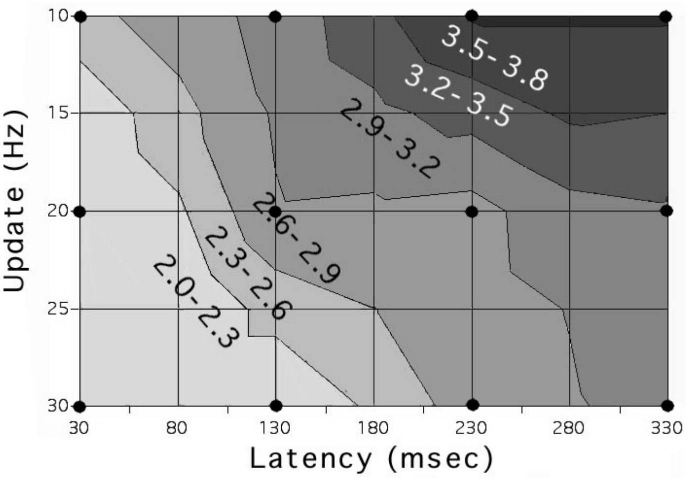

Figure 8. Iso subjective cube stability contours based on interpolated surface passing through measured points.

The systematic appearance of the response surface in Figure 6 invites description and modeling. The approximate parallelism evident in Figure 3, for example, indicates that tracking performance in the different update rate conditions differ by the same proportion for all latency levels.

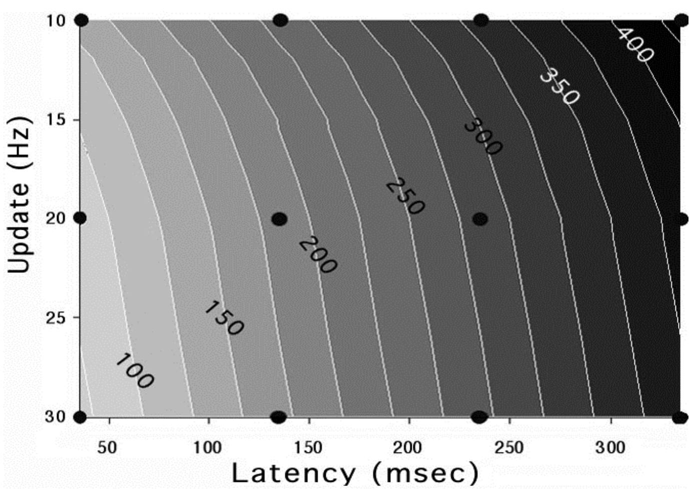

Figure 9. Composite latency (in msec) based on update period plus latency shows a response surface somewhat similar to that observed for nRMS but with reversed curvature. Dots represent points where performance was actually measured in the experiment

We have found empirically that the iso-performance contours can all be plotted by functions of the form $y = a \log(x) + b$, where a and b are themselves linearly dependent and $a = 84.113*(\text{nRMS}) + 6.6664$ and $b = (a + 45.24) / 2.82$. We are beginning to develop an analytic basis to predict the shape of the contours. An initial theory could, for example suggest that the subjects' performance would be a function of the composite delay, d_c, arising both the latency, L, and sample and hold characteristics due the discrete update rate u. Since $d_u = 1 / u$ and is equal to the sampling period, we can propose a simple linear function, $d_c = d_u + L$, as model. Figure 9 shows the contours expected if the response surface of a dependent variable such as nRMS were to be this linear function of d_c. Though the general features of the response surface are captured by this model, it clearly does not incorporate the curvature of the empirical contours and therefore can be only a first step in a modeling effort.

REFERENCES

Barfield, W., & Hendrix, C., (1995) The effect of update rate on the sense of presence within virtual environments, *Virtual Reality: Res., Dev., Appl., 1*(1), 3-15.

Cooper, G.E., & Harper, R.P. Jr. (1969) The use of pilot ratings in the evaluation of aircraft handling qualities. NASA TN-D-5153.

Ellis, S.R. (1996) Presence of mind: a reaction to Sheridan's musings on telepresence. *Presence, 5,* (2), 247-259.

Ellis, S.R., Dorighi, N.S., Menges B.M., Adelstein B.D., & Jacoby R.H., (1997a) In search of equivalence classes in subjective scales of reality, in M.J. Smith, G. Salvendy, and R.J. Koubec, eds., *Design of Computing Systems, Advances in Human Factors Ergonomics, vol 21B*, Elsevier.city,pp. 873-876.

Ellis, S.R., Bréant, F., Menges, B.M., Jacoby, R.H., & Adelstein, B.D. (1997b) Factors influencing operator Interaction with virtual objects viewed via head-mounted see-through displays: viewing conditions and rendering latency. *Proc., Virtual Reality Ann. Internat. Symp*. pp. 138-145

Ellis, S.R., Adelstein, B.D., Baumeler, S., Jense, G.J., & Jacoby, R.H., (1999a) Sensor spatial distortion, visual latency, and update rate effects on 3D tracking in virtual environments. *Proceedings, IEEE Virtual Reality '99*, pp. 218-221.

Ellis, S.R., Young, M.J. , Ehrlich, S.M. , & Adelstein, B.D. (1999b) Discrimination of changes of latency during voluntary hand movement of virtual objects. *Proceedings, HFES*, pp. 1182-1186.

Kim, W.S., Ellis, S.R, Hannaford, B., Tyler, M., & Stark, L. (1987) A quantitative evaluation of perspective and stereoscopic displays in three axis manual tracking tasks. *IEEE Trans. on System Man and Cybernetics,* SMC-17, 61-71.

Jacoby, R.H., Adelstein, B.D., & Ellis, S .R. (1996) Improved temporal response in virtual environments through system hardware and software reorganization. *Proceedings, SPIE: Stereoscopic Displays and Application, 2653*, pp. 271-284.

Liu, A., Tharp G., French, L., Lai, S., & Stark, L. (1993) Some of what one needs to know about using head mounted displays to improve teleoperator performance," *IEEE Trans. Robotics and Automation, 9*(5), 638-648.

Poulton, E.C. (1974) *Tracking skill and manual control*, Academic Press, N.Y., 199-206.

Sheridan, T. B., & Ferrell, W.R. (1963) Remote manipulative control with transmission delay. *IEEE Trans. Human Factors in Electronics*, HFE, *4*(1), 25-29.

Sheridan, T.B. (1992) *Telerobotics, Automation and Human Supervisory Control,* MIT Press.

Ware C. & Balakrishnan R., (1994.) Reaching for objects in VR displays: lag and frame rate, *ACM Trans. Comp. Hum. Interact., 1*(4) 331-356.

Watson, B., Spaulding, V., Walker, N., & Ribarsky, W. (1997) Evaluation of the effects of frame time variation on VR task performance, *Proc., Virtual Reality Ann. Internat. Symp.*, pp. 38-44.

Acknowledgements. This research was partially supported by ONR's VIRTE and NASA's AOS programs.