In IEEE Transactions on Systems, Man, And Cybernetics – part A: Systems and Humans. Vol. 27, No.4, July 1997, pp518-528.

Anisotropic Human Performance in Six Degree-of-Freedom Tracking: An Evaluation of 3D Display and Control Interfaces

Shumin Zhai Paul Milgram Anu Rastogi

ABSTRACT

Motivated by the need for human performance evaluations of advanced interface technologies, this paper presents an empirical evaluation of a 3D interface, from the point of view of both display and control, in a pursuit tracking experiment. The paper derives methods for decomposing tracking performance into six dimensions (three in translation and three in rotation). This dimensional decomposition approach has the advantage of revealing overall performance levels in the depth dimension relative to performance in the horizontal and vertical dimensions. With interposition, linear perspective, stereoscopic disparity and partial occlusion cues incorporated into a single 3D display system, subjects' tracking errors in the depth dimension were about 45% (with no practice) to 35% (with practice) larger than those in the horizontal and vertical dimensions. It was also found that subjects initially had larger tracking errors along the vertical axis than along the horizontal axis, likely due to their attention allocation strategy. Analysis of rotation errors generated a similar anisotropic pattern. By applying the dimensional decomposition method, the paper also analyses the issue of coordinated control of 6 degrees of freedom with one hand. It was found that when the subject could not control all 6 degrees of freedom well, translational aspects of the task were given higher priority than the rotational aspects. After 40 minutes of practice more than 80% percent of subjects were able to control both translational and rotation aspects together.

INTRODUCTION

Three dimensional (3D) human machine interfaces, including 3D displays and multiple degree-of-freedom (DOF) controllers, are being applied increasingly in areas such as teleoperation, virtual environments, data visualization, and computer aided design. Hence a need arises to empirically evaluate human performance with these interfaces. This paper presents one such endeavour to investigate some of the human factors issues associated with 3D displays and 6 DOF input controls.

With respect to 3D visual displays, a major challenge is to provide sufficient depth information to the user. Human perception is sensitive to a variety of depth cues, including occlusion, binocular disparity, perspective, shadows, texture, motion parallax and active movement [1-3]. Many techniques have been developed for realising such depth cues in computer graphic displays (see for example [4] for a review of implementation methods). Different approaches have been proposed for evaluating the effectiveness of these techniques. One conventional empirical approach is to manipulate the presence or absence of various depth cues and measure task performance as a function of each combination of depth cues (e.g. [5-7]). This approach reveals the strengths of the depth cues relative to each other and potential interaction effects between them; however, it does not provide an estimation of the overall efficiency of particular display cues. An alternative approach is to compare how task performances in each dimension compare with each other, by decomposing behavioural data collected in the presence of all available depth cues into separated horizontal (X), vertical (Y) and depth (Z) components. Using such an approach, Massimino, Sheridan and Roseborough [8] found that, with a conventional 3D display comprising perspective projection cues only, tracking errors in the Z dimension were approximately 400% greater than those in the X and Y dimensions. Performance in the depth dimension can be expected to improve as more depth cues are added, but it is also expected to remain somewhat poorer than performance in the horizontal or vertical dimensions. Few studies have been carried out to determine the extent of that remaining difference, however. One of the objectives of the present study was to examine these differences using a display system comprising a collection of readily implementable 3D cues: interposition, perspective, binocular disparity and partial-occlusion.

There is also some reason to believe from the general literature on visual perception that performance differences may exist between the horizontal and vertical dimensions. For instance, in a task that required nursery school children to reproduce lines on a circular background, Berman, Cunningham and Harkulich [9] found that reproductions of the vertical lines were

[

[

significantly more accurate than reproductions of the horizontal and the oblique lines, as measured by orientation differences. Gottsdanker and Tietz [10] found that subjects tended to be more sensitive in judging relative length in the horizontal direction than in the vertical direction. Following up on this issue, another objective of the present study, therefore, was not only to compare performance in the Z dimension relative to the X and Y dimensions, but also to evaluate performance differences between the X and the Y dimensions. In other words, the generalized objective here was to examine *(an)isotropies,* or *asymmetries,* of performance in the X, Y, and Z directions.

There are two approaches to evaluating (an)isotropy in 3D performance. One is to let subjects perform the same task in each of the three dimensions separately and compare the X, Y and Z performances. Two of the foreseeable problems with this approach are (1) The interactions and integration of the three dimensions are missing. Subjects' behaviour and performance in one dimension at a time may not be the same as those in an integrated 3D environment. (2) The particular order of the task dimensions performed may have an effect on the final results, due to asymmetrical skill transfer resulting from learning [11, 12]. The second approach, which was the method used in the present study, is to request subjects to perform an integrated manipulation task in a 3D environment and subsequently decompose the task performance into its dimensional elements for analysis of the (an)isotropies with respect to the 3D displayed information.

No less important than the problem of how to display depth in 3D human-machine interfaces is the issue of how to enable control of systems possessing several (e.g. six or more) degrees of freedom. Much work has been done in designing hand controllers and input techniques to allow efficient translational and rotational manipulation in 3D space (see [13-15] and [16] for reviews). However, the number of published evaluations carried out to compare human performance with various types of multi-DOF controllers is limited. In fact, even such basic issues as whether and how well humans can handle all six degrees of freedom together have not been satisfactorily resolved. In the teleoperation literature, for example, there have been some concerns about whether human operators are able to control all six degrees of freedom with only one hand. Rice, Yorchak and Hartley [17] observed that controlling 6 DOF with one hand is difficult, due to unwanted cross coupling between axes. Some teleoperation systems, such as the Shuttle Remote Manipulator, also known as the "Canadarm", explicitly require two-handed operation, with one hand for rotation control and the other for translation control. O'Hara [18] contradicted Rice's observation, however, and found no differences between two 3 DOF controllers versus one 6 DOF controller. McKinnon and King [19] further argued that a single 6

DOF hand controller should be preferable to control distributed among separate controllers. In conclusion, it is only with empirical data that such issues can be resolved. Decomposition of task performance into six components is one way to provide insights into the issue of 6 DOF controllability with one hand.

METHOD

Experimental Task

Pursuit tracking with 6 degrees of freedom was chosen as our evaluation task. The long tradition of using tracking as an experimental paradigm for studying human skills in motor control research (e.g. [20-22]) has more recently taken root in the teleoperation and virtual environment community. For example, Kim, Ellis, Tyler, Hannaford and Stark [23] examined 3D displays with a three axis tracking task; Massimino and colleagues [8] studied 6 DOF tracking with one hand; Ellis, Tyler, Kim and Stark [24] studied the effects of display-control axis misalignment; and Tachi and Yasuda [7] studied tracking in the depth dimension with various displays.

In the present experiment, subjects were asked to continuously control a 3D cursor and align it as closely as possible in both position and orientation with respect to a 3D target that moved unpredictably in 6 DOF within a virtual environment (see Figure 1). Both the tracking target and the controlled cursor were tetrahedral. Each tetrahedron had two blue adjacent edges and the remaining edges were colored red. The two blue coloured edges ensured that only one possible correct orientation match existed.

The cursor and the target differed in two ways. First, the radius (from the centre to any of the vertices) of the target tetrahedron was 3.55 graphics units[1] while the radius of the cursor was 4.62 graphic units, or 1.3 times as large as the target. Second, the cursor had semi-transparent surfaces while the target was rendered as a wireframe model only. These two differences were

[

[1] In this paper, all lengths are given in terms of graphic units as defined in the display program, where 1 graphic unit = 1.4 cm on the particular display screen used in this experiment.

introduced in order to minimize any potential confusion between the identity of the subject controlled cursor and the independently moving target. As discussed later, the semi-transparent surfaces also served as an important visual cue to locate the cursor relative to the target.

Motion of the target in the experiment was driven by six different independent forcing functions, one for each degree of freedom. Each of the forcing functions consisted of a weighted combination of 20 different sine functions with a random initial phase (similar to the forcing function used in [7]). Such forcing functions are often used in tracking research, since they produce smooth motion which, from the subjects' point of view, is unpredictable, i.e., subjectively perceived low-pass random noise (Poulton, 1974).

(a)

(b) (c)

Figure 1. *6 DOF tracking task. The tetrahedron with semi-transparent surfaces was the controlled cursor. The tetrahedron without semi-transparent surfaces was the randomly moving target· The subjects' goal was to align the cursor with the target. Shown in the figure are examples of (a) a very larger error between the cursor and the target (b) a large translation error and small rotation error, and (c) a small translation error and large rotation error.*

In our experiment, the six degrees of freedom of the target movement, translations along the X, Y and Z axes and rotations about the X, Y and Z axes, were respectively driven by:

$$x(t) = \sum_{i=0}^{19} A\, p^{-i} \sin(2\pi f_0 p^i t + \alpha_x(i)) \qquad (1)$$

$$y(t) = \sum_{i=0}^{19} A\, p^{-i} \sin(2\pi f_0 p^i t + \alpha_y(i)) \qquad (2)$$

$$z(t) = \sum_{i=0}^{19} A\, p^{-i} \sin(2\pi f_0 p^i t + \alpha_z(i)) \qquad (3)$$

[

$$\varphi(t) = \sum_{i=0}^{19} B\ p^{-i} \sin(2\pi f_0 p^i t + \alpha_{\varphi}(i)) \qquad (4)$$

$$\theta(t) = \sum_{i=0}^{19} B\ p^{-i} \sin(2\pi f_0 p^i t + \alpha_{\theta}(i)) \qquad (5)$$

$$\psi(t) = \sum_{i=0}^{19} B\ p^{-i} \sin(2\pi f_0 p^i t + \alpha_{\psi}(i)) \qquad (6)$$

where t is the time duration from the beginning of each experimental test and the constants A = 3.5, B = _/3.0, p = 1.25, f_o = 0.01. These values were determined through pilot testing so that the target remained within the bounds of the display and moved at a challenging but manageable speed. The parameters $\alpha_x(i)$, $\alpha_y(i)$, $\alpha_z(i)$, $\alpha_{\varphi}(i)$, $\alpha_{\theta}(i)$, $\alpha_{\psi}(i)$, (i = 0, ..., 20) were independent pseudo-random numbers ranging between 0 and 2π. Since f_o and p were common, each forcing function therefore had identical frequency characteristics.

Performance Measures

At sampling instant $t = i\Delta t$ (where i is the step number and Δt is the sampling period), the vector from the center of the cursor to the centre of the target is defined as the *translation vector* $\mathbf{T}_i$.. The norm of $\mathbf{T}_i$. i.e., the Euclidean distance from the centre of the cursor to the centre of the target, is the translational error. For each trial, the translational root-mean-square (RMS) error was defined as

$$T_{rms} = \sqrt{\frac{\sum_{i=0}^{N} |\mathbf{T}_i|^2}{N}} \qquad (7)$$

where N was the final step in the trial.

The translation vector $\mathbf{T}_i$ consisted of three components along the horizontal (x_i), vertical (y_i) and depth (z_i) dimensions respectively, i.e.:

$$\mathbf{T}_i = (x_i, y_i, z_i) \qquad (8)$$

For each trial the decomposed RMS tracking errors in the X, Y and Z dimensions were defined according to

$$X_{rms} = \sqrt{\frac{\sum_{i=0}^{N} x_i^2}{N}}, \qquad Y_{rms} = \sqrt{\frac{\sum_{i=0}^{N} y_i^2}{N}}, \qquad Z_{rms} = \sqrt{\frac{\sum_{i=0}^{N} z_i^2}{N}} \tag{9}$$

respectively. X_{rms}, Y_{rms} and Z_{rms} are the decomposed measures used for the subsequent dimensional analysis of translation.

Rotational errors were measured in a similar way. (Parameterization of rotations in 3D space is a relatively complex subject; see Altmann, 1986 or Hughes, 1986 for mathematical treatments of rotational parameterization). At $t = i\Delta t$, the angular error (rotation mismatch) between the cursor and the target can be expressed as [25], p 70)

$$R(\phi_i \mathbf{n}_i) \tag{10}$$

where $R(\phi_i \mathbf{n}_i)$ signifies that at tracking step i, the cursor and the target are angularly mismatched about an axis $\mathbf{n}_i$ by an amount ϕ_i. $\mathbf{n}_i = (n_{xi}, n_{yi}, n_{zi})$ is a unit vector specifying the *direction* of the orientation mismatch and ϕ_i is the *amount* of mismatch. ϕ_i and $\mathbf{n}_i$ can be combined as a single rotation vector:

$$\boldsymbol{\phi}_i = \phi_i \mathbf{n}_i = (\phi_i n_{xi}, \phi_i n_{yi}, \phi_i n_{zi}) = (\phi_{xi}, \phi_{yi}, \phi_{zi}) \tag{11}$$

Since $\mathbf{n}_i$ is a unit vector, the length of $\boldsymbol{\phi}_i$ is the amount of rotation mismatch between the cursor and the target. ϕ_{xi}, ϕ_{yi}, and ϕ_{zi} are the decomposed components of the rotation mismatch in the X, Y and Z dimensions respectively. Note that ϕ_{xi}, ϕ_{yi}, and ϕ_{zi} are not pitch, yaw and roll angles. They are the projections of vector $\boldsymbol{\phi}_i$ onto X, Y and Z axes. The values of ϕ_{xi}, ϕ_{yi}, and ϕ_{zi} relative to each other collectively reflect the *inclination* of $\boldsymbol{\phi}_i$ towards the axes X, Y or Z. For instance, the greater ϕ_{xi} is relative to the other components, the more biased the rotation vector $\boldsymbol{\phi}_i$ is towards the horizontal axis X (See Figure 2).

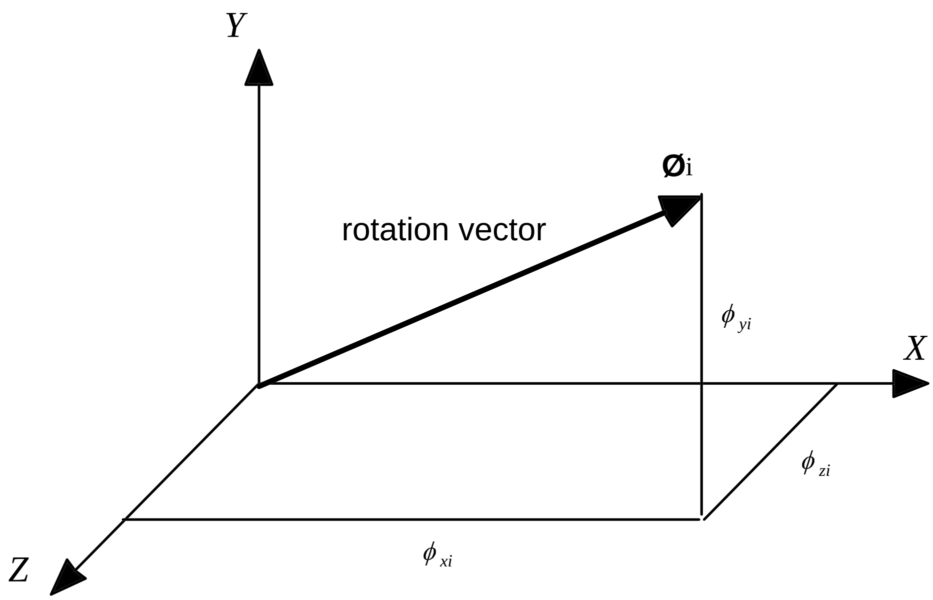

Figure 2. *Rotation vector and its components in the X, Y and Z dimensions*

For each entire trial, the total RMS rotational error is:

$$R_{rms} = \sqrt{\frac{\sum_{i=0}^{N} |\phi_i|^2}{N}} \qquad (12)$$

where N is the final step in the trial.

The RMS values of the individual components ϕ_{xi}, ϕ_{yi}, and ϕ_{zi} are:

$$R_{xrms} = \sqrt{\frac{\sum_{i=0}^{N} \phi_{xi}^2}{N}}, \quad R_{yrms} = \sqrt{\frac{\sum_{i=0}^{N} \phi_{yi}^2}{N}}, \quad R_{zrms} = \sqrt{\frac{\sum_{i=0}^{N} \phi_{zi}^2}{N}} \qquad (13)$$

R_{xrms}, R_{yrms} and R_{zrms} reflect the total rotational errors (from $i = 0$ to $i = N$) of the projections of rotation vector ϕ_i onto the X, Y and Z axes respectively. These measures are used as the decomposed rotation measures for analysis of rotation asymmetries.

Experimental System

Display. In designing the 3D displays used in the experiment, four types of depth cues were chosen, all of which are both powerful and easily realized with current technology: binocular (stereoscopic) disparity, linear perspective, interposition (edge occlusion), and partial occlusion through semi-transparency. Binocular disparity, linear perspective and interposition are conventionally used as strong depth cues [2]. The use of semi-transparency to create partial occlusion, as shown in Figure 1, is a less prevalent technique, but has been shown to be both effective and easy to implement [26, 27].

During the experiment, subjects sat 60 cm away from a Silicon Graphics colour display (Model No. 2086A3SG) and wore CrystalEyes™ 120 Hz stereoscopic glasses (Model No. CE-1), manufactured by StereoGraphics Corp. The experimental room was darkened throughout the experiment.

Input Controllers. Two 6 DOF input controllers were used in the experiment: a Spaceball™ (Model 2003, Spaceball Technologies Inc.) and an EGG (Elastic General-purpose Grip) controller, an egg shaped 6 DOF device designed by the authors (Figure 3). Whereas the Spaceball™ is an isometric, force sensitive device, the EGG is a suspended elastic resistance device whose displacement is proportional to the force and torque applied by the user. Both devices were operated in rate control mode (see [16] for further details). During the experiment each controller was situated proximal to the subject's dominant hand.

(a)

(b)

Figure 3*. The isometric Spaceball™ (a) and elastic EGG (b) input controllers used in the experiment*

One of the fundamental decisions to be made in designing such experiments is to develop a set of transfer functions between the operator's control output and the control device output which permit fair and valid comparisons across different devices. Our prevailing philosophy in this research has been to determine an *optimal* transfer function for each device, which thus supports our claim that all comparisons are made between the *best case design* of each controller configuration. In so doing we therefore maximise the conservativeness of our tests or, alternatively, minimise the extent to which significant effects can be attributed to the design of the controller transfer function rather than experimental treatment. (See [16] for further details.)

To accommodate both fast/coarse motions and slow/fine motions in this experiment, a non-linear exponential transformation was applied to all inputs:

$$\begin{aligned} &\Delta x = K_x i_x^{\alpha}, \quad \Delta y = K_y i_y^{\alpha}, \quad \Delta z = K_z i_z^{\alpha}, \\ &\Delta\varphi = K_{\varphi} i_{\varphi}^{\beta}, \quad \Delta\theta = K_{\theta} i_{\theta}^{\beta}, \quad \Delta\psi = K_{\psi} i_{\psi}^{\beta}, \end{aligned} \tag{14}$$

where K_x, K_y, K_z, K_φ, K_θ, K_ψ are the control gains (sensitivities) for each DOF, chosen empirically by optimal searching; and i_x, i_y, i_z, i_φ, i_θ, i_ψ are the signals generated by each control device (either the Spaceball or the EGG controller) for each of the six degrees of freedom. The coefficients α and β were also determined empirically and set at 2.75 and 2.2 respectively.

Workstation. The experiment was conducted using the MITS (Manipulation in Three Space) system developed by the authors. MITS is a desk-top stereoscopic virtual environment developed for the purpose of investigating 6 DOF motor control performance. In this experiment MITS was run on a SGI Iris 4D/310 GTX graphics workstation, with a graphics update rate of 15 Hz.

Subjects. Thirty paid volunteers were recruited by advertising. All subjects were screened using a Bausch & Lomb Orthorater. Three subjects were rejected for having weak stereoscopic acuity, and one was rejected for having poor corrected near-vision acuity. The accepted subjects' ages ranged from 18 to 37, with 3 subjects under 20, 17 subjects between 20 and 30, and 6 subjects over 30 years of age. Except for one school teacher and one high school student, the rest of the subjects were university students studying Engineering, Science, or Humanities. None of the subjects had prior experience with any 6 DOF manipulation devices. Half of the 26 subjects accepted were assigned to the isometric rate controller (Spaceball) and the other 13 to the elastic rate controller (EGG).

Procedure. Each experimental session was preceded by a 15 minute vision screening test and a handedness check. The data gathering was divided into five phases, as illustrated in Figure 4. Each phase consisted of a practice session, followed by 4 trials of tracking. Each trial lasted 40 seconds. Practice in *Phase 0* proceeded as follows: The subject was first shown how to use the assigned input device to control the cursor for each of the six degrees of freedom, as well as translations and rotations along/about arbitrary axes. After that, the subject performed one trial of tracking. The total duration of Phase 0 practice was about 3 minutes (Figure 4). Practice sessions in Phases 1, 2, 3 and 4 lasted 7 minutes each, and consisted of demonstrations and coaching by the experimenter, together with actual practice trials.

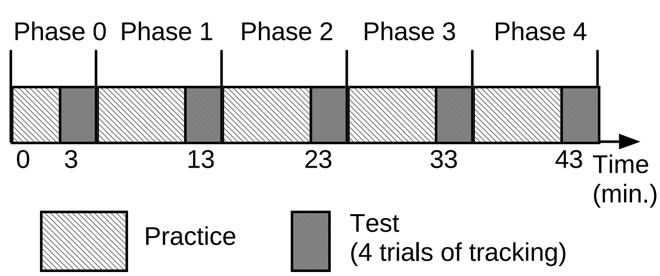

***Figure 4**. Experimental procedure: each phase consisted of practice followed by a test consisting of 4 trials of tracking.*

Each of the four trials in any test had a distinct target trajectory. Each trial began with the cursor coincident with the target (zero error). During the experiment, subjects were instructed to track the target as closely as possible in both translation and rotation.

RESULTS AND DISCUSSIONS

RMS tracking error scores, as defined earlier in equations (7), (8), (12) and (13), were collected for 2 (controller types) x 13 (subjects) x 5 (phases) x 4 (paths) = 520 trials. Nonlinear (logarithmic) transformations were applied to the data in order to meet the model residual distribution requirement for ANOVA analysis [28]. In the following analysis, results are organized respectively according to translation in 3D, rotation in 3D and the controllability of all 6 DOF with one hand.

Anisotropic Performance in Translation

A repeated measure analysis of variance on the translational RMS Errors X_{rms} , Y_{rms} , Z_{rms} as defined in equation (8) was conducted with one between-subject factor (controller type) and three within-subject factors (dimension, experimental phase and tracking path). The significant main effects found included dimension (X, Y, Z) ($F(2, 48) = 59.03$, $p<.0001$), experimental phase ($F(4, 96) = 98.9$, $p<.0001$), and tracking path ($F(3, 72) = 12.73$, $p<.0001$). A significant interaction was also found between dimension and experimental phase ($F(8, 192) = 6.96$, $p<.0001$).

Pairwise contrast comparisons [28] showed that tracking errors in the X, Y and Z dimensions were significantly different from each other. The means of the X, Y and Z RMS errors are shown in Figure 5. As expected, the mean error in the Z direction was significantly greater than both of those in the other two directions (X vs. Z contrast: $F = 114.48$, $p< .0001$; Y vs. Z contrast, $F=48.8$, $p<.0001$). As measured by magnitude, the mean of Z_{rms} was 40% greater than the mean of X_{rms} and 17% greater than the mean of Y_{rms} . It is noteworthy that with the particular 3D depth cues used in the present display, including stereoscopic disparity and semi-transparency, the mean differences found were well below the levels of about 400% previously reported in [8].

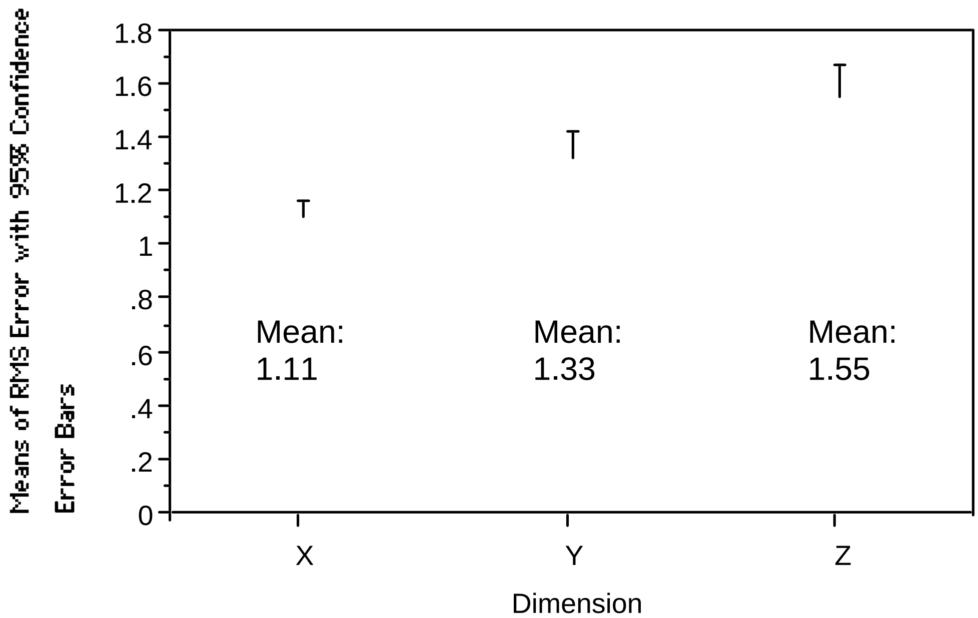

Figure 5*. Means of RMS tracking errors in the horizontal (X), vertical (Y) and depth (Z) dimensions*

Another interesting finding is that the mean error in the Y direction was significantly greater than the mean error in the X direction by 20% (X vs. Y contrast: $F = 13.79$, $p<0.001$). This was somewhat surprising, given that both Y and X are on the plane of the display screen. The first possible explanation for this result was that the difference arose due to differences in resolution between the vertical and the horizontal dimensions in the stereo display. That is, using a standard technique of generating stereo with a 120 Hz CRT display, the 1280x1024 display memory is split into top and bottom halves: one half for the left view and one half for the right view. The vertical resolution in such *stereo* displays is therefore less than half of the horizontal resolution. However, the reduced vertical resolution threshold (equivalent to 0.04 graphics unit in the experiment) was still one order of magnitude less than the mean RMS tracking error, which was at a level of more than 1 graphic unit (Figure 5). This implies that the X-Y resolution difference was not the likely cause of this asymmetry between X and Y error scores.

A second explanation for the performance difference between the vertical and the horizontal dimensions is that perhaps there was a bias in either the input controllers or the muscle groups used, which may have made the horizontal dimension easier to manipulate than the vertical dimension. In this experiment, two types of controllers were used in the experiment, differing both in technical design and in physical features. The elastic device involved hand movements, whereas control of the isometric device required force and torque only. It is therefore very unlikely that there would be identical biases due to sensing technologies or motor anatomy. The fact that the same relative performance pattern in the X, Y, Z directions ($X_{rms} < Y_{rms} < Z_{rms}$) was found across the two types of controllers (Figure 6) therefore leads us to reject this explanation.

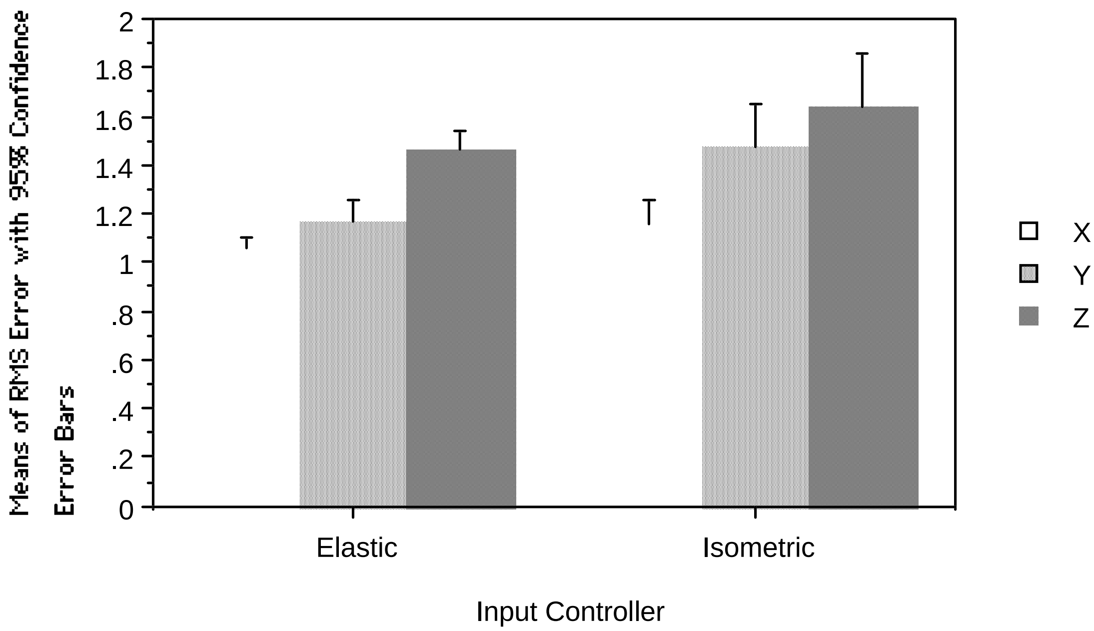

Figure 6*. Translation tracking errors with two types of input controllers.*

Another possible explanation is the particular tracking trajectories presented. Since each tracking path was randomly generated, there exists some probability that movement in the Y dimension might have been more difficult than in the X dimension for a particular trajectory. This possibility was tested and again rejected, however. When the experiment was run with no input control applied to the cursor movement (in the baseline test), the means of X_{rms} were in fact *greater* than the means of Y_{rms} in three of the four tracking paths (Figure 7). However, subjects' relative performance patterns ($X_{rms} < Y_{rms} < Z_{rms}$) were consistent across the four distinct target trajectories (Figure 8), independent of the amount of target movement in each dimension.

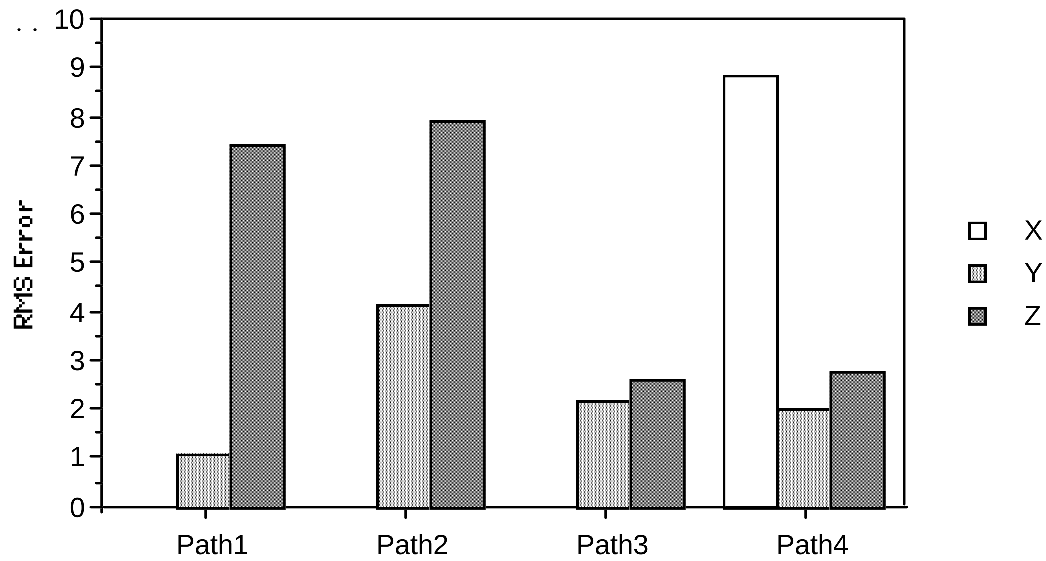

Figure 7. *Baseline test: RMS errors when no input control was applied.*

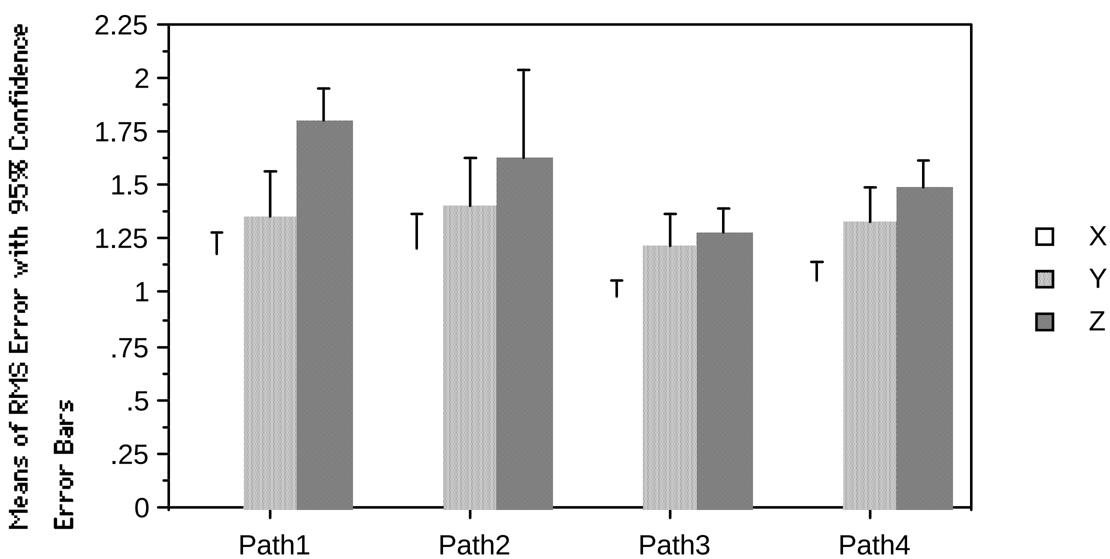

Figure 8. *Consistent performance pattern in X, Y, Z across four tracking paths*

The puzzle of the X and Y differences was better clarified when performance was examined in relation to experimental phase. The differences among the X, Y and Z error components were significantly affected by subjects' practice (dimension x learning interaction: $F(8, 192) = 6.96$, $p<.0001$). As Figure 9 illustrates, errors in the Y dimension initially were as great as those in the Z dimension. As practice progressed, however, Y errors decreased and approached the error level of the X dimension. As indicated respectively in Figure 10 and Figure 11, this pattern was consistent across all four distinct tracking paths and across both types of input controllers. These consistencies were confirmed by the absence of significant interactions between dimension, phase and path ($F(24, 576) = 0.867$, $p = .65$) and between dimension, phase and input ($F(8, 192) = 1.35$, $p = .22$).

This change of performance in Y relative to the other dimensions therefore suggests that the inferior performance in Y is neither due to perception nor due to motor control action per se, but is more likely a manifestation of attentional bias. That is, in the early stages of learning, when subjects had difficulty in managing all of the dimensions simultaneously, they gave higher *attentional priority* to horizontal errors than to vertical errors. In the later stage of learning, however, when their overall performance had improved and attentional resources were presumably freed up, subjects were able to perform equally well in controlling errors in both X and Y dimensions.

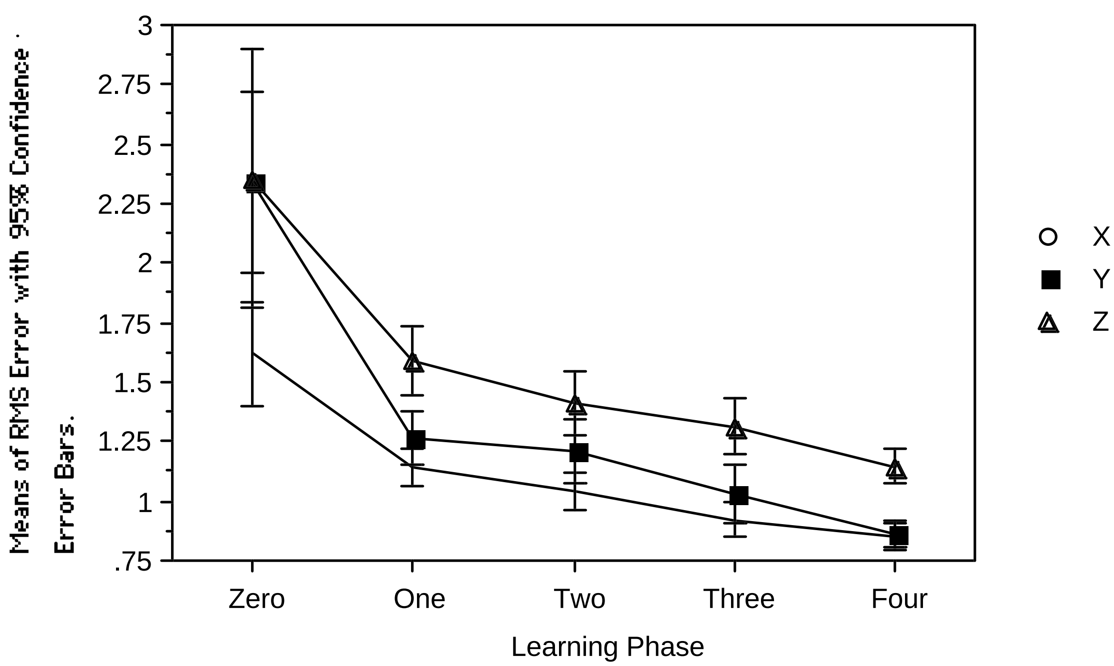

Figure 9. *The evolution of* Y_{rms} *in relation to* X_{rms} *and* Z_{rms} *during learning*

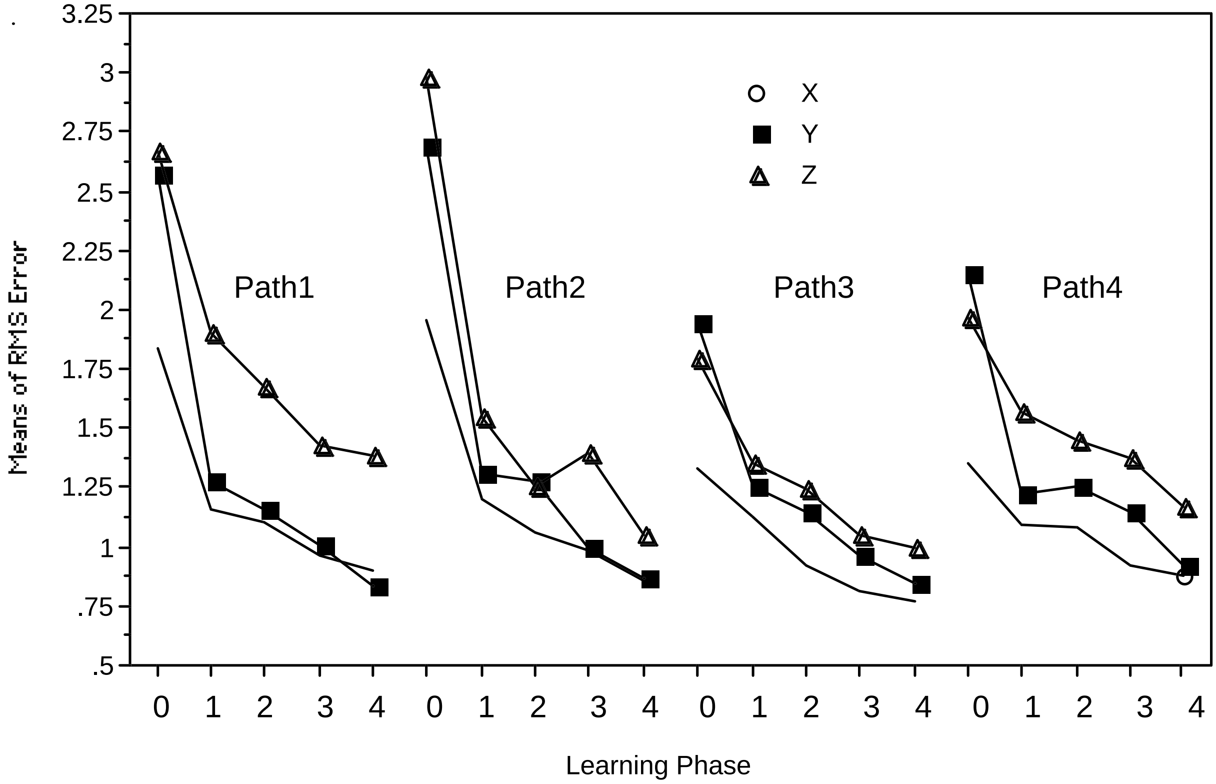

Figure 10. *Error evolution for four distinct tracking paths*

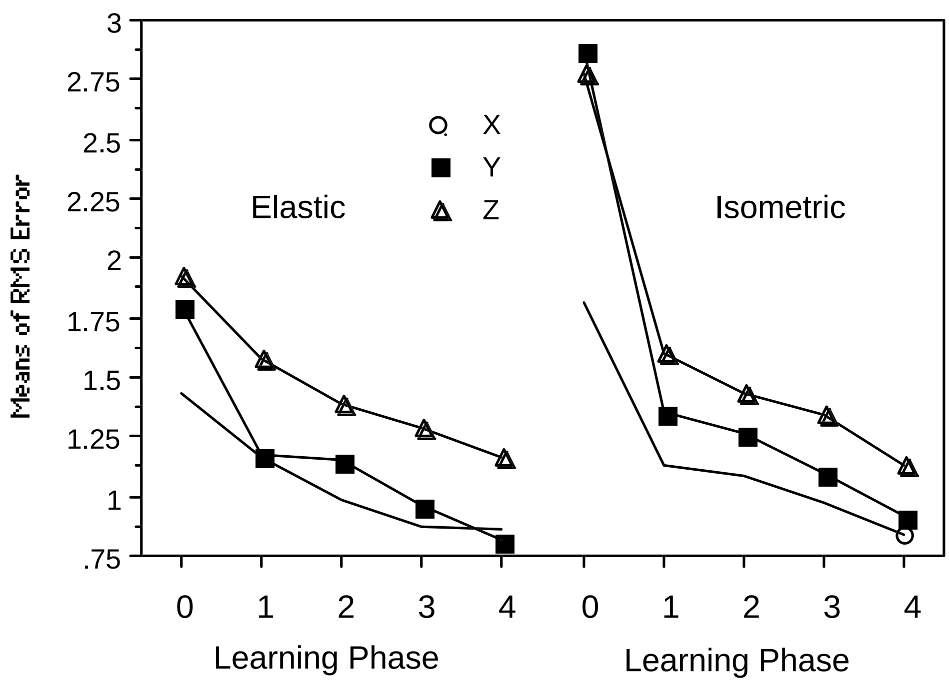

Figure 11. *Error evolution for both input control modes*

The cause of the performance difference between the Z dimension and the X dimension, on the other hand, is certainly perceptual, as is well known in the literature on 3D displays, and presumably also attentional, for the same reasons. In Phase 0, Z_{rms} was 45% greater than X_{rms}. By Phase 4, this difference had reduced to 35%. Since the difficulties of perception in the depth dimension were inherent and thus less likely to diminish so markedly with practice, it is reasonable to assume that the reduction in difference between X and Z dimensions (from 45% to 35%) resulted from more attention paid to the Z dimension in the later learning stages.

This hypothesis of attentional priority in 3D environments is a plausible one. In our daily life, movement visual stimuli are distributed in the horizontal direction much more than in the vertical direction. The literature indicates that, while there is no acuity difference between horizontal and vertical vision, humans tend to be more *sensitive* to horizontal than to vertical length changes, as indicated by shorter reaction times in the horizontal dimension [10].

Performance in Rotation

Based on the decomposed rotational tracking errors R_{xrms}, R_{yrms} and R_{zrms}, as defined in equation (13), a repeated measure variance analysis with one between-subject factor (controller type) and three within-subject factors (R_{xrms}, R_{yrms} and R_{zrms}, experimental phase, and tracking path) showed the same significant main effects as for the translational analysis. That is, dimensional components R_{xrms}, R_{yrms} and R_{zrms} ($F(2, 48) = 5.632$, $p < 0.01$), experimental phase ($F(4, 96) = 48.76$, $p<.0001$), and tracking path ($F(3, 72) = 3.33$, $p=0.25$) were all significant. In contrast to translation, the differences among R_{xrms}, R_{yrms} and R_{zrms} were not affected by learning (dimension x experimental phase: $F(8, 192) = 2.10$, $p=0.66$).

The differences between the R_{xrms}, R_{yrms} and R_{zrms} components are shown in Figure 12. Comparison contrast tests indicated that the rotation vector component along the Z axis (R_{zrms}) was significantly smaller than the components along the X and the Y axes (R_{xrms} vs. R_{yrms}: $F = 27.94$, $p <0.0001$; R_{yrms} vs. R_{zrms}: $F = 14.88$, $P < 0.001$). On the average, R_{yrms} was slightly smaller than R_{xrms}, but this difference was not statistically significant ($F = 2.04$, $p=0.16$).

These results are consistent with those found from the translational analysis. That is, R_{zrms} was the smallest because orientation mismatches about the Z axis do not involve any displacements at all in the Z dimension and therefore were not subject to any of the perceptual restrictions of the depth direction. Orientation mismatches about the X and Y axes, on the other hand, both involve displacements in depth, resulting in greater R_{xrms} and R_{yrms} relative to R_{zrms}. Although the difference was not statistically significant, R_{yrms} was slightly smaller than R_{xrms}

because rotation about the Y axis involves horizontal change while rotation about the X axis involves vertical change. Due to the small size of the target and the cursor, the dimensional differences in rotation were less pronounced.

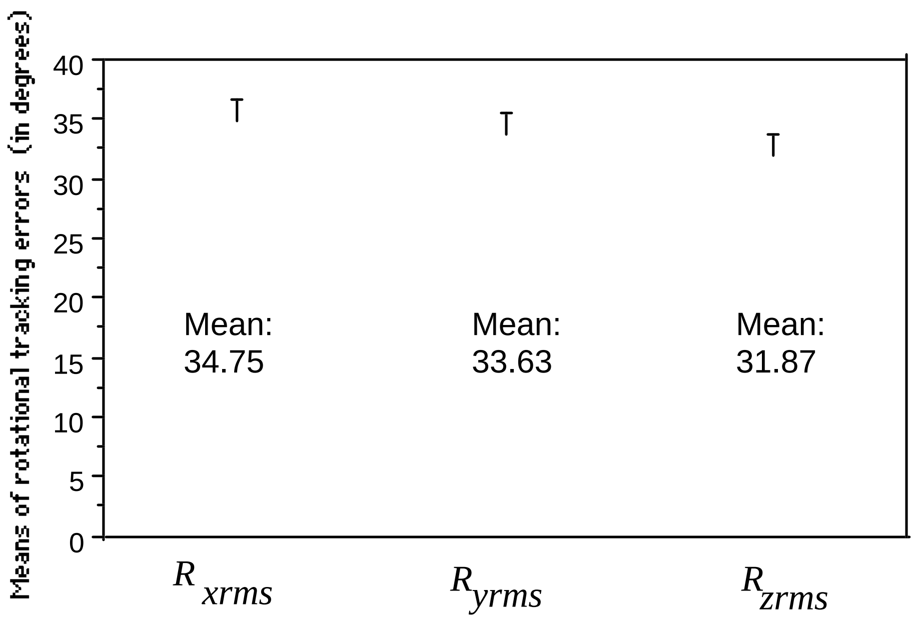

Figure 12*. The means of decomposed rotation errors*

Controlling both translation and rotation with one hand

This section analyses subjects' performance in terms of their ability simultaneously to control all six degrees of freedom with one hand. During the experiments it was observed that when subjects could not do the tracking task very well, especially in the early stages of the experiment, they tended to ignore rotations and concentrated on moving the cursor to catch the target in position only. In spite of the fact that no instructions were given regarding whether to give preference to translation over rotation, or vice versa, this was apparently a reasonable strategy to take, since rotation errors have a limited range (180 degree at the greatest) while translation errors are theoretically unlimited in size. In the later stages of the experiment, however, after significant learning had taken place, the majority of the subjects seemed to be able to control all six degree of freedom concurrently. In order to substantiate this hypothesis quantitatively, 6 DOF tracking errors were decomposed into scores along six separate dimensions. However, since the separate analyses of (a)symmetries within the translational and rotational degrees of freedom have already been presented, we limit ourselves in this section to an analysis of the combined translation errors and combined rotation errors.

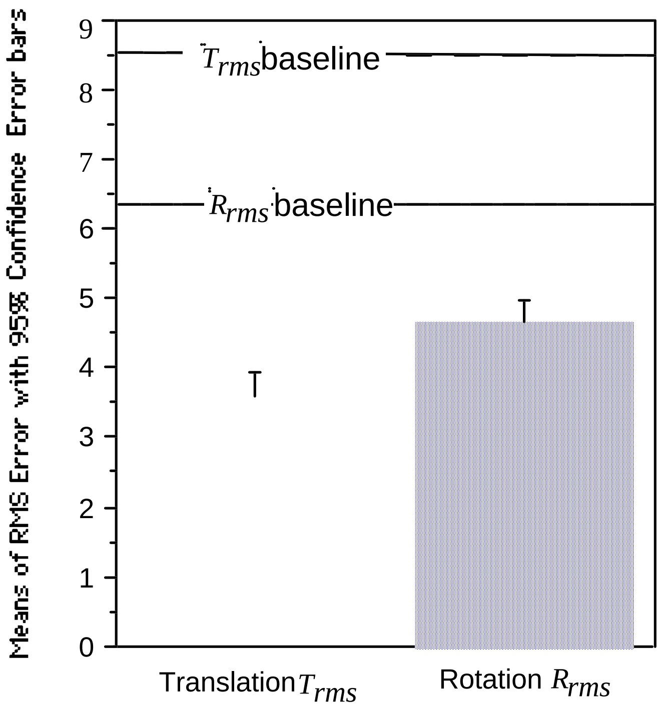

***Figure 13**. Mean tracking errors in translation and rotation during Phase 0 (no practice).*

Figure 13 shows the mean rms values of translation errors and rotation errors during experimental phase 0 (first four trials). In the figure, translation RMS error T_{rms} is defined by equation (7) and rotation RMS error R_{rms} by equation (12). In order to compare the two quantities, which in fact have different units (degrees of rotation vs. graphic units for translation), it was necessary to normalize the two measures. R_{rms} has thus been transformed into equivalent translational units by multiplying each datum by a scale value equal to the radius of the target tetrahedron. In other words, each value rR_{rms}, where r = 3.55, is equivalent to the length of the arc through which the tetrahedron vertices moved during the corresponding rotation. Following this transformation, it is now valid to compare T_{rms} values with rR_{rms} values.

An additional problem, however, is to determine how good the scores were, not only relative to each other, but on an absolute performance scale. *Baseline scores* for both translation and rotation were therefore computed, as standards of comparison. These were obtained by running the experiment several times with *no subject control input* and finding the mean values of T_{rms} and R_{rms} over four trials, each corresponding to one of the four distinct trajectories. As indicated in Figure 13, the two mean baseline scores thus obtained were 8.62 graphic units for T_{rms} and 6.35 graphic units (corresponding to (180/_) *6.35 /r = 119.1 degrees of rotation) for

R_{rms}. Any scores which approach those levels are therefore not very different from chance performance, whereas any scores significantly below those levels represent an improvement over chance.

Large individual differences exist in the ability to simultaneously track both translation and rotation. As shown Figure 14, in Phase 0 two subjects, U and W, did not effectively control either translation or rotation (i.e. RMS errors were close to or even greater than the baseline data). The other twenty four subjects controlled translation with varying degrees of success, i.e. significantly below T_{rms} baseline, but many of them could not manage rotational control at this stage at all. Four of them, subjects B, D, M and P, were no more than 5% below the baseline rotation RMS error. In summary, the data show that without practice, 20 of 26 (77%) subjects were able to cope somewhat with both translation and rotation together; 4 of 26 (15%) were able to control only translation effectively, and 2 of 26 (8%) subjects could control neither translation nor rotation at all.

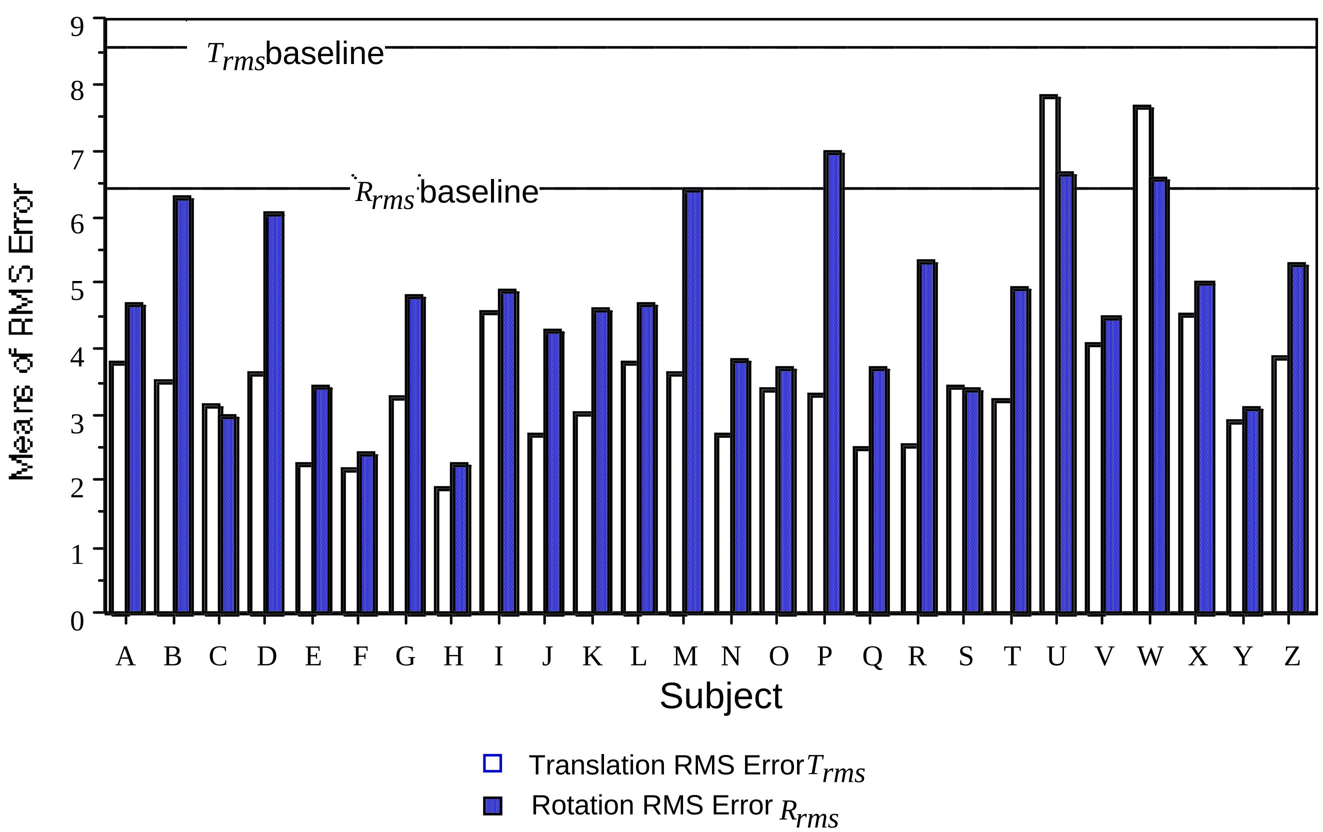

Figure 14 *Individual performance scores in translation and rotation during Phase 0 (no practice).*

Subjects substantially improved their performance with practice. Figure 15 and Figure 16 show respectively the mean and individual performance scores during the final phase of the experiment. After 40 minutes of practice, all subjects were able to control translations to a reasonable degree (more than a 50% reduction from the baseline). Two subjects (B and W), however, still could not effectively control rotation (less than 5% reduction from the baseline), 3 more (R, U, and Z) had more than 5% but less than 50% reduction from the baseline. These five subjects, (B, W, R, U and Z) also had larger performance disparities between translation and rotation. That is, their rotation errors were greater than their translation errors by 185% (B), 83% (R), 83% (U), 197% (W) and 126% (Z) respectively. The rest of the subjects (21 of 26 = 81%) controlled both rotation and translation together, which required all 6 DOF, with some degree of success, as indicated in Figure 16 by the substantial reductions from the respective baselines in both translation and rotation .

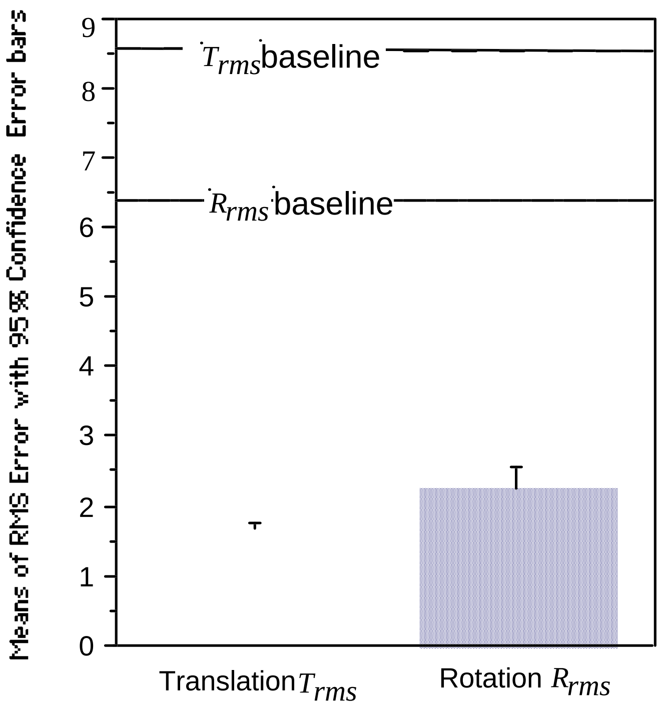

Figure 15*. Mean tracking errors in translation and rotation during Phase 4 (40 minutes practice)*

(initially) to 135% (after 40 minutes of practice) of the error scores for the horizontal dimension. The experimental evidence suggests that subjects tended to give higher attentional priority to the horizontal dimension than to the vertical dimension. The tracking errors in the vertical dimension were larger than those of the horizontal dimension in the early learning stage, but decreased to the same level as the horizontal errors in the later learning stage. On the input side, the analysis indicated that large individual differences exist in the ability of subjects to control six degrees of freedom in a coordinated fashion. After 40 minutes of practice, more than 80% of the subjects could control both translation and rotation effectively (requiring all 6 DOF).

It appears, therefore, that in tracking 6 DOF movement, the subjects tended to adopt a strategy of allocating attention in a certain biased order. During early stages of learning, when sufficient skills to manage all the degrees of freedom may not yet have been acquired, the subjects tended to concentrate on translations while ignoring rotations. Comparing the three translational dimensions, they apparently gave higher attentional priority to reducing horizontal errors than to reducing vertical or depth errors.

It should be noted that through the methodology developed in the paper, only one set of quantitative human performance data from a particular 3D interface experiment has been presented. The results may in actuality not be independent of variables such as experimental paradigm, task difficulty (bandwidth), and methods and criteria for evaluating task performance. More work remains to be done, both in further developing analysis methodologies and in expanding the scope of evaluations to include data from such interfaces as head-mounted-displays.

References

[1] Haber, R.N. and M. Hershenson, *The psychology of visual perception*. 1973, New York: Holt, Rinehart & Winston.

[2] Kaufman, L., *Sight and mind - an introduction to visual perception.* 1974, London: Oxford University Press.

[3] Wickens, C.D., S. Todd, and K. Seidler, Three-dimensional displays: Perception, implementation and applications. (CSERIAC SOAR-89-01). 1989, Wright Patterson Air Force Base, Ohio.: Armstrong Aerospace Medical Research Laboratory.

[4] McAllister, D.F. (ed.), *Stereo computer graphics and other true 3D technologies*. 1993, Princeton University Press: Princeton, New Jersey.

[5] Sollenberger, R.L. and P. Milgram. Effects of stereoscopic and rotational displays in a three-dimensional path-tracing task. Human Factors. 35(3) (1993). 483-499.

[6] Arthur, K., K. Booth, and C. Ware. Evaluating 3D task performance for fish tank virtual worlds. ACM Transactions on Information Systems. 11(3) (1993). 239-265.

[7] Tachi, S. and K. Yasuda. Evaluation experiments of a tele-existence manipulation system. Presence: Teleoperators and Virtual Environments. 3(1) (1994). 35-44.

[8] Massimino, M.J., T.B. Sheridan, and J.B. Roseborough. One hand tracking in six degrees of freedom. In Proceedings of IEEE International Conference on Systems, Man, and Cybernetics. 1989. pp.498-503

[9] Berman, P.W., J.G. Cunningham, and J. Harkulich. Construction of the horizontal, vertical and oblique by young children: Failure to find the "oblique effect". Child Development. 45 (1974). 474-478.

[10] Gottsdanker, R. and J.D. Tietz. Reaction time for horizontal versus vertical line-length discrimination. Bulletin of the Psychonomic Society. 30(1) (1992). 74-76.

[11] Poulton, E.C. Unwanted asymmetrical transfer effects with balanced experimental designs. Psychological Bulletin. 66(1) (1966). 1-8.

[12] Poulton, E.C. Unwanted range effects from using within-subject experimental designs. Psychological Bulletin. 80(2) (1973). 113-121.

[13] Brooks, T.L. and A.K. Bejczy, Hand controllers for teleoperation, a state of the art technology survey and evaluation. 1985, Jet Propulsion Laboratory:

[14] Siva, K.V., et al. Development of a general purpose hand controller for advanced teleoperation. In Proceedings of International Symposium on Teleoperation and Control. (Bristol, England, 1988). IFS Ltd, UK; Springer-Verlag, Berlin, pp.277-290

[15] Jacobus, H.N., Riggs, A. J., Jacobus, C. J., Weinstein, Y., Implementation issues for telerobotic hand controllers: human-robot ergonomics, in *Human-Robot Interaction*, M. Rahimi and W. Karwowski, (Ed's) . 1992, Taylor & Francis: London.

[16] Zhai, S., Human Performance in Six Degree of Freedom Input Control. PhD Dissertation, 1995, University of Toronto. (http://vered.rose.toronto.edu/people/shumin.html/)

[17] Rice, J.R., J.P. Yorchak, and C.S. Hartley. Capture of satellites having rotational motion. In Proceedings of Human Factors Society 30th Annual Meeting. (Dayton, Ohio, 1986) pp.870-874

[18] O'Hara, J. Telerobotic control of a dexterous manipulator using master and six-DOF hand controllers for space assembly and servicing tasks. In Proceedings of Human Factors Society 31st Annual Meeting. (New York City, 1987) pp.791-795

[19] McKinnon, M. and M. King. Manual control of telemanipulators. In Proceedings of International Symposium on Teleoperation and Control. (Bristol, England, 1988) IFS Ltd, UK; Springer-Verlag, Berlin, pp.263-276

[20] Gibbs, C.B. The continuous regulation of skilled response by kinaesthetic feedback. British Journal of Psychology. 45 (1954). 24-39.

[21] Pew, R.W. Acquisition of hierarchical control over the temporal organization of a skill. Journal of Experimental Psychology. 71(5) (1966). 764-771.

[22] Poulton, E.C., *Tracking skill and manual control.* 1974, New York: Academic Press.

[23] Kim, W.S., Ellis, S. R., Tyler, M. E., Hannaford, B., Stark L, W., Quantitative evaluation of perspective and stereoscopic displays in three-axis manual tracking tasks. IEEE Transactions on System, Man, and Cybernetics. SMC-17(1) (1987).

[24] Ellis, S.R., Tyler, M., Kim, W.S., Stark, L., Three-dimensional tracking with misalignment between display and control axes. SAE Transactions: Journal of Aerospace. (100-1) (1992). 985-989.

[25] Altmann, S.L., *Rotations, Quaternions, and Double Groups*. 1986, Oxford: Clarendon Press.

[26] Zhai, S., W. Buxton, and P. Milgram. The "silk cursor": investigating transparency for 3D target acquisition. In Proceedings of CHI'94: ACM conference on Human Factors in Computing Systems. (Boston, 1994) ACM, pp.459-464

[27] Zhai, S., B. Buxton, and P. Milgram. The partial occlusion effect: utilizing semi-transparency in 3D human computer interaction. To appear in ACM Transactions on Computer Human Interaction. (In Press).

[28] Howell, D.C., *Statistical Methods for Psychology*. Third ed. 1992, Boston: PWS-KENT Publishing Company.

The work was supported in part by the Information Technology Research Centre of Ontario, DCIEM, and the Natural Sciences and Engineering Research Council of Canada.

Shumin Zhai is with the IBM Almaden Research Center, 650 Harry Road, K54/B2, San Jose, CA95120

Paul Milgram is with the Department of Industrial Engineering, University of Toronto, Toronto, ON, Canada, M5S 1A4

Anu Rastogi is with Alias|wavefront Inc, 110 Richmond St. E. Toronto, ON, Canada. M5C 1P1

Biographies:

Shumin Zhai received his Ph.D. degree in Human Factors (Industrial) Engineering from the University of Toronto, Toronto, Ontario, Canada in 1995. From 1984 to 1989, he was on the faculty of XiDian University (formerly the Northwest Telecommunication Engineering Institute), Xian, China. From July 1995 to May 1996, he was a research scientist at the Computer Systems Research Institute, University of Toronto. He is currently a Research Staff Member at the IBM Almaden Research Center, San Jose, California. Dr. Zhai has authored numerous publications on human factors issues in teleoperation, virtual and augmented environments, 3D graphical displays and six degree of freedom input control. He is on the editorial board of the Springer-Verlag journal *Personal Technologies*. His current research interest lies in developing advanced computer interfaces based on human factors insights and experimentation.

Paul Milgram received his BASc degree in Engineering Science at the University of Toronto (1970), his MSEE degree in Electrical (Biomedical) Engineering at the Technion in Israel (1973), and his Ph.D. in Human Factors (Industrial) Engineering at the University of Toronto (1980). From 1980-82 he was a ZWO Visiting Scientist and a NATO Postdoctoral Fellow at the TNO Institute for Perception in Soesterberg, Netherlands, and from 1982-84 a

Senior Research Engineer in Human Engineering at the National Aerospace Laboratory (NLR) in Amsterdam. In 1986 he joined the Industrial Engineering Dept. of the University ofToronto, where he is currently an Associate Professor. He is also cross appointed to the Institute of Biomedical Engineering and the Department of Psychology. During 1993-94 he was an Invited Researcher at the ATR Communication Systems Research Laboratories, in Kyoto, Japan. His research interests include display and control issues in teleoperation and telerobotics, augmented reality, stereoscopic video and computer graphics, cognitive engineering, and human factors issues in medicine. He is also President of Translucent Technologies Inc., a Toronto based company which manufactures and markets PLATO liquid crystal visual occlusion spectacles (of which heis the inventor), for visual and psychomotor research. He is a Member of the Human Factors and Ergonomics Society (HFES), the IEEE, the Human Factors Association of Canada (HFAC/ACE), the Association for Computing Machinery (ACM) and the Professional Engineers of Ontario (PEO).

Anu Rastogi received the B.Tech. degree in mechanical engineering from I.I.T. Kanpur, India in 1990 and the M.A.Sc. degree in industrial engineering from University of Toronto in 1995. His research interest include user interface design, geometric modelling and computer graphics. He has published papers in conference proceedings. He is currently working at Alias|Wavefront, Toronto, Ontario, Canada.