Viewpoint Optimisation for Virtual Environment Navigation Using Dynamic Tethering - A Study of Tether Rigidity

Wenbi Wang and Paul Milgram
Ergonomics in Teleoperation and Control Laboratory (ETC-lab)
Dept. of Mechanical and Industrial Engineering, University of Toronto
{wenbi, milgram} @ etclab.mie.utoronto.ca

Much research has been carried out on providing navigational aids and improving users' wayfinding ability across a variety of navigation related tasks. In this study, we focus on investigating users' navigational performance with respect to the display frame of reference in a large scale virtual environment. Dynamic viewpoint tethering is proposed as a means to improve user control and reduce the need for mental rotations, thus facilitating the acquisition of configurational knowledge about the virtual space.

The modelling of dynamic viewpoint tethering is explained and recent research findings are presented. Twelve volunteers participated in an experiment in which they were instructed to control an aircraft-shaped cursor flying through a set of virtual tunnels (local guidance) and to answer questions about the environment (global awareness). Experimental results showed that neither the very loose dynamic tether nor the completely rigid tether supported the best control performance. Rather, an optimal tether configuration lies at the centre of the rigidity continuum. Research results are discussed from the point of view of design of navigational system interfaces.

Keywords: Dynamically tethered displays, viewpoint optimisation, local guidance, global awareness, virtual cameras, navigation, virtual environments

INTRODUCTION

Marcel was driving through Metropolis to visit his friend. This being his first time in the city, he had to follow a map while driving. Marcel knew where his friend's house was on the map, but he had to keep on updating his own location by scanning the street signs and comparing them to the street names on the map. He soon found the task to be quite stressful, especially after the traffic started getting very heavy, forcing him to concentrate on the driving, with little time left for checking the road names. It didn't take long for Marcel to feel lost. Finally, he decided to pull his car over and call his friend for help.

The story above describes a typical navigational problem and Marcel's experience is not at all uncommon. Similar difficulties have been reported in many different environments, both virtual and real (Darken et al., 2001). Internet users commonly have trouble returning to the web pages they have just visited; computer gamers frequently get confused when travelling within their game world; surgeons often feel disoriented when performing endoscopic surgery, etc. The key to solving this class of problem is to understand the cognitive processes involved in navigation, identify the associated navigational requirements and develop appropriate navigational aids.

As defined by Darken et al. (1998), navigation can be regarded as a process of extracting information, forming mental representations, and using those representations for route planning and moving about. Generally there are two goals when people travel through an environment: 1) to perform the actions necessary to get to their destination; and 2) to understand the spatial structure of the area being traversed. The first goal comprises the physical challenges of navigation, especially when the actions are performed indirectly (e.g., using a steering wheel to drive a car, or using a joystick to manipulate a virtual character in a video game). The second goal encompasses the cognitive challenges of navigation with respect to acquiring the configurational knowledge of the space. These two goals furthermore respectively reflect two of the fundamental sub-tasks involved in navigation: local guidance (tasks involving control) and global awareness (tasks involving understanding, planning, problem solving, etc.). In this study, local guidance is operationally defined as the task of manoeuvring along a route, while global awareness is operationally defined as the maintaining of one's position relative to other objects and the world. The same terms can be used with respect to the task of controlling a separate object or avatar relative to a nominal viewpoint / virtual camera attached to that object / avatar.

Lasswell and Wickens (1995) have presented a thorough discussion of how navigational performance is affected by display frames of reference. They point out that global spatial awareness is supported by knowledge with respect to *exocentric* frames of reference, while local navigational guidance is facilitated by knowledge with respect to *egocentric* frames of references. Examples of this distinction can be found in many popular video games. For example, two views are provided in most first-person combat games, where a front forward (egocentric) view is supplied for facilitating local control tasks (obstacle avoidance, aiming, and shooting), and a north-up map (exocentric) view, with a 'you-are-here' marker, is included for maintaining orientation in the maze-like game space.

Such combined dual display array solutions have proved to be sufficient for non-critical applications, such as game-playing. However, the extra effort needed to integrate the information from these two views could become critical in real-life emergency situations, where information from both frames of references is required. In the story at the beginning of the paper, Marcel needed to perform both control and wayfinding tasks at the same time. The fact that the information needed to perform these tasks, from both frames of reference, was not available in a well-balanced, integrated form led to the failure.

Although there is not much we can do in the physical world to change the many constraints which already exist (i.e. we don't typically have the ability to displace our personal viewpoint outside our bodies), we do have a much broader design space in virtual environments, in which real-world physical constraints do not need to exist. Our aim in this research has been to provide an innovative way of seamlessly presenting information from both egocentric and exocentric frames of reference, thus supporting superior navigational performance in both control and wayfinding tasks.

To achieve this goal, a *tethered* viewpoint has been proposed. Rather than providing a purely local guidance display from the observer's point of view, which would correspond to placing an imaginary (virtual) camera within the avatar[1] and looking forward, we create a viewpoint which is positioned behind and above the avatar, as if there existed a tether which attaches the virtual camera to the avatar. (A virtual camera in computer graphics parlance defines the observer's viewpoint relative to the controlled object.) Figure 1 illustrates the spatial relationship between the virtual camera and the avatar.

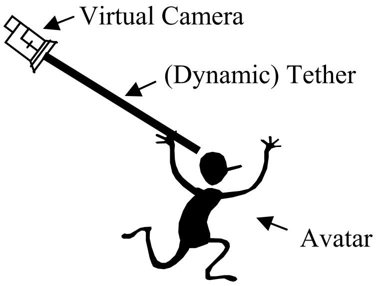

Figure 1. Spatial relationship among (dynamic) tether, virtual camera, and avatar.

In this study, effort has been focused on extending the concept of a simple tethered viewpoint to the modelling of a dynamic tether and evaluating its effects on human navigation and control performance.

Conceptually, the dynamic tether is modelled as analogous to a mass-spring-damper system, with the virtual camera viewpoint coupled to the avatar through this dynamic tether connection. Spring force is proportional to linear displacement, while spring torque is proportional to the angle of rotation directed about an effective axis of rotation. Any movement of the avatar thus influences the tether's displacement and rotation, which generates a virtual force/torque on the virtual camera (and an opposite force/torque on the avatar).

Under most operational circumstances, the position and orientation of an avatar are collectively under the control of either an involved user, or of some external algorithm. When working with an input device, such as a joystick, to control an avatar navigating in the virtual space, the dynamic tether defines the dynamic relationship between the avatar motion and the viewpoint motion, with the operator's direct control of the avatar serving as an excitational force on the tether, through which the viewpoint motion results.

The combined force and torque equations used here are:

$$\boldsymbol{F} = \boldsymbol{K}.\boldsymbol{d} - \boldsymbol{C}.\boldsymbol{v};$$

where

$\boldsymbol{F}$ is a vector comprising all forces and torques;
$\boldsymbol{K}$ is a matrix of spring constants (coefficients of stiffness);
$\boldsymbol{C}$ is a matrix of damping coefficients
$\boldsymbol{d}$ is a vector of relative displacements, either linear or angular, of the avatar
$\boldsymbol{v}$ is a vector of relative velocities, either linear or angular, of the avatar.

Besides mass, spring constant and damping coefficient, the behaviour of a dynamic tether is also characterised by two other important attributes: its natural frequency ω_n and its damping ratio ζ.

The *natural frequency* of any of the vector components of a dynamic tether, calculated from the formula $\omega_n = \sqrt{k/m}$ for each force or torque component (note that we are using the scalar form here, for the sake of simplicity), enables one to distinguish the frequencies of the displayed information from the input command (forcing function) frequencies, based on their spectral characteristics.

According to vibration theory (Dimarogonas, 1996), when the excitational forcing function frequency ω is much less than the natural frequency of the tether, ω_n (i.e. $\omega/\omega_n << 1$), the system behaves as a static system, the *stiffness* becomes the dominant factor, and the mass of the camera has negligible effect. In this case, the dynamic tether behaves like a *rigid pole* and preserves the local avatar frame of reference. In the case of a *very high* frequency excitational force, when the frequency ratio ω/ω_n approaches infinity, the mass of the camera becomes the dominant factor and the spring stiffness has negligible effect, somewhat akin to a "wet noodle". In other words, the virtual camera can not catch up with any fast avatar movements and the tether works like a *low-pass filter*. The display, in this case, takes on the characteristics of a *world frame of reference*. When the frequency ratio equals *unity* ($\omega/\omega_n = 1$), tether *resonance* will occur and damping must be added to reduce the oscillation amplitude.

The frequency separation effect discussed above can also be explained by the relative movement of the pictorial representation of the avatar relative to the surrounding space on the display. When the command frequency is greater than the tether natural frequency ($\omega/\omega_n > 1$), control forces applied to the avatar will result in movements of the displayed avatar in the same direction as the control forces, relative to the surrounding space. On the other hand, when

---

[1] We use the term 'avatar' here to refer to any kind of remote entity under control of the user.

the control frequency is less than the tether natural frequency ($\omega/\omega_n < 1$), control forces applied to the avatar will result in movements of the surrounding environment *in the opposite direction* to the applied control force. In a general sense, when extended to their limits, these two cases correspond respectively to the concepts of *exocentric* and *egocentric* viewpoints, discussed earlier (Colquhoun, 2001).

In the former case, motion of the avatar in the same direction as the command input is compatible with the kind of feedback commonly expected by operators, and is consistent with the principle of compatibility of display movement (Colquhoun, 2001). This states that the moving component on a display should conform to the mental model of the operator by responding in the same direction as the operator's command input. On the other hand, in the latter case, the movement of the whole environment in the opposite direction maintains the proper static display/control (i.e. "out-the-window") alignment and reduces the need for extensive mental rotation when making static directional judgements. This case corresponds in many ways to the well known case of compensatory control, and is important for both mitigating control difficulties and acquiring configurational knowledge of the space (Colquhoun, 2001).

As pointed out above, another important attribute of any dynamic tether is its *damping ratio,* calculated from $\zeta = {}^{c}\!/\!{}_{\sqrt{2mk}}$ (in scalar form). The damping ratio (ζ) determines the amount of oscillation that may occur during viewpoint movement. Viewpoint oscillation refers to movement (up/down, left/right, or forward/backward) of the viewpoint around its equilibrium position. Assuming that the spring constant (**k**) stays fixed, with the change of damping coefficient value, the tether could be in an under-damped, a critically-damped, or an over-damped condition. The reader is referred to Wang & Milgram (2002) for a detailed description of these three conditions.

Because our goal is to provide useful guidelines for directing navigational display design, it is our concern how people's navigational performance is affected by these different tether configurations. In the early stages of this project, experimental results confirmed the hypothesised advantage of tethered viewpoints (either dynamic or rigid) over egocentric (first-person) viewpoints and of exocentric (map) displays in supporting local navigational guidance performance (Wang & Milgram, 2001). A follow-up study also showed an improvement with respect to local guidance performance of critically-damped tethered displays in general over both under-damped and over-damped tethered displays (Wang & Milgram, 2002).

The goal of the experiment reported here is to further explore the design space of critically damped tether configurations and find the relationship between natural frequency and command input bandwidth in supporting navigation. It is our hypothesis that the extreme cases of both very loose and very rigid tethers will affect users' control performance detrimentally and that the best control performance will be supported by tethers that are located at the *centre* of the rigidity continuum.

METHOD

Participants

Twelve students (7 men, 5 women) participated in the experiment. All had normal or corrected-to-normal vision, and satisfied a standard test of stereoscopic acuity. Participants were paid $70 for their participation.

Apparatus and software

As described in the Procedure section below, participants were required to perform a set of navigational tasks in a virtual environment rendered on a SGI O2 workstation. The virtual environment was developed using OpenGL and consisted of a three dimensional virtual winding tunnel. It was regarded as a large-scale virtual space since global spatial information was needed to navigate through it.

Six different tunnel configurations were used in the test, with each tunnel following a different sinusoidally shaped but equally long trajectory from a start point to an end point in the virtual space. Figure 2 shows a (monoscopic) representation of one of the tunnels used in this study, as seen by the subjects. (In order to provide maximal depth cues to the participants, IMAX goggles were used to implement stereoscopic viewing.) As indicated, the centre line of each tunnel was marked by a red line

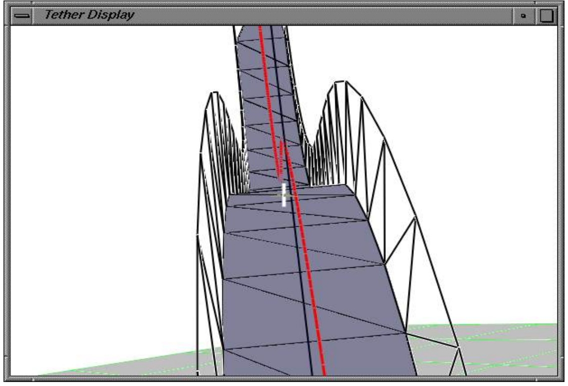

Figure 2. Snapshot from dynamically tethered display used in the study. The white airplane symbol at the centre of the display indicates the avatar controlled by the subjects.

The frequency of each tunnel was determined by both its shape and the flying speed of the avatar through it. In this study, a constant avatar flying speed was used for all testing conditions. As a consequence, the frequency of each tunnel was affected solely by its shape. The six tunnels used were evenly divided into two frequency groups, with tunnels in each group having the same sinusoidal frequency. The high frequency group had a spatial frequency of 1 cycle/cm which, due to the constant speed of traversal, was equivalent to a value of 1 Hz in the time domain. The low frequency group had a spatial frequency of 0.1 cycles/cm, or 0.1 Hz. The only differences among the three tunnels within each group were with respect to their spatial orientation.

To test our hypotheses, 6 tether conditions were compared in this study, as summarised in Table 1. Five of the tethers (T1-T5) were critically damped, with various spring constants, and thus different natural frequencies. A rigidly tethered display (T6, infinite rigidity) was also included as a control condition. The same tether length was used in all displays. The selection of these parameter values was based on two considerations: 1) covering the whole spectrum of rigidity; and 2) ensuring that the tasks (both local guidance and global awareness) were viable.

Tether	Spring constant	Damping coefficient
T1	1	2
T2	100	20
T3	400	40
T4	900	60
T5	1600	80
T6	Infinite	n/a

Table 1. Parametric definition of the six tethered displays.

Procedure

Participants individually completed six one-hour sessions spread over a three-day period. During the first session, which was used for training, participants received both written and oral instructions describing the simulation and the tasks. The experimenter remained in the room and answered general questions when necessary. In the ensuing four sessions, the participants completed four blocks of active flying-through-the-tunnel trials.

The active flying task required the participant to control an avatar, represented by an aircraft symbol, flying along the centre of the virtual tunnel with two wings parallel to the tunnel floor, while simultaneously mentally keeping track of the shape of the tunnel. A Spaceball was used to control the three rotational degrees of freedom of the aircraft (i.e. pitch, yaw and roll). Forward motion of the aircraft was independently controlled by the software; that is, the aircraft automatically flew forward at a constant speed.

The order of display presentation was randomly decided. Within each block there were 15 experimental trials, each of which lasted up to one minute.

For each of these active flying conditions, RMS tracking error was used to measure local guidance performance. The error scores were measured in terms of graphical distance units. A perfect score (zero error) would have meant that the participant had flown the aircraft exactly along the centre of the tunnel, with the wings continuously parallel to the tunnel floor.

After completion of each flying trial, participants were also presented with a set of six physical tunnel models. One among those six physical tunnel models had the same shape and orientation as the virtual tunnel used in the just completed trial and participants were required to pick it out. The accuracy of this judgement, in terms of percent correct judgements, was used as an indication of the completeness of the cognitive map developed by the participants (Golledge, 1999). In other words, these scores were used as a measure of global awareness performance.

The sixth session of the experiment was a passive navigation session, in which we attempted to assess how the different display formats affected participants' judgement of their own performance on this task. The purpose of this was to elicit a subjective evaluation of the efficacy of the different tether conditions from this new population of "experts" in the task. For this final condition, the airplane automatically flew perfectly through the tunnel along the centre red line, but with differently tethered viewpoints. Without telling the participants that the airplane had been flying perfectly, we asked them to make subjective evaluations of (local) tracking performance for the different display conditions, on a scale of 0 to 100. In addition to this, the tunnel shape recognition task described above) was again carried out following each trial. Finally, a brief interview was carried out at the end of this session.

RESULTS

Local guidance performance

Local guidance performance was analysed relative to the two levels of tunnel frequencies. An overall RMS error was calculated, to represent a comprehensive measurement of guidance performance with respect to both location and orientation.

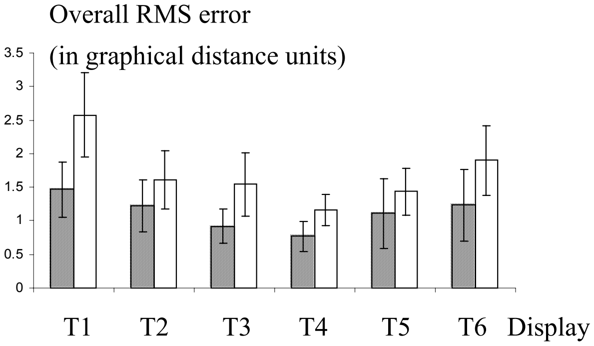

Figure 3. Local guidance performance across six tethered displays under high (white bars) and low (hatched bars) input command frequency conditions. (Horizontal axis is in increasing order of tether rigidity.)

For the high frequency tunnels, a U-shaped performance curve was clearly observed for RMS error across the six display conditions (F(5, 360) = 5.62, $p < 0.001$), as shown in Figure 3. (Error bars in this and subsequent graphs represent 95% confidence intervals.) This finding suggests that the best performance (smallest RMS error) was supported by tethered displays in the centre of the rigidity continuum, in this case in the vicinity of T4. Tukey's test showed that T1 (the loose tether) was significantly worse than the rest of the conditions, except the rigid tether (T6).

For the low frequency tunnels, superior local guidance performance was observed under all six display conditions as compared to high frequency task conditions, as shown in Figure 3. The mean value of overall RMS errors also appeared to follow a U–shaped curve here, with the apparently lowest mean error again corresponding to T4. However, statistical analysis did not indicate that any of performance scores across the six displays were significantly different (F(5,413)=2.02, p=0.075). This, we believe, is due to the fact that the relative simplicity of this particular task condition masked the potential advantage of tethered displays in the centre of the centricity continuum.

Global awareness performance

Figure 4 shows the users' global awareness error performance scores across the six tether conditions. Note that, in general, subjects correctly selected the proper tunnel more than 50% of the time, and in some cases (e.g. high frequency T5) almost 90% of the time. The correct tunnel shape recognition scores (global awareness measure), however, didn't reveal any significant differences across display conditions for either the low frequency tasks (F(5,413)=3.02, p=0.011) or for the high frequency tasks (F(5,413)=0.17, p=0.974). In addition to this result, no significant differences were evident either for any of the subjective passive-flying evaluations. This absence of detectable significance suggests that global awareness was not significantly affected by the particular dynamic tether properties used in this experiment.

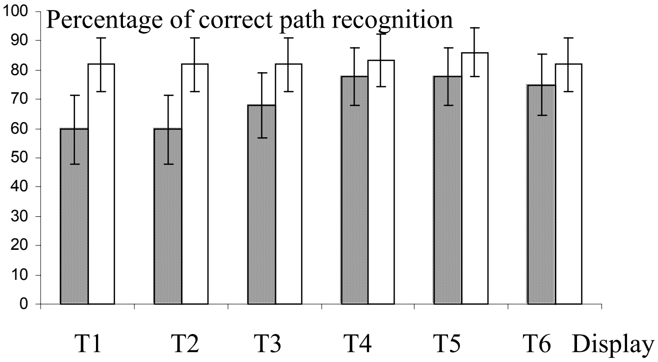

Figure 4. Global awareness performance across six tethered displays under low (hatched bars) and high (white bars) input command frequency conditions (horizontal axis in increasing order of tether rigidity).

DISCUSSION

Task bandwidth and local guidance performance

Through theoretical analysis and empirical observation, two factors have been identified as affecting users' local guidance performance: control/display alignment and control/display congruence. Display formats which violate these factors will normally cause deterioration in control performance. Most of the experimental findings in this study can be explained in this light. For example, the oscillations caused by the loosely tethered viewpoints (i.e. dynamically tethered displays with small rigidity values) failed to maintain good control/display alignment, thus, leading to the poor RMS score.

The experimental results can also be explained from another perspective: by the *amplification factors* associated with each dynamic tether filter. The larger the amplification factor, the greater the viewpoint oscillation. However, the amplification factor, or *control gain* G, is a function of the ratio between the task frequency and the undamped natural frequency of the tether, that is, $G = f(\omega/\omega_n)$. When the input forcing frequencies were relatively low, the differences in amplification factors across tether conditions should have been small. In fact, in our experiment the amplification factors of all six displays for the low frequency conditions were all close to unity, and it is thus not surprising that no significant control performance differences were found for that case. On the other hand, for the high frequency task conditions, the theoretical amplification factors of all the tethers also increased, but to different values. This likely explains why local control performance scores were lower overall and variances for the high frequencies were higher.

Tunnel shape recognition as a secondary task

Tunnel shape recognition scores were used in this study as a measure of participants' global awareness performance. This same measurement was used in our earlier study, where the results showed improved performance with increased display exocentricity (Wang & Miligram, 2001). It is our belief that the display centricity characteristic in a tethered display is determined by the tether *length* and *nominal viewpoint* position, neither of which were investigated here. These two factors together determine the amount of context information the user can perceive, as well as the amount of preview information the user can obtain. Since the same tether length and nominal viewpoint position were used for all trials in the experiment reported here, all the displays compared in this study had essentially the same centricity characteristics. It is not surprising therefore that no significant performance differences were observed in global awareness performance. Our follow-up research to the study reported here (Wang & Milgram, 2003) is thus investigating how these other characteristics affect global navigation performance.

REFERENCES

Colquhoun, H. (2001). Dynamic tethering for enhanced remote control and navigation. Unpublished Master of Applied Science dissertation, University of Toronto.

Darken, R.P., Peterson, B. (2001). Spatial Orientation, wayfinding, and representation. *Handbook of Virtual Environment Technology.* Stanney, K. (Ed.) Erlbaum.

Darken, R.P., Allard, T., Achille, L.B. (1998). Spatial orientation and wayfinding in large-scale virtual spaces: An introduction. *Presence: Teleoperators and Virtual Environments*, 7(2), pp.101-107.

Dimarogonas, A. (1996). *Vibration for Engineers. 2nd Ed.* Prentice Hall.

Golledge, R.G. (1999). Human wayfinding and cognitive maps. In: Golledge, R.G. (Ed*), Wayfinding behavior: cognitive mapping and other spatial processes*. The Johns Hopkins University Press.

Lasswell, JW & Wickens, CD (1995) The effects of display location and dimensionality on taxi-way navigation. *TR-ARL-95-5/NASA-95-2.* Univ Illinois Aviation Res Lab.

Wang, W. and Milgram, P. (2001). Dynamic viewpoint tethering for navigation in large-scale virtual environments. *Proc 45th Annual Mtg of Human Factors and Ergonomics Society*. Minneapolis, Minnesota.

Wang, W. and Milgram, P. (2002). Viewpoint optimisation for navigation using dynamic tether. *Proc 46th Annual Mtg Human Factors & Ergonomics Society*. Baltimore.

Wang, W. and Milgram, P. (2003). Effects of viewpoint displacement on navigation performance in virtual environments. *(Proc. 47th HFES Annual Mtg, Denver.)*