Telepresence and Artificial Reality: Some Hard Questions

Thomas B. Sheridan
Massachusetts Istitute of Technology
Cambridge, MA 02139, USA

Abstract. This paper proposes three measurable physical variables which determine telepresence and virtual presence. It discusses several aspects of human performance which might be (differentially) affected by these forms of presence. The paper suggests that distortions or filters in the afferent and efferent communication channels should be further tested experimentally for their effects on both presence and performance. Finally it suggests models by which to characterize both kinematic and dynamic properties of the human-machine interface and how they affect both sense of presence and performance.

Introduction

There are many new technologies emerging which are likely to have large impact on both teleoperation (human remote control of vehicles, manipulators and other systems using video, audio, kinesthetic and tactile feedback from the remote site) and the experience of artificial reality (computer-simulated environment from which the user can try out alternative actions and get realistic feedback). In either case, given sufficiently high-fidelity display, a mental attitude of willing acceptance, and a modicum of motor "participation" (see discussion below) the human operator experiences "telepresence" (sense of being physically present with virtual object(s) at the remote teleoperator site) or "virtual presence" (sense of being physically present with visual, auditory or force displays generated by a computer). Figure 1 illustrates these terms and their interrelations.

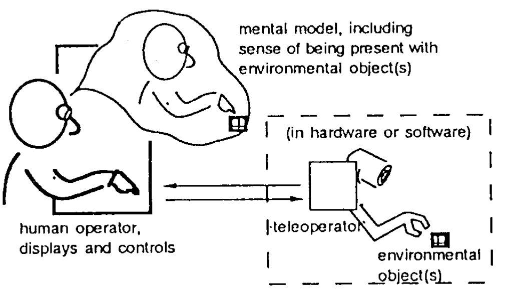

Figure 1. In telepresence, the displays and operator actions result in a mental model which includes the sense of being present with actual environmetnal object(s). In virtual presence, software substitutes for actual teleoperator and environmental object(s).

The new and driving technologies are high-density video, high-resolution and fast computer-graphics, head-coupled display, instrumented "data gloves" and "body suits", and high-bandwidth, multi-degree-of-freedom force-feedback and cutaneous stimulation devices. Virtual environment devices are now being integrated into actual control stations in the field, so that the operator can use such virtual "what-would-happen-if" exercises for *in-situ* planning and embedded training. Perhaps we are at the threshold of a dramatic improvement in remote control and training capability. Or perhaps we are simply victims of the technological imperative ("invention is the mother of necessity").

Questions needing answers

Allegedly, such sophisticated, high fidelity human interfaces, characterized by terms like "telepresence" in the case of teleoperation (feeling like you are actually "there" at the remote site of operation) and "virtual presence" (feeling like you are present in the environment generated by the computer), will improve sensory-motor or cognitive performance and efficiency of training and planning. Yet at present we have very little fundamental understanding of what effect such conveyors of "presence" do for us. The literature does not even offer us a useful measure of "presence", though Newell (Sheridan et al, 1987), Held and Durlach (1987) and others have stressed the need for an objective measure.

At present we have no theory of presence, let alone a theory of telepresence or virtual presence. This is spite of the fact that students of literature, the graphic arts, the theater arts, film and TV have long been concerned with the observer's sense of presence. In fact, one might ask, what do the new technological interfaces add, and how do they affect this sense, beyond the ways in which our imaginations (mental models) have been stimulated by authors and artists for centuries?

In controlling actual vehicles or telerobots, where we absolutely depend upon feedback, we have inadequate knowledge of how the operator's sense of presence contributes either to ultimate performance -- apart from providing necessary feedback for control. Is what really matters the bits of information, coded with sufficiently high resolution in stimulus magnitude and time and space, and displayed to the appropriate sensory modality? And is sense of "presence" simply a concomitant benign phenomenon, or even a distraction? Or is the quality of "presence" the critical psychological indicator of physical stimulus sufficiency?

Similar questions are appropriate for training and learning. Knowledge of results is known to be essential. A sense of participation and involvement is known to help a great deal. When simulation and virtual environments are employed, what is contributed by the sense of presence *per se?* When simulation and virtual environments are used in conjunction with actual vehicles or dynamic systems in the field, providing embedded or *in situ* training/planning capability, we do not have design/operating principles for how best to use the virtual and actual capabilities together.

Suggested here are some measures of presence, the crude beginnings of a theory of presence, and some experimental approaches to help sort things out.

Toward an operational measure of presence

There is an obvious need for a measure of "presence" which is operational (repeatable by anyone using proper procedure), reliable, useful (can be applied) and robust. "Presence" is a subjective sensation, much like "mental workload", and mental model -- it is a mental manifestation, not so amenable to objective physiological definition and measurement. Therefore, as with mental workload and mental models, subjective report is the essential basic measurement.

Subjective measures need not be unidimensional. Like so many rating scales, if the dimensions (factors) are carefully qualified, subjects should be able to give reliably different responses for the different factors to the same stimulus. This has certainly been true of mental workload, and multifactor rating procedures (Sheridan and Simpson, 1979) now form the basis of mental workload applications. Multiple dimensions can also be inferred from unidimensional ratings of "psychological distance" between stimuli (Shepard et al, 1972).

One likes to have objective measures as well. A good (by the above standards) objective measure can often be obtained more conveniently than a subjective one. Some people, e.g., physical scientists and engineers tend to feel more comfortable with objective measures. One type of measure which has face validity is response to unexpected and perhaps threatening stimuli, as suggested by Held and Durlach (1987). If a virtual object is suddenly seen (and/or heard binaurally to be) on a collision course with one's head does the subject blink, or duck? If the subject is controlling a virtual automobile displayed to be heading for a crash, or a virtual aircraft showing signs of stalling, does the operator take immediate appropriate action? Another type of measure might be a socially conditioned response to virtual social encounters: unpremeditated reaching out to grasp something being handed over, or to shake hands or cheer, or an utterance of "Hello", "I beg your pardon" or "Gesundheit", as appropriate.

Three determinants of presence

I propose that there are three principal determinants of the sense of presence:

(1) **extent of sensory information** (the transmitted bits of information concerning a salient variable to appropriate sensors of the observer);

(2) **control of relation of sensors to environment** (e.g., ability of the observer to modify his viewpoint for visual parallax or visual field, or to reposition his head to modify binaural hearing, or ability to perform haptic search); and

(3) **ability to modify physical environment** (e,g., the extent of motor control to actually change objects in the environment or their relation to one another).

These determinants may be represented as three orthogonal axes (see Figure 2) since the three can be varied independently in an experiment. Perceived extent of sensory information is sometimes regarded as the *only* salient factor. Sometimes the other two are lumped together as "user interaction" (Zeltzer, 1990). Figure 2 shows "perfect presence" as the maximum of all three, though it is far from clear by what function "presence" is determined by combinations of the three. It surely is not a simple vector sum.

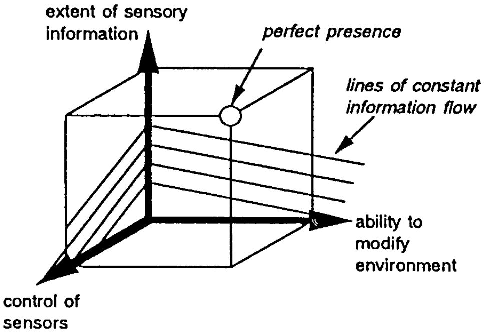

Figure 2. Principal determinants of sense of presence for a given task

Lines of constant information communicated are suggested in the figure to indicate that the "extent of sensory information" is a much greater consumer of information (bits) than are the two control components, "control of sensors" and "ability to modify environment".

Major task variables

I am not suggesting that the three principal determinants of presence operate alone. They are surely task-dependent. It seems to me there are two major properties of tasks which affect behavior, both subjective and objective. These I call: (1) **task difficulty**; and (2) **degree of automation**. Task difficulty may be defined in terms of entropy measures, such as Fitts' *index of difficulty* (Fitts, 1954). Degree of automation means the extent to which the control of the task (the ability to modify the environment) is automatic as contrasted to being manual. There is a scale from manual to automatic, where intermediate levels of automation are normally called supervisory control (Sheridan, 1987).

Dependent variables: from presence to performance

Given the three independent determinants of presence I see the larger research challenge to be the determination of the dependent variables: (1) **sense of presence**, as measured by (a) **subjective rating**, and (b) **objective measures** suggested above; (2) objective **training efficiency**; and (3) ultimate **task performance**. This mapping is illustrated in Figure 3.

Tasks in which presence, learning and performance interrelate take many forms in time and space and stimulus intensity (e.g., brightness, loudness, force). Sometimes the

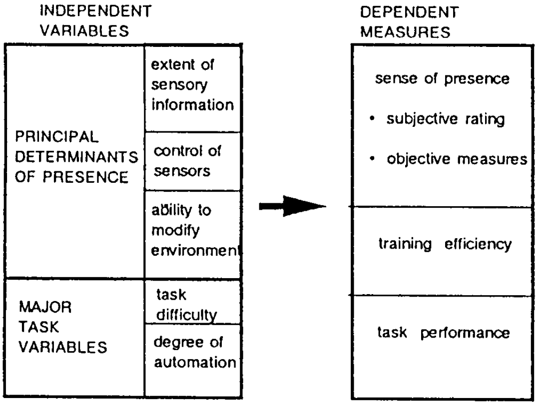

Figure 3. Experimental determination of presence, learning efficiency and performance

dynamic aspects of sensing and control are the most important. Sometimes it is the relation of stimuli in space. It may be useful to distinguish these and consider several cases. However, before considering examples of either the temporal or spatial relations, it is important to distinguish the relative effects of efferent information and afferent information upon presence, training and performance.

Afferent and efferent filtering: independence and coupling

Causality of events in a human-machine system is commonly characterized as a closed information/control loop: the efferent or motor loop from human response to environmental state, plus the afferent or sensory loop from environmental state to human senses. This mapping from operator to (actual or virtual) vehicle or teleoperator environment and back again can be generalized as two filters, an **efferent filter** and an **afferent filter** respectively, where the concept of filter can embody all forms of distortion in mapping, including distortions in both time and space. This is represented in Figure 4.

The following general question is immediately suggested: What are the relative effects of efferent or afferent distortions on sense of presence, training efficiency and performance? I submit that we have very poor experimental answers to this question, for other than relatively trivial distortions such as pure time delay. When, in purely mechanical systems, distortions can be couched in sufficiently tractable mathematical form (e.g., linear differential equations) control theory can predict the relative effects of forward loop or feedback loop filtering on closed loop performance. Such analyses can provide a starting point for modeling effects of afferent and efferent distortions in human-machine systems.

In the case of human control loops, we have relatively poor information, to a large extent because sensation and perception tend to be characterized in very different ways from motor responses. However the fact remains that the relative sensitivity of closed-loop performance may be enormously different for the same effort or dollars spent on improving the feedback as compared to improving the feedforward.

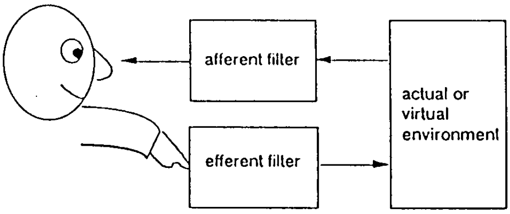

Figure 4. Loci of efferent and afferent distortions in human interaction with actual and virtual environments.

We do know, for example, that in very simple continuous visual-motor tasks it usually does not matter with respect to final closed loop control whether a pure time delay is in the forward position-command loop or in the visual feedback loop. However for any given element of human response the time of closed-loop environmental state change is obviously different by one time delay.

Coupling in touch and kinesthesis: effects of dynamic filtering

When touch and kinesthesis are involved, unlike vision and hearing, there is an energy-port couple relation between force and velocity (or displacement) so that what constitutes input to and what is output from the human is an arbitrary. In any case, if one is considered input the other must be output.

In such a case it is probably more appropriate to resort to a filter model in which what is forward and what is feedback are determined by the governing equations for the coupling, with an arbitrary assignment of causality between force and position. A simple linear lumped-parameter model based on this idea is shown in Figure 5, where manual position Xm and applied force Fm form the energy couple at the human end, while manipulator position Xe and force Fe form the corresponding couple at the environment end. In a teleoperation context such a model helps us answer questions about how the various parameters (which pose different

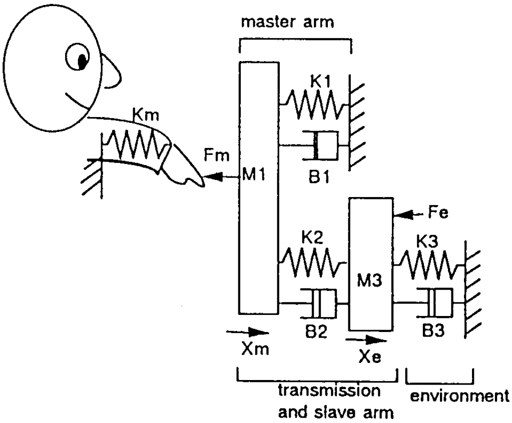

Figure 5. Lumped parameter model of tactile / kinesthetic coupling between afferent and efferent filters.

forces and dynamics) affect an operator's ability to (1) reproduce positional changes at the slave by repositioning the master (requiring absolute judgments of position or force) or (2) discriminate small position-force changes of the slave by feeling position-force changes in the master (differential judgments). From a model such as is shown below one can predict, for example, Fm / Fc force sensitivities for given filter parameters and to some extent predict results of psychophysical experimental. A simple linear dynamic analysis is given in the Appendix. Tactile-kinematic performance (and sense of presence and learning) will be affected much as vision or hearing are affected by the dynamic conditions of the visual-auditory stimulus pattern and surround.

A more general energy-port model of a master-slave manipulator was performed by Raju et al (1989), considering the dynamics of the operator neuromuscular system as well as those of the master and slave and task environment. They put to rest simple claims that the impedance of the master arm (as felt by human arm) should always be the same as the impedance of the slave arm (as felt by the environment) or the impedance of the environment (as felt by the slave arm). Clearly what is best depends on the task to be done. Hammering a nail, threading a needle, and detecting small changes in environmental stiffness are best done with quite different different combinations of master and slave impedances.

Spatial filtering: effects of distortions from spatial isomorphism on sense of presence, training and performance

Figure 6 portrays as vectors the position and orientation of elements local to the operator (operator's body, head, control, display screen, displayed arm and object) and those remote from him, whether real or in a computer model (vehicle, sensor, arm of teleoperator, object manipulated).

The research question here is: how do the geometrical mappings of body and environmental objects both within the perceived (virtual) environment and the true one, and relative to each other, contribute to sense of presence, training and performance?
Control by the human operator which requires some such mappings may be easy, while others may be very difficult. Some own-body to teleoperator/remote-environment control tasks (or own-body to virtual-operator/virtual environment tasks) may demand a high degree of isomorphism. In some cases there may be a need to deviate significantly from strict geometric isomorphism because of hardware limits, or constraints of the human body. At present we do not have design/operating principles for knowing what mapping or remapping from the lower set of vectors to the upper, or back again for feedback are permissible, and which degrade performance.

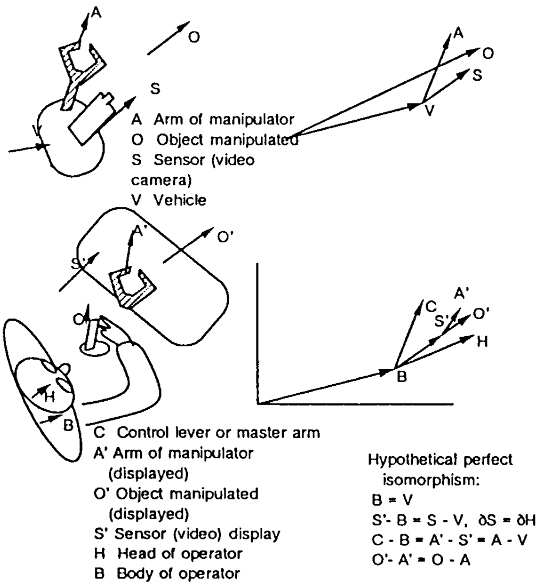

Figure 6. Vector representation of spatial isomorphisms.

References

Fitts.P.M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *J. Exp. Psychol.*, 47, 381-391.

Held, R. and Durlach, N.(1987). Telepresence, Time Delay and Adaptation, *NASA Conference Publication 10032*, chapter 28.

Raju, G.J., Verghese, G., and Sheridan, T.B. (1989). Design issues in 2-port network models of bilateral remote manipulation, *Proc. IEEE Intl. Conf. on Robotics and Automation.*

Shepard, R.L., Romney, A.K. and Nerlove, S.B. (1972). *Multidimensional scaling,* Vol I: Theory, NY: Seminar Press.

Sheridan, T.B and Simpson, R.W. (1979). *Toward the definition and measurement of mental workload of transport pilots.* Flight Transportation and Man-Machine Systems Laboratories, Technical Report DOT-OS-70055. Cambridge, MA: MIT.

Sheridan, T.B. (1987). Supervisory control. In Salvendy, G., *Handbook of Human Factors.* pp.1243-1268, New York: Wiley.

Sheridan, T.B., Kruser, D.S. and Deutsch, S. , Eds.(1987), *Human Factors in Automated and Robotic Space Systems: Proceedings of a Symposium,* Washington, DC: National Research Council, p.324.

Zeltzer, D.(1990). Presentation given at Symposium on Media Technology, Cambridge, MA: MIT, Oct. 2 and 3.

Appendix: Linear dynamical analysis of figure 5

A dynamical analysis of the linear model reveals the following:

$$F_m = (M_1S^2 + B_1S + K_1)\,x_m + (M_3S^2 + B_3S + K_3)\,x_c + F_e \qquad \text{and}$$

$$(B_2S + K_2)(x_m - x_c) = (M_3S^2 + B_3S + K_3)\,x_c + F_e$$

Solving for x_c in the second equation and substituting for x_c in the first yields

$$F_m = (M_1S^2 + B_1S + K_1)\,x_m + \frac{[M_3B_2S^3 + (B_2B_3 + M_3K_2)S^2 + (B_2K_3 + B_3K_2)S + K_2K_3]\,x_m + (B_2S + K_2)\,F_e}{M_3S^2 + (B_2 + B_3)S + (K_2 + K_3)}$$

If all terms are finite this means the feeling of every environmental force component is modified by properties of the intermediate teleoperator mechanics and filtered through a damped oscillatory filter.

For K_2 large (i.e., a rigidly connected master and slave),

$$F_m = [\,(M_1S^2 + B_1S + K_1) + (M_3S^2 + B_3S + K_3)\,]\;x_m + F_e$$

and so the local and environmental damping and stiffness terms simply add. There is no way to distinguish slave from master forces in this case.

For K_3 and B_3 both = 0 (i.e., the slave has no contact with the environment),

$$F_m = (M_1S^2 + B_1S + K_1)\,x_m + \frac{(M_3B_2S^3 + M_3K_2S^2)\,x_m + (B_2S + K_2)\,F_e}{M_3S^2 + B_2S + K_2)}$$

In this case if $B_2 = 0$,

$$F_m = (M_1S^2 + B_1S + K_1)\,x_m + \frac{(M_3K_2S^2)\,x_m + K_2\,F_e}{M_3S^2 + K_2)}$$

which means environmental mass and stiffness are felt in combination and through an undamped oscillation, and so too are the unbalanced forces.

If for the no contact situation $K_2 = 0$,

$$F_m = (M_1S^2 + B_1S + K_1)\,x_m + \frac{(M_3B_2S^2)\,x_m + B_2\,F_e}{M_3S + B_2}$$

which means environmental mass and damping are felt in combination and through a first order lag, and so too are the unbalanced forces.