Further Musings on the Psychophysics of Presence

Thomas B. Sheridan
Massachusetts Institute of Technology

Abstract

This is an extension of an earlier paper (Sheridan 1992) which considered alternative meanings and significance of "presence", the experience of "being there", commonly called "telepresence" in the case of remote control or teleoperation, and called "virtual presence" in the case of computer-generated simulation. In both cases presence can include feedback to the human senses of vision, hearing and haptics, both kinesthetic and cutaneous. Presence is discussed here in terms of alternative subjective meanings, operational measurements, and meaningful experimental comparisons. Three practical approaches to measurement of presence are discussed, including elicitation of "natural" neuromuscular or vocal resposes, single or multidimensional subjective scaling, and ability to discriminate the real and immediate environment from that which is recorded /transmitted or synthesized, under varying levels of constraint. The author also opines on the stimulus magnitude, space and time attributes of human interactions with a tele- or virtual environment.

Introduction

In recent years there has been intense interest in "presence" in its two manifest forms, telepresence, wherein the human participant feels herself to be present at a location other than that which is actual (real and immediate), and virtual presence (or virtual reality or artificial reality), wherein the human participant feels herself to be present at a location which is synthetic, created only by a computer and various visual, auditory or haptic displays. However novel and compelling are the experiences wrought by the new technology, and in spite of the fact that philosophers, theologans and drama teachers have for many years contempated "presence", the fact remains that currently there is no accepted theory or operational measurement of presence (Newell 1987).

Figure 1, from Sheridan 1992, illustrates the notion that from the human operator's vantage point, whether it is a teleoperation or an interaction with a virtual environment, the human interface is likely to be the same, and the human's mental model in either case is likely to be coping with suspension of disbelief and/or the compulsion to believe that one is apparently located in space other than where one physically exists.

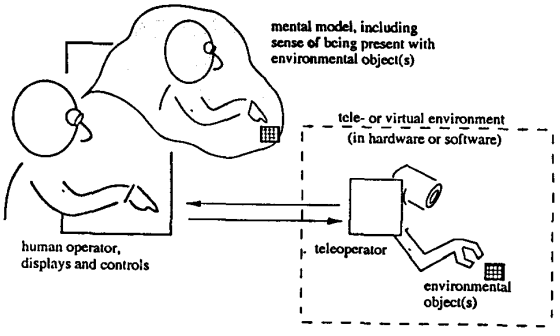

Fig. 1. Telepresence and virtual presence.

Physical presence
- physical evidence for existence
- physical evidence for location "there"

Divine presence:
- miracles (or unexpected events)
- salvation (or luck)

Personal presence:
- attention elicitation
- good communication

Telepresence and Virtual Presence
- sense of "being there"
- natural (expected) responses of human and environment to each other

Fig. 2. Common meanings of "presence".

Philosophers for centuries have been concerned with physical presence, including (Figure 2) evidence that one exists at all, as well as the basis for determination of one's physical location. The theologian's as well as the layman's concept of divine presence is often premised on apparent miracles, or events in which one is somehow saved from ill fate through otherwise unexplainable forces.

0-7803-2129-4/94 $3.00 © 1994 IEEE

The drama or speech teacher refers to presence in terms of a speaker's or actor's ability to "project", to keep audience attention and interest, and in general to communicate well. Tele- and virtual presence, by contrast, have to do with perception of where one is physically located and the responsiveness of both the human and the environment to one another.

In this paper we mean presence to include feedback from the tele- or virtual environment to all of the salient human senses, including visual, auditory and haptic (including both kinesthetic senses in the muscles, tendons and joints, as well as the tactile senses mediated by sensors in the skin.

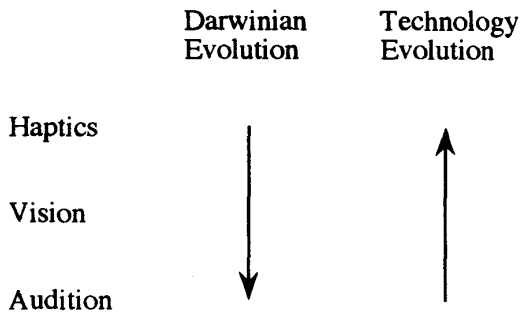

Fig. 3. Darwinian vs. technological evolution.

It is interesting to note that, as illustrated in Figure 3, while Darwinian evolution saw force senses in the limbs and skin of primitive animals long before the exteroceptors of light and sound developed, human technology has developed in the reverse order. First we got the telephone, radio and stereo "hi-fi", then video and computer graphics, and only recently have artificial tactile sensors been developed. No satisfactory tactile display exists even now. In other words, man seemingly has not understood or been able to duplicate what nature accomplished quite early in evolution.

Measurement of Presence

Thus far, to the writer's knowledge, three methods have been proposed to measure presence, whether tele or virtual. Purported contributions of this paper include suggestions of practical means to effect the latter two methods.

1. Reflexive responses. The first of these is the measurement of human response to events which in the natural world normally evoke "reflex" response (Held and Durlach, 1987). If a ball is thrown to a person he or she will automatically reach out to catch it. If a fist or other missile threatens a collision course with one's head, one will "duck" without thinking. Similarly there are reflexive social responses. If one inadvertently collides with someone on the sidewalk or in a meeting room, one immediately utters "Pardon me" or something similar.

2. A three-attribute subjective category rating scale. A second type of measurement, quite obvious and surely proposed by some one or more authors, is subjective scaling. The most obvious scale would be one-dimensional, perhaps of five to ten categories characterized by verbal descriptions of the experience or the degree of realism experienced. However, since tele/virtual presence has many attributes, it would seem that a multi-dimensional subjective category scale would be appropriate. The question then is: what are the dimensions with respect to which a presence experience varies? If one asserts the applicability of a multi-dimensional scale, then one must justify that those dimensions are functionally orthogonal. This is not unlike the scaling of mental workload, another phenomenon which is essentially experiential and not easily parameterized. For example, Sheridan and Simpson (1979) first suggested scaling mental workload of air transport pilots in three orthogonal dimensions, one having to do with time availability or "busy-ness", one with complexity of the problem being solved, and one with emotional stress. Though these dimensions are not themselves appropriate to tele or virtual presence, they are suggestive. Let us turn now to scaling the experience of presence, tele or virtual.

Three seemingly orthogonal interactive properties of presence were suggested (Sheridan 1992) to be: (1) information content of the stimulus independent of the observer; (2) ability of the observer to freely modify the "viewpoint" of the eyes, ears or haptic senses relative to the tele- or virtual environment; and (3) ability of the observer to modify the configuration of the environment.

Information content of a visual stimulus, for example, means: frame-rate (frames per second), resolution (pixels per frame), and color or gray scale (bits per pixel) — the product of which is bits per second. There are auditory and haptic counterparts.

In vision, viewpoint is normally modified by casting one's gaze in different directions, or changing one's head position laterally, so as create parallax and thereby achieve a better sense of depth (range) than is provided by accommodation, stereopsis, size constancy, shadows or other cues. Realistic response to a viewpoint change is what a hologram allows that an ordinary photograph does not. Viewpoint change is also what a head-mounted display (HMD), the position-rotation of which drives the video camera or computer graphic image, does. Therefore the HMD is regarded as a principal provider of "presence" and has become the manifest symbol of presence.

Auditory "viewpoint" is achieved by moving the head (and ears) in relation to an auditory source, or electronically manipulating the read-related transfer function to create the image of head movement or sound source movement or both. Haptic "viewpoint" change is produced by moving the hand to a new configuration or pattern of contact with an environmental object (thereby

changing the set of tactile receptors that are exposed to some pattern of forces).

The third interactive property of presence, ability of the human to modify the configuration of the environment, is not provided by a hologram or a head-mounted display, but only by telemanipulation of objects in a real environment or manipulation of a virtual environment, wherein the operator can actually move objects or otherwise reshape the environment.

It is claimed that these three interactive properties are operationally orthogonal, that is, any one can be changed without affecting the other. For this reason it is here suggested that a subjective scale be based on these three properties, with three to five categories verbally defined for each. Figure 4, as an example, suggests some simple verbal descriptors which might be used for three categories on each dimension. No doubt others will add more levels and/or modify the descriptors.

3. Presence measurement based on discrimination. A third measure of presence has been proposed by my colleague Schloerb (1994), namely the inability to discriminate between a real and a tele- or virtual environment. In this case the "perfect" tele or virtual environment is one which cannot be discriminated from a real and immediate environment. Schloerb describes an experimental test of discrimination, where the quantitative measure of interest is the probability of correct identification, real vs. "artificial" (tele- or virtual). Schloerb admits, however, that with current systems it is unlikely that one will confuse the real with the tele/virtual, and herein lies a problem. How can the current level of "somewhat-compelling-but-still-obviously-not-the-same-as real" tele- and virtual environment be evaluated for its "degree of presence"?

To cope with this problem I propose that something akin to a masking experiment be used as the measurement basis, where the magnitude of some standardized "noise" signal added to the stimulus is taken as an index of presence. However, the proposed measure is quite different from masking in the conventional psychophysical sense. In a conventional masking experiment, as the noise is added the signal strength must be increased in order for the stimulus to be just barely seen, heard or felt. The increase in signal magnitude to just reach an absolute threshold of detection or perception is defined as the measure of the masking effect of the noise. One can also say that the magnitude of the noise which must be added to reduce the signal threshold to zero (where discrimination of signal from no signal is no better than guessing) is a measure of the initial perceptual strength of any stimulus.

In what is proposed here the stronger the stimulus (in the above sense), the more easily the tele- or virtual stimulus is discriminated from the real. Therefore the magnitude of the "noise" which must be added to just reduce discrimination of tele- or virtual from real to no-better-than-guessing is a measure of initial discriminability or, equivalently, lack of presence. To say it another way, the less the "noise" which must be added to suppress discrimination to threshold the more the "presence".

FIDELITY OF PICTURE, SOUND OR TACTILE IMAGE

1. Virtual image resoution (pixels or taxels per frame), refresh rate (frames per second) and gray- or color-scale (bits per pixel or taxel) are too few to convey realism.

2. Virtual image fidelity is fairly realistic. Resolution (pixels or taxels per frame), refresh rate (frames per second) and gray- or color-scale (bits per pixel or taxel) are enough to convey good sense of reality.

3. Virtual image is compelling. Difficult to discriminate the virtual from the real based on a given image.

ABILITY TO MODIFY SENSOR POSITION / ORIENTATION TO GET A DIFFERENT VIEWPOINT

1. Viewpoint with respect to the virtual environment is fixed. Unable to move head or hand to get a different image.

2. Viewpoint with respect to the virtual environment can be changed, but is limited to only a few positions or orientations or head or hand relative to the virtual environment.

3. Viewpoint with respect to the virtual environment can be changed easily to any desired position or orientation of head or hand relative to the virtual environment.

ABILITY TO CHANGE CONFIGURATION OF ENVIRONMENT

1. Unable to effect any changes in the environment such as move objects with respect to one another, change shape of an object, etc.

2. Able to modify environment in a limited way, either as to what aspects may be reconfigured or the difficulty in effecting the change.

3. Able to modify environment freely as easily as one would with bare hands or tools in a real environment.

Fig. 4. A three-attribute scale of presence.

Questions which then arise regarding the proposed measurement are whether the "noise" should be added to both real and tele- or virtual images or just to the real image, and moreover, what type of "noise" to use. One might simply degrade the real image only, until, presumably, it was indistinguishable from the tele- or virtual. The difficulty with that procedure is that one would have to select just the right attribute or attributes with respect to which the two images differed, such as resolution, frame-rate, or gray or color scale, and one might predict that in practical simulations such as are used for training, etc. it would be extremely difficult to find those attributes and effect the degradation to the point of no-better-than-guessing. It seems that it would be better to apply the same "noise" to both images, and that in this case it would not take too long before discrimination was impossible, even with initially highly discriminable images.

With regard to what type of "noise" to use (the quotes around "noise" are intentional), i.e., what stimulus attributes to degrade in both images simultaneously (as the "noise" is increased in "magnitude"), I believe it best to degrade multiple attributes simultaneously, for example resolution, frame rate, gray scale, color scale, and maybe others. A computer would be used to perform this multi-attribute filtering operation so that each attribute was gradually constrained by small increments. Thus the "noise" would not be additive noise at all, but actually would be a multi-dimensional filtering operation. I believe that if both real and tele- or virtual images are gradually degraded simultaneously with respect to several standardized attributes (for example so that the each contributed in equal proportions to the bit-rate degradation), the initially "compelling" tele- or virtual image would very soon become indistinguishable from the real image, and the initially not-so-compelling image would take systematically longer to distinguish from the real — no matter what the relative fidelity of the tele- or virtual image in each component attribute.

Fidelity of the interaction in magnitude, space and time

This last discussion of the nature of the "noise" which might be imposed upon the real and tele- or virtual images to be discriminated quite naturally begs the question about the independent effects of reducing tele- or virtual environment interaction fidelity in magnitude, space or time — which we represent qualitatively in Figure 5. The fact is, we have very little controlled experimental evidence about these effects (which certainly is one reason for the suggestion in the last section that the standardized degradation procedure include a component of degradation in all three).

Let us now examine how the interaction with a tele- or virtual environment might be distorted intentionally (experimentally) or unintentionally (due to imperfect hardware/software) in each of the three dimensions (magnitude, space and time) independently. We might be concerned with measuring the subjective sensation of presence (using any of the three measures

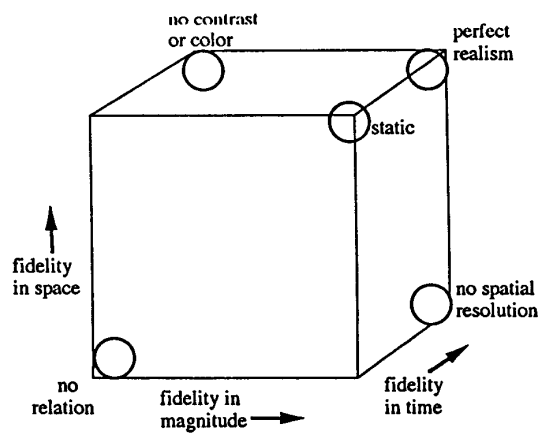

Fig. 5. Components of interaction fidelity.

proposed above) or we might be concerned with measuring performance, or we might be interested in comparing performance to subjective presence. Let us further consider the relevance of magnitude, space and time fidelity on the three interactive properties of virtual or tele-environment (posited in the 1992 paper), which I claim contribute to the subjective experience of presence. There are, to repeat: (1) image information independent of the observer, (2) observer sensor orientation, and (3) observer capability to change the environment.

Perfection of a tele- or virtual image with respect to all three properties must be located at the upper right corner, where the stimulus magnitude at each point in the stimulus field is the same as the real reference (X axis), the image is continuous in space (Y axis), and is continuous in time with no delay (Z axis). Any degradation beyond threshold of noticeability is expected to affect the sense of presence. Extremes of infidelity are labeled, and one might designate, at least ordinally, intermediate levels of degradation.

(1) Image degradation of that information which is independent of the observer. This will be affected in obvious ways by degradation in magnitude resolution in any of the various stimulus dimensions (e.g., gray-scale or color, sound loudness, mechanical force on the skin), by spatial resolution (e.g., pixels for a visual image, taxels for tactile image, and I'm not sure the counterpart for an auditory image even has a name) and by time resolution (frame rate for a visual image, or bandwidth of auditory signal or tactile vibration).

(2) Degradation of the image response to the observer-sensor orientation. This interaction concerns a mapping relation from position/orientation of human sensors (eyes, ears, skin) to a realistic sensed image. While the fidelity of the image itself is subsumed under (1) above, fidelity in this category (2) is fidelity of measuring the sensor position/orientation. Thus only fidelity in space and time are relevant. One is typically concerned here with the lag or time delay of the tele-environment "catching up" to what is expected, due to electromechanical tracking dynamics of the teleoperator visual or tactile sensors relative to the operator's head or hand movement. In a virtual environment the culprit is usually the speed with which a new image can be generated. One may also be concerned with the tracking offset, where the tele-or virtual image does not come to rest where it should, or with jerkiness in its motion.

(3) Degradation of the observer capability to change the environment realistically. Fidelity in this case means that forces are applied to the tele- or virtual environment at a magnitude, position in space (orientation) and time corresponding to what the operator's limb or hand tool does in the local or real environment, and those forces have the same effect upon modifying the environment they would have in a local or real environment. Here one is typically concerned about real-time dynamic fidelity of the telemanipulator arm/hand, vehicle, or other process the operator is controlling. Is it behaving as it should, given the operator's inputs? Accuracy of equations of motion for an aircraft of driving simulator would be considered here.

The above arguments could be extended to other stimulus modalities of a tele-or virtual environment, including vestibular, thermal, olfactory, gustatory, and so on, but clearly the technology is not ready for such extension.

Conclusion

Three psychophysical measures of the subjective experience of presence (as in telepresence or virtual presence) have been discussed, and practical implementations proposed. The components of the fidelity of human interactions with such environments, namely stimulus magnitude, space and time, have also been discussed in relation to previously cited properties of interaction between humans and tele- or virtual environments.

References

- Sheridan, T.B. (1992) Musings on telepresence and virtual presence, *Presence* Vol.1, No. 1, Winter, 120-126.
- Newell, A., a comment in Sheridan, T.B., Kruser, D.S. and Deutsch, S. Eds. (1987) *Human Factors in Automated and Robotic Space Systems: Proceedings of a Symposium*, Washington, D.C., National Research Council, p. 324.
- Held, R. and Durlach, N. (1987) Telepresence, time delay and adaptation, *NASA Conference Publication 10032*, Ch. 28.
- 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, Tech. Rep. DOT-OS-70055, Cambridge, MA: MIT.
- Schloerb, D.W. (in press) A quantitative measure of telepresence, *Presence*.