Towards Transparent Telepresence

Gordon M. Mair

Transparent Telepresence Research Group, Department of Design, Manufacture, and Engineering Management, University of Strathclyde, 75 Montrose Street, Glasgow G1 1XJ, Scotland, UK
g.m.mair@strath.ac.uk

Abstract. It is proposed that the concept of transparent telepresence can be closely approached through high fidelity technological mediation. It is argued that the matching of the system capabilities to those of the human user will yield a strong sense of immersion and presence at a remote site. Some applications of such a system are noted. The concept is explained and critical system elements are described together with an overview of some of the necessary system specifications.

Keywords: Telepresence, teleoperation, displays, immersion, mixed-reality.

1 Introduction

This paper is concerned with immersion in a remote real world environment, literally tele-presence, as opposed to immersion in a computer-generated environment as found in virtual reality. The author will therefore argue that it is justifiable to assert that a state of transparent telepresence, in which the technological mediation is transparent to the user of a telepresence system, can be closely approached through the use of technology alone. It is recognised that “immersion” is not necessarily “presence” and that other factors also contribute to a sense of presence. However it is suggested that in this telepresence context they are of secondary importance to the fidelity of the human–computer/machine/product interface and associated communication and remote site equipment.

1.1 Defining Transparent Telepresence

Over half a century ago the science-fiction author James Blish [1] explained the concept of telepresence in a short story about a worker, located on a moon of Jupiter, who carried out work through remote control of a vehicle in the planet’s atmosphere. He even used the equivalent of a head-mounted-display. The term itself was first mooted in print by Marvin Minsky [2]. Today the term has come to be used in a manner that has wider connotations. This is despite a number of writers providing useful and practical definitions throughout the last one and a half decades, e.g. [3], [4], [5]. What is apparent is that today “presence” is often, but not always, used as an abbreviation for “telepresence” and it is also inclusive of virtual reality and the experience of being immersed in a virtual world. The explication of “presence” is still
Towards Transparent Telepresence

a matter for discussion with many scholars offering interesting perspectives, for example the International Society for Presence Research web site [6] has a comprehensive statement on the topic. Also Sadowski and Stanney discuss the characterisation of presence in [7], in this they provide a comprehensive literature review and suggest that “one must be afforded both technological and experiential immersion to maximize the likelihood of high levels of presence”.

This paper acknowledges the widely accepted view that in the general case technological fidelity is insufficient for full presence, e.g. Heeter [8] states “perfectly mediated sensory stimuli do not automatically induce continuous presence”. She discusses how the sense of presence can be different for different individuals and vary from moment to moment. Sanchez-Vives and Slater [9] and Welch [10] point out that the amount of control we have over the remote environment is also important. However it is this author’s contention that **transparent telepresence** [11] is a special case and may be defined as **the experience of being fully present interactively at a live real world location remote from one’s own physical location**. In this situation the technological mediation should provide both the multi-sensory input necessary to represent the remote environment, and the ability to interact with that environment in a natural and instinctive manner, thus providing most of the criteria for a sense of presence. Transparent telepresence is therefore a goal rather than an existing state, it is a goal that is unlikely to be achieved for decades but it does provide an objective and quantifiable target.

Another justification for adopting this technological approach is that the experience of presence in itself is difficult to measure, particularly if it is “second order” [6], [12], i.e. technologically mediated experience as opposed to “first order” mediated only by human senses. In attempting to measure presence numerous questionnaires have been created, e.g. [13] and their validity discussed e.g. [14] which includes a comment that other approaches such as ethnographic studies could be more appropriate. Witmer et.al [15] provide a review of the chronological development of questionnaires since the early nineteen nineties, also Ma and Kaber [16] include comments on the difficulty of using questionnaires and measuring presence in general. Finally in [17] Slater argues that post experience questionnaires alone cannot be used to assess presence in virtual environments. However in adopting the ultimate goal of transparent telepresence the question of how to measure presence becomes redundant since transparent telepresence will, by definition, allow the system user to feel present at the remote site. This technological approach can ultimately only provide perceptual, rather than physical, equivalence of the remote site. To provide even apparent physical equivalence a means of bypassing our normal sensory system and sending information on the remote site directly to the appropriate locations in the brain would be necessary as happens with VR in fictional stories such as Neuromancer [18].

As a final comment in this section it should be noted that, particularly in Europe, the term “Mixed Reality” is currently popular. This follows from the observation that technologically mediated experiences occur along a Reality-Virtuality continuum [19]. At one end of this is unmediated reality and at the other is a fully computer generated virtual environment. As we move along the continuum from the real environment we experience Augmented Reality, this occurs when computer generated information can be superimposed on the real world. We can then move further along
and find Augmented Virtuality where the computer generated environment may have real world entities superimposed. Moving further we find at the extreme end we have a fully computer generated environment. An interesting example of a mixed reality system is described in [20] where a tripod mounted display is used to provide real world images of a site of historical interest combined with computer generated images of the scene as it would have existed in historical times. Currently a number of large-scale European funded projects focusing on mixed-reality are underway, information on these and further links can be found at [21]. Transparent telepresence will sit very close to Reality on this continuum since although it constitutes a technologically mediated experience it is the real world in real time that is being considered albeit at a site physically removed from the system user.

Ultimately a full Product Design Specification should be achievable for a Transparent Telepresence System based on technological transparency. This goal perhaps appears contrary to the zeitgeist in presence research however since transparent telepresence is a special case of presence the author suggests that this approach of attaining a high fidelity representation of the mediated environment will provide an “absolute” specification or datum, from which to compare other systems at an objective and unambiguous level.

1.2 Application and Justification of Transparent Telepresence

The ability to apparently exist interactively in a remote real world location has many applications, a more detailed account can be found in [22]. Applications include the ability to carry out work in remote hazardous or unpleasant environments, e.g. bomb disposal and decommissioning of radioactive nuclear facilities; remote surgery, particularly where a specialist would find it impossible to be physically present such as in a battlefield; space and planetary working although transparency will never be fully achieved due to the distances and subsequent communication time delays involved. Entertainment and education could be exciting applications where the system user could experience coral reefs, rain-forests, foreign cities and people, as though physically there. A trend that is emerging in the developed countries is that their demographics are changing such that average age the population is increasing as the percentage of older people increases. With advancing age also often comes a decrease in mobility, thus transparent telepresence could provide a means for this part of the population to continue contributing to society, earning income, and end generally enjoying a better quality of life. This would also apply to those with disabilities that may prevent them participating in many activities they would otherwise pursue. Finally business applications where eye-to-eye contact and one to one conversations or round table meetings are involved could produce economic and environmental rewards. For example CO$_2$ emissions would be reduced by reducing the need for air travel, also the actual cost of the travel would be removed. Additionally unusual situations can arise that restrict travel, e.g. where there is a terrorist alert, disruptive weather, or even concern over a possible pandemic.

1.3 Structure of a Transparent Telepresence System

The basic elements of a transparent telepresence system are shown in Figure 1. In a basic system there is the home site, the remote site, and a communications link
Towards Transparent Telepresence

In order to address the problem of technological transparency the elements have to be able to provide a high fidelity interface between the hardware and the human at the home site, and between the hardware and the environment at the remote site. The human machine interface requires careful consideration of the displays and controls, a wide variety of both are currently available. Similarly the sensors at the remote site have to acquire information to a degree of precision that will be suitable for technological transparency of the display and the actuators have to respond in such a manner to very closely emulate the control signals. The computer interface at both home and remote sites needs to be able to process the data being transferred at a speed sufficient for transparency, and the communication link needs to have a bandwidth capable of transmitting all of the data at a fast speed although of course over very long distances the speed of light is obviously a limiting and noticeable factor. These aspects are now considered more fully.

![Diagram](image)

**Fig. 1. Basic Elements of a Transparent Telepresence System**

## 2 Transparent Telepresence System Elements Overview

### 2.1 Home Site Human - System Interface Ergonomics

Technologically mediated experiences involve real world ergonomic problems for the system users. For example for extended operation, the working environment should usually be comfortable with respect to ambient temperature, airflow, and noise levels. However if fully transparent telepresence is to be experienced then it may be reasonably argued that these factors should be the same as the remote environment. Thus we have a dichotomy, full immersion in the remote site may cause uncomfortable working conditions for the system user, but making them comfortable will destroy the fidelity of the representation of the remote site. The solution will be dependent on the application, i.e. if it is important for application then the remote conditions can be emulated, if not then the comfort of the user will be paramount. Other aspects of the system user must be fully considered such as anthropometrics, physiology, anatomy, and psychology. For example if a head mounted display is being used then consideration must be given as to how its weight and balance on the head affects the joints and muscles of the wearer, otherwise strain and fatigue may
result. Unnatural forces on joints, tendons, and muscles must also be avoided. MacCauley Bell provides a discussion on Ergonomics in Virtual Environments in [23], and more general information on ergonomics can be found in [24] which includes established design guidelines. The basis of this technological approach is that by considering the human senses as having discrete just noticeable differences (JNDs) based on the Weber fraction [25] we can compare these directly for each sense with a technological capability that can provide a perceptual equivalent.

### 2.2 Home Site Displays and Remote Site Sensors

A wide range of displays are available for technologically mediated experience however only a few types are appropriate for transparent telepresence. For our purposes we require displays that are capable of providing immersion and also of displaying a live representation of the remote site. This should apply for as many of the senses as possible and for transparent telepresence should include visual, aural, haptic, olfactory, and vestibular, (gustatory sensing will normally be unnecessary).

Similarly at the remote site a wide variety of sensors may be used. These should be compatible with the home site displays and provide enough information for the fully immersive experience. In this paper we will assume that the resolution of the remote site sensor and the home site display are approximately the same. The exception to this will be the aural system where the Nyquist sampling theorem states that the sampling rate must be at least twice the frequency of the original signal. This theorem can be applied to all the senses but audio is the most critical in this respect.

**Visual Displays.** Considering firstly the visual sense the likely displays will either be of the head mounted display (HMD) or wide angle screen type. Humans have a field of view of just over 180 degrees horizontally and 120 degrees vertically [26], [27]. To the best of the author’s knowledge at the time of writing there is no head mounted display capable of providing this. It is possible to achieve wide-angle immersion through the use of projection systems as can be found in large format cinemas and theme parks, however none of these provide the live video required for transparent telepresence. Additionally we have the ability to perceive depth through the use of stereoscopy. This means that the display will also require the ability to present images as they would be seen by both the left and right eye of the user. With respect to resolution, humans can detect the separation of two points that are separated by 60 seconds of arc [28], this can provide an indication as to the resolution required for the display. The dynamic range of the display is also important, at any level of light adaptation the eye can see detail at a contrast ratio of 10,000:1 and therefore the ability to produce a display capable of this [29] is important for transparent telepresence. Additional aspects that need to be considered include colour fidelity, frame rate, and the need to provide a natural operation of both accommodation and vergence of the eye.

**Visual Sensors.** In order to create a stereo image two perspectives will be required. Had we simply required a relatively narrow field of view then this would not present a significant problem since two matched cameras could be used separated by the average pupillary pitch of approximately 63mm. If the system is to be for a single
user wearing a head mounted display then the problem is ameliorated since the cameras can be mounted on a motorized sensor platform whose movements are slaved to the movements of the user located at the home site, an early visual telepresence system designed by a member of the author’s Transparent Telepresence Research Group (TTRG) is described in [30]. However for a real time fully immersive large screen display for a number of viewers there are considerable difficulties. For example the cameras have to acquire an image with a wide enough field of view to provide an immersive experience to all of the viewers simultaneously. To achieve this at a high enough resolution, without image distortion, for technological transparency it is likely that a cluster of cameras would be required. However two of these clusters would be necessary for stereoscopic viewing and if a spherical field of view was necessary then one camera cluster would be seen by the other and stereoscopy would be lost at certain angles. Another alternative is to use a single cluster with a very large number of small cameras pointing in directions that will allow two spherical stereoscopically separated images to be acquired. These cameras will also need to be able to acquire the images at the appropriate resolution and dynamic range suitable for display at the home site.

Aural Displays. Hearing is particularly important for a sense of presence [31]. In the transparent telepresence context the displays would be either binaural headphones or surround sound speakers. The headphones would be most likely to be used in conjunction with a head mounted display for single user applications. Binaural headphones provide the listener with a sense of direction for the sound source that, unlike stereo headphones, is external to the head and can be located on a 360 degree sphere. The surround sound speakers with the immersive screen would be suitable for multiple users.

Aural Sensors. To acquire binaural sound for the single user headphones a dummy head fitted with microphones located where the tympanic membrane of the user’s head would be. Ideally the head could be fitted with artificial pinnae corresponding the shape of the user’s external ear and the head would have the same head related transfer function (HRTF) as the user’s head. This allows accurate recording of interaural time differences (ITDs) and interaural intensity differences (IIDs). Head movements are also important for identifying the direction of sound [32], see [33] for a practical experiment using artificial pinnae on a mechatronic sensor platform.

Haptic Displays. The haptic sense is comprised of touch and kinaesthesia. Touch or tactile information results from cutaneous stimulation that provides the brain with knowledge of external stimuli acting on the skin. Whereas kinaesthetic information provides the brain with knowledge from internal sensors indicating the position of, and forces acting on, the body’s limbs and joints. Thus haptic displays need to provide both tactile information such as the shape, roughness or smoothness, temperature, wetness or dryness, etc. of an object; and kinaesthetic information such as the forces being applied to grasp the object, it’s weight, and the position or velocity of the grasped object etc. There is significant ongoing research work in this field, e.g. [34] provides one example. An early example of a TTRG system can be found in [35].
Haptic Sensors. In order to detect information at the remote site that can inform the haptic displays of the home site complementary sensors are required. For tactile sensing the field of robotics has yielded a variety of touch sensors that provide information on the shape of an object through arrays of pressure sensors on gripper fingertips. Again they do not approach the resolution required for transparency. Similarly a remote slave arm could be instrumented at appropriate locations with strain gauges to act as force and torque sensors. These would provide information to be communicated to the home site which would alter the forces being experienced by the user who would be wearing a master arm with motorized joints which could emulate the forces being experienced by the slave arm at the remote site.

Olfactory Displays. Estimates of the number of identifiable sensations of odour vary between $10^4$ [36] and $10^6$ [37]. A number of factors are important in producing an effective olfactory display, some of these are; the odorant concentration, the duration of exposure, the flow rate, the directionality, and the adaptation time. Very little odorant will be necessary to create an olfactory image [38]. Adaptation occurs with all senses but it is particularly noticeable with olfaction and gustation, it is the decrease in sensitivity that occurs when a sense is exposed to constant stimulus. Some very specific “displays” exist at present, for example the generation of the smell of freshly baked bread in a supermarket, however there are no existing full olfactory telepresence displays and work is still at the research stage on this.

Olfactory Sensing. In order to detect “smells” chemical sensors would need to be used. Arrays of sensors have been used to produce input for neural networks and other systems in an attempt to identify smells [39]. This leads on to an interesting problem for telepresence olfaction. Should the specific odour be identified at the remote site, the identity of the odourant signaled to the home site, and then requisite odour selected for display from an existing library. This would be the simpler solution but would only be applicable for a finite number of odours. Alternatively the exact chemical composition and structure could be identified at the remote site, this information communicated to the home site, where it would then be synthesized to replicate the smell. This would be extremely difficult since the precise nature of how the chemical composition affects the odour is not fully known. At the moment there is no known method of creating an olfactory telepresence system that can operate in real time across a wide range of odours.

Vestibular Displays and Sensors. The vestibular sense is part of the body’s proprioceptive system and various studies have been carried out to determine JNDs and sensitivity, e.g. [40]. These are most familiar to visitors of theme parks where they experience virtual rides on flying vehicles or submarines, and to pilots using flight simulators. These motion platforms, usually of the six axis Stewart Platform type, are essentially vestibular displays since they convey a sense of motion and orientation to our bodies through manipulation of our vestibular system. When combined with visual cues the brain can be tricked into believing it is moving over long distances and accelerating, decelerating, or simply traveling at a constant velocity. These displays are not yet normally used with telepresence systems. Sensors would be located on the vehicle or other sensor platform containing the cameras etc. at the remote site. They would be likely to be gyroscopes and accelerometers to
provide absolute information on velocity, acceleration, deceleration, and orientation of the remote system

2.3 Home Site Controls and Remote Site Actuators

The controls should adhere to all ergonomic principles for ease of use. An interesting possibility is the use of a direct EEG based brain-computer interface (BCI) for control. As well as being investigated to aid physically disabled individuals [41] it has also been used for navigation in virtual environments [42], [43]. The possibility of hybrid systems incorporating BCI for control with more conventional technology for displays is an interesting one. At the remote site the actuators may be motors controlling the movements of an anthropomorphic arm and gripper. As such they will require resolution comparable with home site controls and good repeatability. The technology for these elements is similar to that of industrial robotics, with the exception that for transparency anthropometric manipulators should be used.

2.4 Home and Remote Site Computer Interfaces and Communications Link

At the home site these will handle the signals from the controls and change them into a format for transmission to the remote actuators, and also handle the incoming signals from the remote sensors and translate them into suitable values for use by the displays. At the remote site the data from the sensors will be processed before transmission to the home site, for example image compression algorithms [44] will probably be used to reduce the data required for the video and audio information. The control signals will be received from the home site and translated into appropriate signals for the actuators. For transparent telepresence a broadband connection will be necessary, mostly to cope with the video information. Audio, haptic, olfactory, and vestibular senses will require much lower bandwidth. The processing time for all of this information introduces delays into the system. For transparency these need to be minimised as various writers have presented values for the maximum delay allowable. One of the lower values indicated is 17ms [45] for HMD operation although delays up to 300 ms could be tolerated as a maximum for teleoperation [46].

3 Discussion

The paper has postulated that if appropriate technology can be created, then the mediation between the human and the telepresence system can approach transparency. This would allow full immersion in the remote environment, and the author strongly suggests, although this is open to argument, that this would also allow full “presence” in the remote environment to be experienced.

References