

EC/NSF Position Paper

Toward tightly-coupled human interfaces

Dr. Thomas A. Furness III

Human Interface Technology Laboratory

University of Washington

Seattle, WA 98195 USA

tfurness@u.washington.edu

<http://www.hitl.washington.edu>

As we stand at the portal of the next millennium, I am both excited and terrified about the future. I feel that as a modern civilization we may have become intoxicated by technology, and find ourselves involved in enterprises that push technology and build stuff just because we can do it. At the same time we confronted with a world that is increasing needful of vision and solutions for global problems relating to the environment, food, crime, terrorism and an aging population. In this information technology milieu, I find myself being an advocate for the human and working to make computing and information technology tools that extend our capabilities, unlock our intelligence and link our minds to solve these pervasive problems.

Some assumptions about the future

It was estimated that in 1995 there were 257.2 million computers in the world (96.2 million in US, 18.3 million in Japan, 40.3 million in Europe). Collectively, these computers produced a computing capacity of 8,265,419 million instructions per second (mips). By the year 2000, the number of computers relative to 1995 is expected to more than double, with a combined computing capacity of 246,509,000 mips.[1]. That's about 41,000 instructions per second for every person who lives upon the earth. Ray

Kurzweil [2] predicts that by 2010 we can purchase for \$1000 the equivalent information processing capacity of one mouse brain and by 2030 the equivalent computing capacity of one human brain. Continuing this extrapolation, he predicts that by 2060 digital computing (again purchased for \$1000) will equal the processing capacity of all the human brains on the earth (and Kurzweil has been pretty good at predicting).

These trends suggest that the following assumptions will be (for all intents and purposes) realized in the coming decades:

- Computing capacity will continue to increase at least Moore's law rates (i.e. doubling every 18-24 months) [3].
- Dramatic advances will be made in high resolution digital imaging, compression algorithms and random access mass storage.
- Broadband communications will be available worldwide.
- There will be a rich mix of available wired and wireless communications.
- Reduction in size, cost, and power consumption of computational and communications hardware will continue.
- There will be continued advancement in portable power generation and storage.
- AI heuristics will continue to develop including natural language and learning.
- World's knowledge resource will be digitized and placed in accessible locations.
- Computers will continue to be connected to people.

My colleagues and I also anticipate an emerging trend toward a "power company" model of networked system architecture, in which "thick" local processing (focused largely on interface) communicates with "thick" computing and content services through relatively "thin" network devices and servers. A final (and key) assumption is that although humans may be slow relative to the speed and growth of computation, we have an incredible ability to think out of the box and make 'cognitive leaps' in solving problems. So humans are not obsolete yet.

Within this context we envision a future in which the boundaries of human thought, communication, and creativity are not defined by the design, location, and proximity of

information technology, but by the human endeavor which these devices support. Tightly-coupled human interface technology will produce a symbiotic relationship, supporting and facilitating reflective and experiential thought. Emotional and motivational factors will prove to be as important as cognitive factors in many domains, and natural human behavior will be the predominant mode of interaction. Future media will be personal, flexible, emergent, and universal.

Interface Challenges

While these trends will greatly expand our use of digital media, they will not on their own produce a fundamental shift in the way we conceptualize and interact with media and information technology systems. I feel that the greatest near term challenge of the information age is that of being able to really use the phenomenal capacity that will be achieved in digital media, computing and networking. How will humans tap and process all that can flow...like drinking from a fire hydrant with our mouths too small?!

Herbert A. Simon, the 1978 Nobel Laureate in Economics and the recognized father of artificial intelligence and cognitive science, stated that:

“What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.”

(It should be added parenthetically, that the lack of a good interface also consumes much more resources than an intuitive one.)

Even though we have made great progress in developing computing technology, the concomitant development of the interfaces to those media has been lacking. Television is still two dimensional, telephony is still monophonic and we are still using a highly coded symbol interface (the keyboard) and a small screen to interact with computers. In the last 20 years about the only improvement in the human to computer interface has

been the mouse invented by Douglas Englebart in 1965. The mouse, as a spatial input device, has made a dramatic improvement in working with desktop and windowed screens; but as for the rest, little progress has been made.

This concern about lagging interfaces has been echoed by the United States National Research Council who recently published a report of its steering committee on computer interfaces titled *More than Screen Deep*. [4] There were three main recommendations of the Committee. The first was the need to break away from 1960s technology and paradigms, and develop new approaches for immersing users in computer mediated interactions. The second conclusion was the need to invest in the research required to provide the component subsystems needed for every-citizen¹ interfaces. The third conclusion was to encourage research on systems-level design and development of human machine interfaces that support multiperson, multimachine groups as well as individuals.

Computers generally give us a way to create, store, search and process vast amounts of information rapidly in digital domains and then to communicate this information to other computers and/or to people. To fully exploit the potential power of the computer in unlocking and linking minds, I believe that we have to address computation and humans as a symbiotic system.

To achieve this vision of a radically different model of our relationship to information systems we will need to address the following research challenges:

- (1) What are the most useful and effective methods of integrating the information system interface of the future?
- (2) What are the most appropriate metrics and methods for determining when we're on the right track?
- (3) What innovative component appliances will be possible and how will they be used?

- (4) How will we get bandwidth to the brain and expand human intelligence to make use of the media and information processing appliances of the future?

Some fundamental assertions

In an attempt to answer these questions, I propose the following assertions or principles that we should follow in developing better interface appliances:

1. We must exploit the fundamental 3D perceptual organization of the human in order to get bandwidth into the brain.
2. We must exploit the fundamental 3D organization of our psychomotor mechanisms to get bandwidth out of the brain.
3. We must use multiple sensory and psychomotor modalities increase the effective bandwidth to and from the brain.
4. We must observe the human unobtrusively and infer intent and emotions, so that we can adapt the information channel to tune the flow of information in/out of the human based upon these measures.
5. We must remember that humans build mental models to predict and conserve bandwidth.
6. We must remember the power of place (e.g. people generally remember ‘places’ better than text.)
7. We must put people in “places” in order to put “places” in people.
8. Machines must become more human-like (rather than humans machine-like) in order to advance together.
9. In the future we can expect machines to learn and adapt to humans.
10. We can progress no faster than our tools to measure our progress.

Matching machines to humans

The term *interface* can be described as *what exists between faces*. At the basest level, the

¹ In the context of this report, the term every citizen includes all ages and physical abilities.

role of the human interface is to transfer signals across human and machine boundaries. (One may think of this is where the silicon and carbon meet.) These signals may exist in the form of photons, mechanical vibrations, electromagnetic and/or chemical signals and may represent discrete events such as key presses and status lights, as well as continuous events such as speech, head/eye movement, visual and acoustic imagery, physiological state etc. The physical interface is intended to be a means to an end, and not the end itself, and thus it should be transparent to the user in performing a particular task with the medium. Ideally, the interface provides an 'impedance match' between human sensory input and machine signal output while simultaneously providing efficient transduction of human intent as reflected in psychomotor or physiological behavior of the user. The end goal is to create a high bandwidth signal channel between the human cognitive processes and the machine signal manipulation and delivery processes.

To create an ideal interface or 'impedance match' between the human and the machine, it is first necessary to understand the saliencies of how humans function. Much can be said on this topic. The reader is encouraged to explore the references at the end of this paper for further information. To summarize the author's experience in interface design, human capabilities can be boiled down into the following statements:

#1 Humans are 3D spatial beings. We see, hear and touch in three dimensions. Although providing redundancy, our two eyes and two ears, along with feedback (i.e. proprioceptive cues) from arms, legs etc., allow us to localize ourselves in three dimensional space. Light rays emitted or reflected from the three dimensional world reach the retinae of our eyes and are transduced by a two dimensional receptor field. The then brain uses the signals from both eyes containing vergence, stereographic and accommodative cues to construct three dimensional understanding. From birth we develop these spatial skills by interacting with the world. Similarly, our ears individually receive and process sound. Depending upon the location of the sound, the brain compares the interaural latencies and sound wavefront (having been manipulated by the pinnae of the outer ear) to create a three dimensional interpretation of the sound field reaching our ears. If we use interfaces that do not represent signals naturally or in

3D, we have to build new mental models to operate and interpret the signals from these interfaces.

#2 Humans have two visual systems. Our eyes are amazing. The light sensitive organ of our eye, the retina, is composed of two receptor types: cones and rods. The cone receptors (of which there are about 7,000,000) are sensitive to color and high spatial detail, and are located primarily in the macula or fovea of the eye. This region only subtends a 2-4 degree visual angle. The peripheral retina is populated with about 120,000,000 rod receptors, which are not color sensitive, but have a shorter time constant, are highly sensitive to movement and can operate at lower light levels. Even though certain portions of the peripheral retinal have a greater density of rod receptors than that of the cone receptors in the fovea, these rod receptors are connected together such that they are 'ganged' to integrate light. It is interesting that these two receptor fields are processed at different regions of the brain and thereby perform different functions. The foveal region provides the detailed spatial information to our visual cortex so that we can read. This necessitates that we rotate our eyes often by rapid eye movements called saccades in order to read. The function of this region is to provide what we call our *focal vision*, that tells us the 'what' of things. Simultaneously, the signals from our peripheral retina are processed in the lateral geniculate and other portions of the brain and do not have as dominant a connectivity to the visual cortex. The function of the peripheral retina is to help us maintain a spatial orientation. It is our peripheral vision or *ambient vision* that tells us the 'where' of things. In essence the ambient visual system tells the focal visual system where to fixate.

To build a visual interface that takes advantage of the architecture of the human visual system, the display first must be wide field-of-view (e.g. subtend a large enough visual angle to allow the ambient visual system to work in conjunction with the focal visual system) and second, the information needs to be organized so that the spatial or 'where' content is in the periphery while the 'what' or detail is in the center of vision.

#3 Humans build mental models that create expectations. William James, the 19th century philosopher/psychologist stated that: "...part of what we perceive comes from

the object before us, the other part always comes out of our own head.” This is saying that much of what we perceive in the world is a result of prestored spatial models that we have in our heads. We are mental model builders. Pictures spring into our mind as we use language to communicate. Indeed, our state of learning can be attributed to the fidelity of our mental models in allowing us to understand new perceptions and to synthesize new things. The efficiency with which we build mental models is associated with the intuitiveness of the interfaces and environments we inhabit. Highly coded interfaces (such as a computer keyboard) may require that we expend too much mental energy just to learn how to use the interface (the context) rather than concentrating on the content. Such an interface is not transparent and gets in the way of the task we are really trying to perform.

#4 Humans like parallel information input. People make use of a combination of sensory stimuli to help reduce ambiguity. The sound of a letter dropping in a mailbox tells us a lot about how full the mailbox is. The echoes in a room tell us about the material in the fixtures and floors of a room. We use head movement to improve our directional interpretation of sound. We use touch along with sight to determine the smoothness of a surface. Multiple modalities give us rich combinatorial windows to our environment that we use to define and refine our percept of the environment. It is our way of reducing ambiguity.

#5 Humans work best with 3D motor control. Generally, people perform motor control functions most efficiently when they are natural and intuitive. For example, using the scaled movement of a mouse in a horizontal two dimensional plane to position a cursor on a screen in another vertical two dimensional plane is not naturally intuitive. Learn it we can, and become proficient. Still, this may not be as effective and intuitive as pointing a finger at the screen or better yet, just looking at the item and using eye gaze angle as an input mechanism. Anytime we depart from the natural or intuitive way of manipulating or interacting with the world, we require the user to build new mental models, which creates additional overhead and distracts from the primary task.

#6 Humans are different from each other. People are all different. We have different shapes, sizes, physical and cognitive abilities, even different interests and ways of doing things. Unfortunately, we often build tools and interfaces, expecting everyone to be the same. When we have the flexibility of mapping the way we do things into the tools we use, chances are we will use them more effectively.

#7 Humans don't like to read instructions. This is the laughable position in which we now find ourselves, especially in the age of fast food and instant gratification. It is painful to read instructions, and often they are not paid attention to. The best interfaces are those that are natural and intuitive. When instructions are to be given, it is best to use a tutorial, or better yet, on-screen context sensitive help. Maybe best would be an intelligent agent which watches our progress and mistakes and (politely) makes recommendations.

Table 1 : Status of Current Computer Interfaces

- information is still highly coded
- presentations are not three dimensional (vision & audition)
- display fields-of-view too small (e.g. not immersive and don't take advantage of the peripheral retina)
- the user is outside looking in (do not exploit the perceptual organization of the human)
- inflexible input modalities (e.g. such as using speech and eye gaze)
- presentations are not transparent (cannot overlay images over the world)
- interfaces require the user to be 'computer like'
- interfaces are not intuitive (i.e. takes a while to learn)
- it is difficult to involve multiple participants

Shortfalls in current computer interfaces to humans

If we use these characteristics of the human to examine the current incarnation of

computer interfaces², we find that they fail dismally the impedance matching test and don't take advantage of even the basic criteria of how humans work. I have listed just a few of these in Table 1.

In summary, the current state of computer interfaces is poor. There is no match to the capabilities of the human thereby greatly limiting the linking of digital data streams in and out of the brain.

Virtual interfaces----a better way?

To overcome the human interface difficulties enumerated above, and to restore our lost vision in the information age, it is clear that a paradigm shift is needed in the way we think about coupling human intelligence to computation processes. The end goal of any coupling mechanism should be to provide bandwidth to the brain through matching the organization and portrayal of digital data streams to sensory, perceptual, cognitive and motor configurations of the human. Since it is difficult to change the configuration of human sensory and psychomotor functions (except through training), it is more advantageous to arrive at a computer interface that is designed from the human out. Indeed the goal is to 'make the computer human-like' rather than our past of making the human computer-like.

Virtual reality is a new human interface that attempts to solve many of these interface problems. Much (and perhaps, too much) has been written about virtual reality. It has been called the ultimate computer interface by many. [5] At least in theory, virtual interfaces attempt to couple data streams to the senses ideally, and afford intuitive and natural interaction with these images. But the central key to the concept of virtual reality is that of sensory immersion within a 3D interactive environment.

Definitions of the Virtual

The term *virtual reality* originated from the concept within optical physics of virtual

² For the sake of discussion, I use the typical computer interface as a flat panel or CRT display that provides a 40 cm diagonal face, a alphanumeric keyboard and a conventional mouse.

images. A *virtual image* is an image that appears to originate from a place in space but does not. It is a projection. In the field of virtual reality, virtual images are extrapolated to include visual, auditory and tactile stimuli which are transmitted to the sensory endorgans such that they appear to originate from within the three dimensional space surrounding the user. The term *virtual interface* adds the input and output modalities to interacting with virtual images and is defined as: A system of transducers, signal processors, computer hardware and software that create an interactive medium through which: 1) information is transmitted to the senses in the form of three dimensional virtual images; and 2) psychomotor and physiological behavior of the user is monitored and use to manipulate the virtual images. A *virtual world* and or *virtual environment* is synonymous with virtual reality and is defined as: the representation of a computer model or database in the form of a system of virtual images which creates an interactive 3D environment which can be experienced and /or manipulated by the user.

The components of a typical virtual interface system are shown in Figure 1. Virtual interfaces and the information environments they produce provide new alternatives for communicating between humans and within databases. Instead of viewing directly a physical display screen, the virtual display creates only a small physical image (e.g., nominally one square inch) and projects this image into the eyes by optical lenses and mirrors so that the original image appears to be a large picture suspended in the world. (Note that the virtual retinal display, discussed below, does not use a two dimensional array of photon emitters but instead, projects a single photon stream directly on the retina of the eye.)

A personal virtual display system, termed a head-mounted display, usually consists of a small image source (e.g., a miniature cathode-ray tube or liquid crystal array) that is mounted on some headgear, and small optical elements which magnify, collimate and project this image via a mirror combiner into the eyes such that the original image appears at optical infinity. The perceived size of the image is a function of the magnification of the optics and not the physical size of the original image source. With two image sources and projection optics, one for each eye, a binocular virtual display can be created providing a 3D or stereoscopic scene.

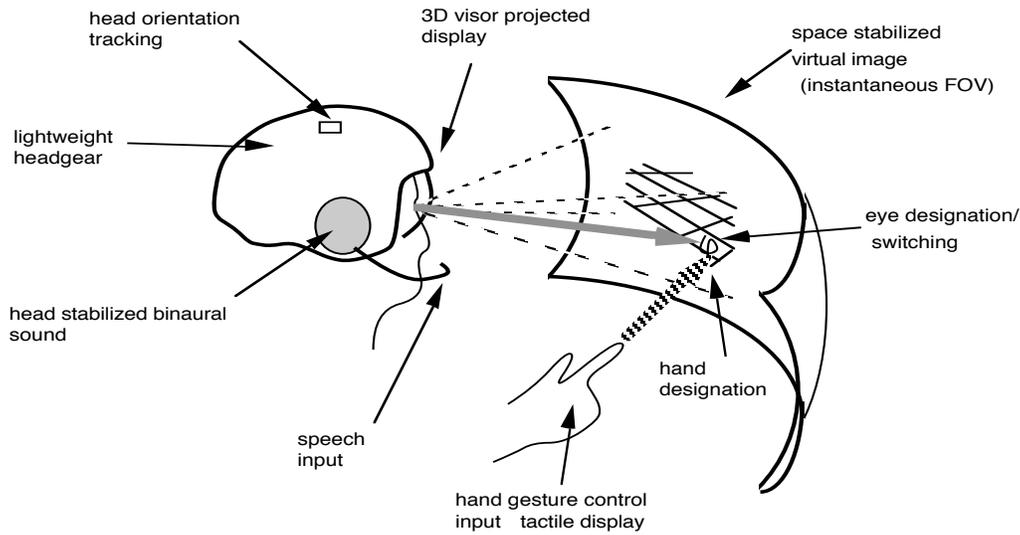


Figure 1. Virtual Environment Display System

With a partially reflective combiner (a mirror that reflects light from the image source into the eyes), the display scene can be superimposed onto the normal physical world. The user can also position the image anywhere (i.e., it moves with the head). When combined with a head position sensing system, the information on the display can be stabilized as a function of head movement, thereby creating the effect of viewing a space-stabilized circumambience or "virtual world" which surrounds the user.

An acoustic virtual display can also be created by processing a sound image in the same way that the pinnae of the ear manipulate a sound wavefront. A sound object is first digitized and then convolved with head related transfer function (HRTF) coefficients which describe the finite impulse response of the ears of a generic head to sounds at particular angles and distances from the head. Monaural digitized sound can thus be transformed to spatially localized binaural sound presented through stereo headphones to the subject. By using the instantaneously measured head position to select from a library of HRTF coefficients, a localized sound which is stable in space can be generated. These sound objects can be used either separately or as an "overlay" of 3D visual objects.

Similarly, a tactile image can be displayed by providing a two-dimensional array of vibration or pressure transducers (termed *tactors*) which are in contact with the skin of the hand or body. Tactors may be actuated as a function of the shape and surface features of a virtual object and the instantaneous position of the head and fingers. Force or inertial feedback can also be provided through a mechanically coupled motor system that senses hand position and provides virtual stops or modulation to hand movement given instantaneous hand position relative to an object.

Potential advantages of virtual interfaces

Since virtual displays can surround users with three dimensional stimuli, under ideal conditions of high image resolution and low update latencies, users feel a sense of "presence", that we are "inhabiting" a new place instead of looking at a picture. This aspect is illustrated by Figure 2 below. Normally, when we look at a display terminal, for example, we see an object (i.e. the terminal) embedded in our three dimensional world through which a separate message is being conveyed. In order to interact effectively within this world, we have to use three cognitive models: a model of our immediate environment, a model of the functionality of the medium (the terminal, in this case) and a model of the message and its heuristics as conveyed through this medium.

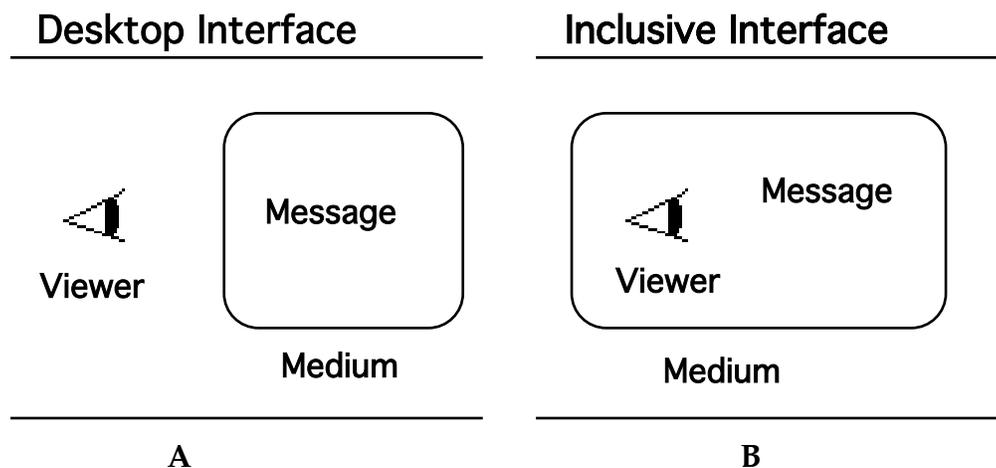


Figure 2. In (A) above, the viewer is separate from the message; in (B) the viewer and message occupy the same virtual space.

Conversely, when we are immersed in an inclusive virtual environment, we in effect become a part of the message. The original environment and presentation medium disappear and we are required to draw upon only a single model of the new environment which represents only the message. Then we can interact within this virtual environment using the same natural semantics that we use when interacting with the physical world. These factors empower the virtual interface as a medium with an unprecedented efficiency in communicating computer-generated graphical and pictorial information, making it ideal for a transportation system for our senses, i.e. an electronically mediated presence. This is the best way to increase substantially bandwidth to the brain.

Other advantages of the virtual environment include its flexibility in conveying simultaneously three dimensional information into several modalities, such as using visual and acoustic representations of an object's location and state in three space. Multiple modality displays have a greater promise of reducing ambiguity in complex displays and perhaps a more effective way of attracting attention to critical conditions during high workload tasks.

It should be noted here that virtual environments can also serve as a good means of simulating other systems, both physical and virtual (e.g., using a virtual display to simulate another virtual display). We will exploit this aspect extensively in our proposed testbed and experimental activities.

My supposition is that virtual environments can provide a three dimensional medium through which complex information can be communicated rapidly to humans and between humans. Virtual reality also provides an empirical context for exploring theories of cooperation between human groups and software configurations. In Table 2 I summarize these advantages of virtual interfaces. As can be seen from the discussion above, most of the attributes of humans can be matched by virtual interfaces thereby creating, at least in theory, an ideal interface.

Table 2: Potential Advantages of Virtual Interfaces

| | |
|------------------------|--|
| • Multisensory display | --hear, see, touch in a 3D circumambience |
| • Multimodal input | --speech, motion (head, hand, body, eyes), facial expressions and gestures |
| • Presence | --visual and acoustic sensory immersion becomes a 3D place |
| • Transparency | --can superimpose and register virtual images over the real world |
| • Intuitive | --behave naturally rather than symbolically --easy to learn |
| High bandwidth | --between the computer and the brain --between brains mediated by the computer |
| Virtual Worlds | --surrogate to real world --abstraction of dataspace |
| Virtual Physics | --can manipulate time, scale and physical laws to create a new physics in the virtual world. |

So if VR is so good, why don't we have more of it?

This is an easy question to answer...the technology simply does not exist yet and what does exist is too expensive. Most virtual world demonstrations to date have been clunky, in that the quality of the worlds have been limited due to the need to render

millions of polygons and texture maps in real time (e.g. 20-30 Hz update rates) while maintaining low latencies in image stabilization (e.g. <16 ms). The headgear and other equipment are often large, heavy and oppressive, and generally get in the way of natural behavior. Furthermore, we may be up against a showstopper, that of cybersickness: This phenomenon, akin to motion sickness, occurs when there is a sensory conflict between the visual and vestibular cues reaching the brain. The potential for cybersickness, along with the need to improve the VR technology, are slowing progress in adopting virtual interfaces.

But perhaps the single greatest hindrance in the widespread development and adoption of virtual interfaces has been economic. As yet, there are no driving applications for VR, especially ones that would stimulate growth of the industry and encourage risk taking among hardware and software developers to make the necessary investment. Most VR installations exist at academic, industrial and government research organizations and have not been practical and affordable in the commercial or consumer marketplace except in limited arcade games. Consumer applications especially have been limited by the cost of the VR components and the lack of computer clout to render in real time the enormous number of polygons and texture maps that make the virtual world a satisfactory experience.

Progress in virtual interfaces

In spite of the hindrances to growth and development of virtual interfaces discussed above, VR is inherently appealing and embodies the potential for the impedance matching between computers and humans described above. Because of this potential, progress is also being made is various fronts. But to generate the ideal virtual environment, much work is needed all modalities, including the transducers and underlying computational hardware and software components. But above all, we need to understand how people really work in virtual environments.

Understanding humans. The Human Interface Technology Laboratory (HITL) at the University of Washington, along with other academic centers at MIT, University of

North Carolina, University of Central Florida, University of Illinois (NCSA), Carnegie Mellon University and other military and industrial concerns, are continuing to work the technology. We have made notable advances in understanding how humans respond to virtual environments. For example, research at the HIT Lab has been aimed at determining the effects of latency and geometric scale on adaptation and readaptation of the vestibulo-ocular reflex and the propensity of users to be affected adversely by these conditions.[6] These efforts have been extended to include methods for overcoming cybersickness and how to measure objectively the sense of presence in a virtual environment.[7] We have obtained overwhelming evidence of the importance of a wide field-of-view in generating a sense of immersion and presence within the computer generated environment. Indeed, immersion substantially increases our ability to form robust mental models that aid navigation in complex virtual environments. We have also found the importance of using the whole body in navigating within virtual environment...kinesthetic and proprioceptive feedback reinforce the visual and provide a better mapping of virtual 3D space in the brain...much as the real world.[8]

Acoustic Displays. Digital signal processing has brought spatialized sound into practical realization. The only limitation is the need to increase the fidelity of individualized head related transfer function so that the fidelity of 3D sound production is the best for each user (otherwise, one would be listening through someone else's ears.) So 3D sound is gaining great maturity, but we need a fast way to determine an individual 'earprint'. [9]

Force Feedback. Bertrand Russell stated: "...it is touch that gives our sense of "reality"Not only our geometry and physics, but our whole conception of what exists outside us, is based upon the sense of touch." [10] Force feedback and tactile representations are perhaps the least mature virtual interface technology, but some progress is being made in force feedback effectors which sense 6 degrees-of-freedom of the hand and provide force modulation based upon the interaction of the position of the force point with a virtual object. [11] To be realistic, a force feedback subsystem needs to update at a 1000 Hz rate. Fast finite element modeling of deformable surfaces such as skin and other tissue will open new areas for surgical simulation involving

virtual reality containing force feedback elements. Several studies have shown that realism in virtual environments can be substantially enhanced by incorporating tactile feedback along with visual representation. [12,13]

Visual Displays. New advances in the resolution of small image sources such as active matrix LCDs and miniature CRTs have improved somewhat the image quality of virtual displays. Typically, these displays provide a SVGA resolution (800X600 elements). The tradeoff is to create a sufficient number of picture elements so that the virtual scene subtends a large visual angle (to generate a sense of presence) without making each pixel appear too large thereby sacrificing the detail that can be seen. Generally the best rule is that more pixels the better, but magnifying an individual pixel to beyond a 3 minutes-of-arc subtended visual angle makes the virtual scene unusable. Ideally, displays would be built that would stimulate the peripheral retina while also coupling detailed color information to the fovea. This would necessitate fields-of-view of 120 degrees horizontal by 60 degrees vertical, with an ideal pixel density of 4000 X 3000 elements. These performance requirements far exceed the 1000 X 1000 elements that will be available in the near future using matrix element image sources.

Virtual Retinal Display. Perhaps one of the most exciting developments to address the issue of display resolution is the development of the virtual retinal display (VRD).[14,15] The VRD is truly a breakthrough technology which scans a modulated photon stream directly on the retina of the eye. The VRD eliminates the need for an image forming source such as a CRT or matrix element device. Figure 3 illustrates the mechanism of the VRD. Photons are generated by low power gas lasers, lasing diodes or light emitting diodes with red, green and blue spectra. These photon sources are modulated with a video signal from the computer or other video source. The three modulated color beams are then combined into one beam and focused onto a monofilament fiber which relays the light to the scan head. The scan head contains two mirrors. A resonant mirror oscillates sinusoidally at a frequency of 15-20 KHz and a galvo mirror oscillates in a sawtooth pattern at 60 HZ rate. The light enters the first mirror and is deflected in the high speed horizontal direction and then bounces to the second mirror where it is scanned in the vertical direction to create a raster. The light then bounced off

the first mirror again to double the scan angle. The rasterized photon stream is then reflected through spherical optics and a combiner such that it enters the eye through a very narrow crossover point which acts like an exit pupil. When the user positions the entrance pupil of their eye into the exit pupil of the VRD the retina is flooded with the scanned light. The beam forms an instantaneous spot of approximately 30μ on the retina at a power level of about 300 nw. Figure 4 illustrates a color optical bench version of the VRD with see-through optics, along with a more portable monochrome unit.

The VRD is remarkable in that it can create a brilliant, high resolution and wide field-of-view image with very few components. The performance of the display is mainly a function of how fast the photon stream can be modulated. Table 3 gives the projected performance parameters for the current VRD development. Multiple beams can substantially increase the overall resolution and field-of-view. The VRD was invented by the HITL and is licensed to Microvision Inc., a Seattle, Washington company who is developing commercial version of the VRD.

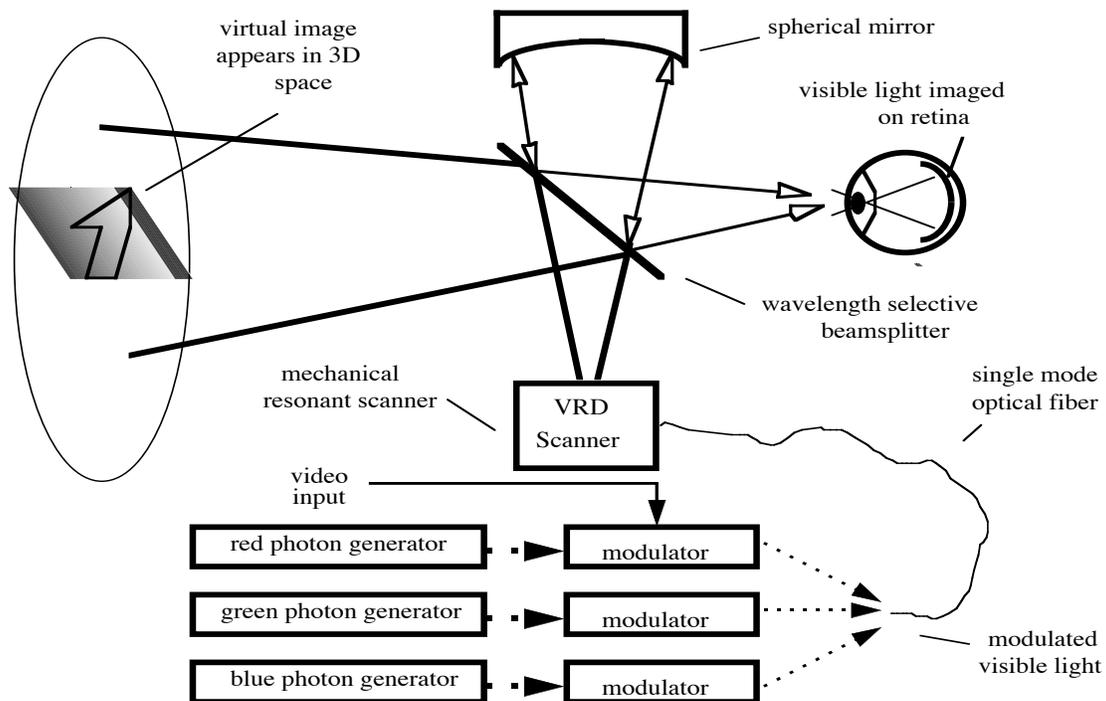


Figure 3 : Virtual retinal display optical schematic



Figure 4 : Left: optical bench configuration of the virtual retinal display
Right: Portable monochrome VRD demonstrator

Table 3: Performance predictions for the virtual retinal display

- | | |
|-------------------------|-----------------------|
| • Resolution (goal) | 2000 X 1500 elements |
| • Luminance | >>10,000 ft. lamberts |
| • Dynamic Range | >3000:1 |
| • Field-of-view | 60 degrees |
| • Accomodation (static) | -2.0 to +2.0 diopters |
| • Color | red, green, blue |
| • Binocular | |
| • Weight | ~100 gms |

VR applications. Applications that are promising now include those settings where visualization of spatial information is important such as automotive and architectural design and medical visualization....especially when it helps to be inside a computer generated model. The Human Interface Technology Laboratory has been exploring the use of virtual environments for teaching complex scientific and mathematical concepts to children. The results have been profound. We have found that students who typically have been 'slow learners' tend to catch up with the 'smart' kids. [16] We have also used VR for the treatment of Parkinson's disease [17], phobias (e.g. fear of spiders) [18] and for pain alleviation. [19] We have built and demonstrated distributed virtual worlds, such as the GreenSpace project wherein we have placed two people in Japan and two people in Seattle into a shared 3D immersive space, wherein participants could see, hear and interact with each other in three dimensions. [20] We have built virtual architectural worlds such as Virtual Venice where we mapped part of the actual city and built a virtual representations which could be use to visualize the prospects of new construction and rehabilitation of old structure.

Other laboratories (e.g. University of North Carolina, NCSA at the University of Illinois, Ford Motor Company) have been using virtual reality for medical imaging, molecular modeling and automotive design. Disney Imagineering has created a virtual world simulating the movie *Alladin* where participants can fly on a virtual magic carpet and visit a 3D world of fantasy. Applications for virtual worlds truly defy the imagination, going from building real world like models to abstractions involving non spatial data sets. [21]

Grand Challenges to advanced virtual interfaces

The sensors and actuator technology community have a great opportunity to contribute to the future of virtual interfaces. Below I have listed some of the areas where innovative solutions are needed to further the utility and performance of virtual interfaces.

Tactile Displays. Even though the skin is the largest organ in the body (e.g.

approximately 2 m², creating a sense of touch within virtual environments is a real challenge. Early work at Wright-Patterson AFB in the Super Cockpit program used a system of pneumatic bladders in a glove to apply pressure to the fingertips, based upon the instantaneous position of the hand relative to a virtual object. Some of the best pioneering work in the field was performed by Margaret Minsky and her colleagues at MIT. Prof. Grigore Burdea at Rutgers has also been pioneering methods of providing touch and force feedback.[22] SensAble Technologies Inc. [23] is currently marketing force feedback devices, which measure 6 degrees of freedom and provide force feedback to a stylus. But providing surface simulation to the receptors of the fingers and other parts of the body is difficult. This is an area where microactuators coupled to the skin could make realistic representations of the roughness of surfaces as a function of hand or finger position. Tactors may also be placed on other parts of the body to simulate surfaces.

Smell/Taste. Perhaps the least developed area in virtual interfaces is the sense of smell and taste. Certainly chemical sensors have been around for a long time, but the real time synthesis of fragrances and/or the delivery mechanism to the human is underdeveloped. Morton Heilig made an attempt to build in odor delivery in his invention of the Sensorama in 1963. Using solenoids, wicks and vials of fragrances, the user would smell appropriate scents while sitting on a motion platform, viewing a wide field-of-view 3D movie, hearing 3D sound, and feeling the wind in their face.

The sensors community should continue to build robust detectors and parse the fundamental elements of taste and smell, but also strive to build those delivery mechanisms which can convey remotely the same chemicals or mimetics to the sensory endorgans. It is likely that delivery side of taste and smell will continue to be difficult and may necessitate in the short term a cross modal representation. The idea here is to use another sensory modality convey smell or taste, such as 3D pattern of light and color to represent a particular combination of tastes and 3D sound patterns to represent smell. Since humans are good pattern recognizers, cross modal patterns might be recognizable after some training. Of course this training should include the real smell and/or taste which correlates with the sound or visual display of such. But there may

be times when it is not desirable to smell or taste the real world as in the case of hazardous environments or being the king's poison control taster. Still, virtual smell and taste without the stimulation of the olfactory and taste receptors may not be too satisfying.

Photon generation & manipulation. Although new display technologies such as the virtual retinal display provide a better coupling of visual images to the retina, there are still many areas for improvement. Even the VRD only covers fraction (about one sixth) of the total retinal receptor field. New microelectromechanical scanning engines are needed which can steer photon streams at rates up to 30-50 KHz. We also need small solid state devices that can deliver blue and green light and be modulated at bandwidths exceeding 100Mhz.

New technology is also needed to manipulate the light wavefront in order to modulate the apparent distance of each display picture element. Currently, all virtual displays present a field of pixels that appear at the same viewing distance (usually optical infinity). This is like sitting inside a sphere upon which the image projection is being made. But the object of an binocular virtual display is to generate vergence cues (or stereographic cues) between the eyes so that objects will appear at different distances and in three dimensions. This situation creates a conflict between the two sets of cues...the monocular accomodative cues of pixel distance which is based upon the collimation of the light versus the binocular vergence cues (between the eyes). A condition may exist wherein the user is viewing an object that has vergence cues of a meter but the pixels appear to be originating from optical infinity (i.e. more that 10 meters away). In order to create a true voxel (volume picture element), the accomodative and vergence cues should match. This necessitates the modulation of the light wavefront at video rates to create an appropriate monocular viewing distance that corresponds with the binocular viewing distance. The problem now becomes, how can we cause a wavefront of light to change its collimation at 30 to 100 MHz rates? Depending upon the optical design, it may only be necessary to defect a surface a few wavelengths of light to obtain the desired effect.

Summary

My purpose in writing this position paper is to emphasize the need for better interfaces with computers and to describe potential improvements that virtual interfaces can offer. Virtual interface technology and virtual environments if developed systematically, may allow us to meld the inductive and rational thinking of the human to the data manipulation power of the computer. Effectively configured, these tools of our age can transport our minds to new places where our vision will yet be extended.

Acknowledgements

The author would like to acknowledge the contributions of Suzanne Weghorst, Joseph King, Maxwell Wells and other staff and students at the Human Interface Technology Laboratory for insights and ideas that are related in this position paper. Research reported herein has been supported in part by the National Science Foundation, Defense Advanced Research Projects Agency, Microvision Inc., and the Virtual Worlds Consortium.

For further information contact: tfurness@u.washington.edu or visit the Human Interface Technology Laboratory website: <http://hitl.washington.edu>.

References

- [1] *Annual Computer Industry Almanac, 1996*. Computer Industry Almanac, Inc. c.1992-
- [2] Kurzweil, R. *The Age of Spiritual Machines*. Viking 1999.
- [3] <http://www.intel.com/pressroom/archive/speeches/GEM93097.HTM>
<http://www.intel.com/intel/museum/25anniv/hof/moore.htm>
- [4] National Research Council. *More Than Screen Deep: toward every-citizen interfaces to the nation's information infrastructure*. National Academy Press, 1997.
- [5] Sutherland, Ivan 1965. *The Ultimate Display*.
- [6] Draper, M. *Exploring the Influence of Virtual Bodies on Spatial Awareness*. Master's Thesis. University of Washington, College of Engineering. 1995.
<http://www.hitl.washington.edu/publications/draper>

- [7] Prothero, J. *The Role of Rest Frames in Vection, Presence, Motion and Sickness*. Dissertation. University of Washington, College of Engineering. 1998. <http://www.hitl.washington.edu/publications/r-98-11>
- [8] Peterson, B 1998. *The Influence of Whole-Body Interaction on Wayfinding in Virtual Reality*. Masters Thesis. University of Washington. <http://www.hitl.washington.edu/publications/r-98-3>
- [8] Peterson, B; Wells, M; Furness, T; Hunt, E (submitted). *The Effects of the Interface on Navigation in Virtual Environments*. In proceedings of Human Factors and Ergonomics Society 1998 Annual Meeting: Human Factors and Ergonomics Society. <http://www.hitl.washington.edu/publications/r-98-5>
- [9] Baldis, J 1998. *Effects of spatial audio on communication during desktop conferencing*. Masters Thesis. University of Washington.
- [10] Russell, Bertrand *The ABC of Relativity*. George Allen & Unwin. 1969.
- [11] <http://www.sensable.com/haptics.htm>
- [12] Hoffman H, Groen J, Rousseau S, Hollander A, Winn W, Wells, M and Furness T. 1996 Tactile augmentation: enhancing presence in virtual reality with tactile feedback from real objects. Paper presented at the meeting of the American Psychological Society, 1996, San Francisco, CA.
- [13] Ikei Y, Wakamatsu K and Fukuda S 1997. *IEEE Computer Graphics and Applications* 17:6.
- [14] Tidwell, M 1995. *A Virtual Retinal Display for Augmenting Ambient Visual Environments*. Masters Thesis. University of Washington. <http://www.hitl.washington.edu/publications/tidwell>
- [15] Wright, Giles 1998 *New Scientist Magazine*. 2146, 33-35.
- [16] http://www.hitl.washington.edu/projects/learning_center/vrrv/
- [17] Prothero, J 1993. *The Treatment of Akinesia Using Virtual Images*. Masters Thesis. University of Washington, College of Engineering. <http://www.hitl.washington.edu/publications/prothero>
- [18] Carlin, A S ; Hoffman, H ; Weghorst, S (1997). *Virtual Reality and Tactile Augmentation in the Treatment of Spider Phobia: A Case Study Behavior Research and Therapy*, 35(2), 153-158. see also: <http://www.hitl.washington.edu/research/therapeutic/burn.html>
- [19] Hoffman, H G ; Doctor, J N ; Patterson, D R ; Weghorst, S ; Furness, T (1998). *VR for Burn Pain Control During Wound Care*. In proceedings of Medicine Meets VR 6, San Diego, CA: IOS Press. see also: <http://www.hitl.washington.edu/research/therapeutic/exposure.html>
- [20] <http://www.hitl.washington.edu/research/greenspace/greenspace.html>
- [21] <http://www.hitl.washington.edu/>
- [22] Burdea, Grigore, *Force and touch feedback for virtual reality*. July 1996. Wiley

[23] <http://www.sensable.com>

Further readings

Barfield W and Furness T 1995 *Virtual Environments and Advanced Interface Design*. Oxford University Press.

http://www.hitl.washington.edu/projects/knowledge_base/guides.html

<http://www.media.mit.edu:80/~marg/haptics-bibliography.html>