# Enhancing Human Computer Interaction in Networked Hapto-Acoustic Virtual Reality Environments on the CeNTIE Network

Tony Adriaansen[1], Alex Krumm-Heller[1], and Chris Gunn[2]

[1] CSIRO, ICT Centre,
P O Box 76, Epping, 1710, NSW, Australia.
http://www.ict.csiro.au
[2] CSIRO, ICT Centre
Building 108 North Road, ANU campus,
Acton, 2601, ACT, Australia

**Abstract.** We discuss some recent advances made to our Virtual Reality system which includes touch feedback in a networked environment. Some issues which affect the realisation and implementation of the Hapto-Acoustic Virtual Reality Environment and human–computer interaction (HCI) are mentioned, as well as networking issues. Our system is called a "haptic workbench" which combines 3D graphics visualisation, touch (haptics) and audio interaction with a computer generated model. The aim is to produce a semi-immersive 3D environment, experienced in real time by the user. Networking two or more systems makes it possible to collaborate in the virtual space and achieve a sense of perceived virtual presence. Recent additions to the system in the form of real time video with audio is demonstrated and shown to enhance the level of virtual presence in the networked environment. Such face to face communication with the remote user in addition to haptic and audio collaboration with the virtual model adds to the sense of presence between users. Two methods of sending video over the internet are compared and our example implementation is described. These additions make it possible to add value in learning scenarios and opens up the possibility of increasing participation during a training session.

## 1 Introduction

Our virtual environment is known as a "haptic workbench"[1] which combines a Virtual Reality (VR) system with the sense of touch. Other typical VR systems enable user interaction by tracking and monitoring a user's head or body position[2], altering the displayed output based on the updated orientation and position. Such systems may use a heads-up display worn by the user, be displayed on a monitor or projected onto a screen for applications where a wide angle view is necessary like flight simulator or similar. The haptic workbench uses a semi-immersive virtual environment where the interactive space is the volume occupied by the hands and parts of the arms when seated, unlike other systems. This space is behind a mirror and contains a PHANToM haptic feedback device. Such an arrangement obviates the need for user position tracking, but navigating around the virtual scene requires an extra input. We use a 3D

M. Bubak et al. (Eds.): ICCS 2004, LNCS 3038, pp. 905–912, 2004.

space-ball mouse that can be directly controlled by the user. This means the user is stationary and the scene is rotated in space for alternate viewpoints. The user therefore has direct control of both viewpoint position and zoom, and is not constrained by machine dependent measurements from a tracking device.

Human interaction with the virtual world requires adequate tactile, visual and audio feedback so that the user's perception of the world is as natural and intuitive as possible. Ideally the user should be focussed on manipulating the virtual object and performing the interactive task rather than being concerned with the intricacies or limitations of the machine. Since there is some interaction in human perception between the senses, the response of each of these should not be considered in isolation. Rather, factors relating to a unified view leading to perception must be taken into account when designing a VR system, to ensure maximum useability.

Networking haptic workbenches provides a shared environment where two or more users can interact with a single model. The Australian CeNTIE (Centre for Networking Technologies in the Information Economy) network* is a high speed research network which allows user-perceived real-time communication between systems possible and makes remote collaboration a reality. CeNTIE is made up of a consortium of Australian companies and research institutions, dedicated to research utilising high bandwidth networks. The aim of our research is to enable a sense of presence with the remote user during collaborative interaction. Adding real-time video as well as audio to the system enhances the sense of presence between users. In effect the users may collaborate on two levels. Firstly, together they may interact with and manipulate the virtual model. Secondly, and concurrent with model interaction, they have direct face to face communication available. Factors affecting real time video and audio on the network are discussed and our implementation is described.

## 2 System Overview

The haptic workbench system enables human interaction using the senses of sight, sound and touch. 3D stereo vision is made possible with the aid of LCD shutter glasses and by using a graphics card where the left and right image fields are alternately displayed. The graphics card generates a left/right frame signal which allows synchronisation with the 3D shutter glasses. The virtual 3D object appears projected through the mirror. By co-locating a force feedback PHANToM device behind the mirror the user perceives a 3D virtual object without occlusion which can be touched as shown in Fig 1.

The haptic tool is a PHANToM premium 1.5 with an effective workspace size of 19.5cm x 27cm x 37.5cm. This has 6 degrees of freedom in position sensing (x, y, z, roll, pitch, yaw) and 3 degrees of freedom in force feedback (x, y, z). It is capable of supplying a continuous force of 1.4N, a maximum force of 8.5N, updates at a frequency of 1KHz and has a position accuracy of 0.03mm. The system was originally developed on a Silicon Graphics computer system and has recently been ported to a PC with dual 2.4Ghz processors, 512KB level 2 cache, 4GB RAM, a dual head graphics card and monitors able to support 120Hz refresh rate for stereo video.

* Details on CeNTIE can be found at http://www.centie.net.au/

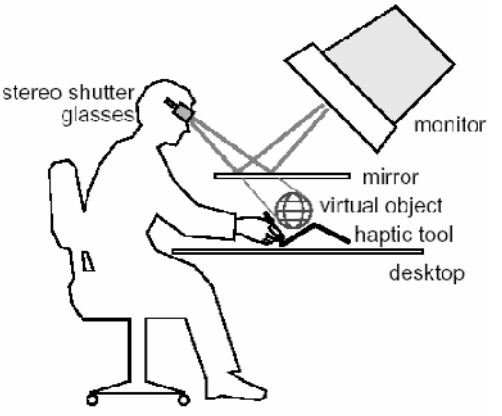


**Fig. 1.** Configuration of the Haptic Workbench

## 3 Human Computer Interaction

Human sensory response and computer/machine performance dictate the users perception and interaction with the virtual model. Not surprisingly, most VR systems focus on visual cues as the sense of sight generally tends to dominate relative to sound or touch when all three are present, even though some studies have shown auditory dominance[3] over visual. However, some studies have shown visual processing to be instrumental in tactile perception [4].

In terms of spatial resolution, a typical monitor is capable of displaying about 2,000 pixels in either dimension, or 5pixels/mm for a 53cm display, enough for relatively smooth visual perception of graphics. In contrast, the spatial resolution capability for control of fingertip position is about 2mm[5] or about 20 times less sensitive, while fingertip displacement touch detection resolution is much more sensitive at about 1 micron.

For video dynamics, the relatively slow video frame update rate of 100-120Hz for 3D stereo has been found to be perceptually flicker free when used with liquid crystal shutter glasses. This together with scene refresh rates of about 30Hz[6] for the movement of virtual objects in the 3D space gives the illusion of smoothly varying object motion. Conversely, the dynamics of human touch perception require much faster update rates and frequencies up to 1KHz can be sensed[7]. However, combining the finger touch perception response time with (arm-hand-finger) position-feedback control means that response rates of between 1-10Hz are adequate for touch dynamics[8].

The synchronisation of audio with both video and haptics is important, delays of generally less than 200ms of audio with respect to video is required for concurrent perception. Good audio quality relies on ensuring no appreciable gaps in transmission. Poor audio/video synchronisation in some television news broadcasts and audio communication drop-out gaps when using mobile phones are common everyday examples highlighting the importance of synchronisation and missing data.

The points outlined above show that improving human interaction and perception with the virtual world relies on matching human characteristics and expectations to

machine capabilities. As the available computer and hardware resources are limited, some shortcuts are necessary in order to develop a system which responds in perceived real time. Briefly, the haptic workbench splits the model into two parts, one for haptic interaction and the other for video, before re-combination. The visual model has high spatial resolution and relatively low update rate while the haptic model has low spatial resolution and a high update rate. In this way the user perceives a continuous real-time experience in all three modes of interaction with the virtual model and when synchronised gives the illusion of seeing, touching and hearing interaction with a single entity.

## 4 Networking

Networking systems makes a collaborative environment possible where both users can interact with the same model in perceived real time.

To successfully realise a networked virtual environment the issue of communication delay (latency) needs to be addressed. For haptics, work by Matsumoto et al.[9] shows that a latency of greater than 60ms prevents the effective use of haptics across a network. For audio communication, normal telephony has a goal of less than 250ms latency, a greater delay causes people to speak at the same time and “infringe on each other’s words”[10]. An adequate network then must have sufficient bandwidth, low latency and quality of service capabilities in order to provide a workable system. The CeNTIE network is a high speed layer 2 switched research network, which uses no routing and therefore reduces latency compared to routing networks. The core network is a dedicated fibre-optic research network which now stretches across Australia and links Sydney, Brisbane, Canberra, Melbourne, Adelaide and Perth, and environs, stretching about 4,000 km east to west and covering some 1,500 km on the east coast. CeNTIE is currently configured for carrying traffic at speeds of 1GBit/s and up to 10GBit/s bandwidth has been demonstated. The CeNTIE network makes collaborative communication possible and fulfils bandwidth and latency requirements. The system uses both TCP/IP and UDP protocols to transfer model data to fulfill real-time perception.

For adequate face to face user communication in addition to the real time model collaboration, and for the same reasons outlined above, there should ideally be adequate resolution, no missing video frames and synchronisation between video and audio streams.

## 5 Adding Real Time Video and Audio

Increasing the perception of being present between users working collaboratively will make user interaction more useful and appealing. This can be achieved by the addition of real time video and audio streams to the system.

Enabling both video and voice communication with the remote user while simultaneously interacting with the virtual model creates a more engaging and natural communication environment compared to not seeing the remote person. This enhances the experience of both users and is of real benefit in aiding the

understanding between users during collaborative tasks. In the same way that moving from purely written communication to speech enables a user to convey extra meaning by intonation and emphasis rather than just content, adding video enables better communication via means such as facial expression. Real time video also enables context dependent communication and provides an understanding of what is happening around the remote person. Creating the enhanced environment also makes it possible to involve others by having additional displays, while incurring no extra computational overhead.

## 6 Comparison of Two Methods

We examined two methods of implementing video and audio conferencing; Digital Video over IP (DV) and MPEG2. DV is a recent video format that offers broadcast quality audio and video, created by a consortium of 60 manufacturers and is gaining in popularity due to agreements with the major players.

DV employs a lossy compression algorithm, based on Discrete Cosine Transform and Huffman coding and provides 5:1 compression of the video image. It has a lower compression ratio compared with MPEG2 and uses a purely intra-frame compression technique. As MPEG2 does intra-frame compression as well as inter-frame compression, latency is higher than for DV. For our application, some data loss in compression/de-compression is tolerable. Table 1 compares the two de/compression algorithms and although DV does not have as high compression values, we chose DV due to lower latency and cost.

**Table 1.** Comparison of DV and MPEG2

| | DV | MPEG2 |
|---|---|---|
| Compression | 5:1 | between 8:1 and 30:1 |
| Compression Type | Intra frame | Both Intra and Inter frame |
| Bitrate | Fixed:~30Mbit/s | Scalable: 2-15 Mbit/s |
| HW Encoder Cost | ~$600 | ~$2000-$5,000 |

By using software that divides each DV video frame into discrete parts we can put this data into packets and send over an IP network such as CeNTIE.

## 7 Implementation of DV

DV data is sent over the Internet using Real Time Protocol (RTP), a protocol for real time applications. Our implementation is based on work by Ogawa et al.[11]. The sending sequence is: Camera receives the DV stream, encapsulates DV packet and transmits to the IP network. The receiving sequence is: data received from network, DV frame reconstructed, attach header, decode and display. With the overhead of encapsulation into IP packets, the data rate is slightly more than 28 Mbit/second. A 100baseT (CAT-5) network cable is able to carry four streams, two in each direction. DV over IP can be considered as an elementary stream implemented as a library that

can be added to any application. We can manipulate this stream to offer the services required for each application. For example altering the frame-rate allows us to alter bandwidth but image quality is retained.

## 8 Ergonomics

Having a stationary user seated at the haptic workbench allows us to add the additional monitoring hardware like a fixed camera and enables positioning of a display not constrained by the orientation or position of the user. In other VR systems where the user's head moves around this would not be possible without elaborate camera motion control.

In order to maintain a realistic viewpoint of each user, the camera should ideally be positioned at the centre of the display so it appears users are facing each other. It is intuitive and natural to look directly at the person on the monitor rather than at an off-axis camera. However, placing a camera at the front center of the display would partly obscure the display. Space constraints also influence the type and positioning of the extra hardware. A compromise using a very small camera mounted close to a small LCD display minimises the off-axis viewpoint position while allowing the user to work without altering head position.

Audio echo cancellation also turns out to be important to avoid audio feedback. We use a special echo canceller which compares outgoing audio with incoming and performs automatic subtraction which almost entirely eliminates round trip echo.

## 9 Results

**Combined Network Traffic Load**

Preliminary testing of the system shows network traffic loads of just over 30Mb/s for combined haptic workbench interaction together with real time video and audio, with no user perceived delay. Similar "current" and "maximal" values for incoming and outgoing traffic are shown in green and blue in figure 2 below.

The system on the CeNTIE network is able to sustain real time activity continuously. The above 3 hour period shows typical data throughput with no loss of continuity, no gaps in data transfer or loss of perceived real time interaction. The resulting implementation of the enhanced workbench is shown in figure 3.

The camera is centrally mounted and aimed directly at the seated user so that the field of view captures a head and shoulders image while the monitor size and position was chosen so that each user's viewpoint is as close as possible to the on-axis position within the confines of the space available. This arrangement eliminates the need for turning away from the virtual world while performing collaborative interaction and simultaneously allows face to face communication. A surgical model can be seen reflected in the mirror as well as the remote user video image with the haptic device partly obscured below.

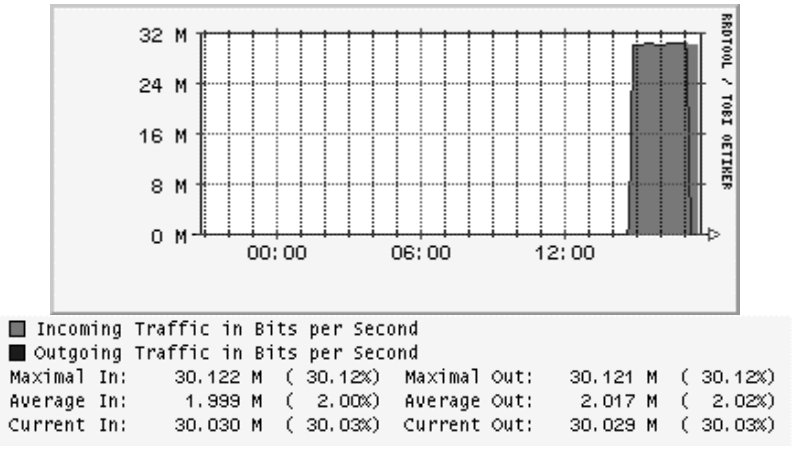


**Fig. 2.** Combined network traffic load (ignore average values)

**Fig. 3.** Enhanced haptic workbench showing small CCD camera (top center), LCD monitor (top right of center) and microphone (right foreground).

Preliminary results on two students and one experienced user confirm increased useability and user performance on a task initiated by the remote user. While it is still too early to generalise, all three showed an improvement in understanding a task when the real time video and audio was present. Objective measurements are planned for the near future and initial comparisons will be made on task completion times. In terms of synchronisation, the real-time perception parameters mentioned in section 3 have been achieved, resulting in a perceptually smooth interaction of haptics, video and audio.

## 10 Conclusion

The system has been operational for only a few weeks, however preliminary tests show that both real time collaborative interaction with a virtual world as well as concurrent face to face communication between users is possible. We also confirm that the implementation of the extra hardware does not detract from the collaborative VR task but in fact enhances the experience. The extra hardware is sufficiently well laid out for the users to feel comfortable in face to face communication and results in substantial improvements in interaction, while not detracting from concurrent collaborative model interaction. The aim of enhancing the networked haptic workbench has been demonstrated.

## References

1. Stevenson,D.R., Smith,K.A., McLaughlin,J.P., Gunn,C.J., Veldkamp,J,P., and Dixon,M.J. “Haptic Workbench: A multi-sensory virtual environment”. In Stereoscopic Displays and Virtual Reality Systems VI, proceedings of SPIE Vol. 3639, 1999. Pages 356–366.
2. LaViola,J.J.Jr., Feliz,D.A., Keefe,D.F., Zeleznik,R.C. “Hands-Free Multi-Scale Navigation in Virtual Environments”. Proceedings of 2001 ACM Symposium on Interactive 3D Graphics, pp 9-16, March 2001.
3. Repp,B.H. and Penel,A. “Auditory Dominance in Temporal Processing: New Evidence From Synchronization With Simultaneous Visual and Auditory Sequences”, Journal of Experimental Psychology-Human Perception and Performance, 2002, Vol. 28, No. 5, 1085–1099
4. Zangaladze,A., Epstein,C.M., Grafton,S.T., Sathian,K., Nature, Volume 401, 7 October 1999, pp587-590.
5. Durlach,N.I., Delhorne,L.A., Wong,A., Ko,W.Y., Rabinowitz,W.M. and Hollerbach,J. Manual discrimination and identification of length by the finger span method. Perception and Psychophysics, 1989, 46(1), 29-38.
6. Chen,W., Towles,H., Nyland,L., Welch,G. and Fuchs,H. "Toward a Compelling Sensation of Telepresence: Demonstrating a portal to a distant (static) office" in Proceedings of IEEE Visualization 2000, October 2000.
7. Bolanowski,S.J., Gescheider,G.A., Verrillo,R.T., Checkosky,C.M. “Four channels mediate the mechanical aspects of touch”, J. Acoustic Society of America, 84 (5), November 1988, pp. 1680-1694.
8. Brooks,T.L. Telerobotic response requirements. Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics. Los Angeles, California, 1990, pp 113-120.
9. Matsumoto,S., Fukuda,I., Morino,H., Hikichi,K., Sezaki,K. and Yasuda,Y. “The Influences of network Issues on Haptic Collaboration in Shared Virtual Environments”. Fifth Phantom Users’ Group Workshop, 2000.
10. DeFanti,T.A., Brown,M.D. and Stevens,R. “Virtual Reality Over High-Speed Networks”. IEEE Computer Graphics & Applications, Vol. 16, No. 4, July 1996, pp. 42-43.
11. Ogawa,A., Kobayashi,K., Sugiura,K., Nakamura,O., Murai.J. “Design and Implementation of DV based video over RTP”, Packet Video Workshop 2000.