The Design of Perceptual Representations for Practical Networked Multimodal Virtual Training Environments

Matthew Hutchins, Matt Adcock, Duncan Stevenson, Chris Gunn and Alexander Krumpholz

CSIRO ICT Centre
GPO Box 664 Canberra ACT 2601, Australia
Matthew.Hutchins@csiro.au

Abstract

This paper is a discussion of our experience in designing perceptual representations for virtual environments for surgical training. Networked virtual environments can be used for training and education, and indeed can have some benefits over hands-on training in the real world. In designing content for such an environment there is a tension between including realistic models of real world objects and choosing other types of representations. We present AFRADERVITE, a framework for describing representations in virtual training environments. The framework identifies which representations are needed, and classifies them according to directness. The paper presents examples of representations taken from surgical training systems for cholecystectomy and temporal bone drilling.

1 Introduction

1.1 Networked multimodal virtual environments for communication

The term "virtual reality" has many definitions, and for the purposes of this paper we will take the point of view that virtual reality (VR) is essentially a communication tool (Riva 1999; Steuer 1993). In particular, we are concerned with the design of networked virtual environments that act as enhanced communication channels between two (or more) participants. That is, the virtual environment is a conceptual place where people go to communicate. They communicate "in" and through the virtual environment. We assume that the virtual environment is populated with computer-generated objects and events that serve to enrich the communication experience of the participants. The objects and events in the environment, and the participants themselves, are represented in such a way as to be perceived by the participants using one or more direct senses, typically vision, hearing and touch, and higher order constructions such as read language and interpreted speech. Communication and action within the environment could use any of several modes, including selection, manipulation, gesture, expression, speech, non-speech sound, written (or typed) language, and drawing. We will not be too concerned with the boundary between real and virtual in the environment, and later discuss the inclusion of real or actual objects, events and experiences within a virtual environment. Milgram and Kishino (1994) have introduced the notion of the virtual reality continuum, ranging from real environments, through augmented reality and augmented virtuality (virtual environments augmented with some real objects), to completely virtual environments. Our discussion will assume environments somewhere along this continuum, but with at least some virtual content.

1.2 Virtual training environments

Practical virtual environments are, typically, designed spaces that have a specific purpose. One of the most compelling purposes for this technology is education and training (Bricken 1991). If we compare training in virtual reality with hands-on training in the real world, we can identify several advantages to training in VR.

- **Safety**. In training situations where mistakes could lead to danger to the trainee or instructor, consumers, bystanders, equipment or the environment, mistakes can not easily be tolerated. However, making mistakes and learning from them is an important part of training. VR training provides the trainee with the freedom to fail safely.
- **Economy and accessibility**. VR training allows repetition without the consumption of physical resources. Also, in situations where training requires access to rare or valuable facilities, VR training could lead to less downtime and higher accessibility for students (assuming the VR environment itself is not a rare and inaccessible resource).

- **Controlled exposure**. Students training "on the job" will often have to learn based on whatever cases occur during a particular training period. The instructor may have little control over the order, frequency, variety or complexity of cases that the student is exposed to. Rare and abnormal cases are, by definition, hard to find in the real world. A VR training environment can give the instructor more control, so at the end of the course, the student has been exposed to a precisely determined set of circumstances. Moreover, VR allows a complex procedure to be broken down into its component parts, so that the amount of training for each component can be varied. In some situations, a VR training environment can be developed for a specific case, so that a complex instance of a procedure can be rehearsed virtually before it is attempted.
- **Measurement**. Virtual reality environments are usually required to track the actions of the participants. This basic tracking can be used as measurements of student performance, and these measurements can be used to build a record of performance over time. This can give an instructor more control over the structure of the course, and could form the basis of objective certification of students.
- **Capabilities beyond real**. Because it is not real, a VR training environment can provide capabilities that go beyond those that exist in the real world. These include simplification, calling attention to only what is significant, providing arbitrary and controllable viewpoints, visualising the normally invisible, annotation of models and events, and the ability to undo and redo steps and save state. Because the physics in a virtual environment is arbitrary, idealised or inaccessible environments (such as zero gravity) can be presented. A virtual environment can provide forms of communication and interaction for groups, or students separated by distance or time, that are impossible in the real world.
- **Engagement**. We believe good VR is fun. Providing students with an engaging and enjoyable environment for learning should increase their attention on the subject. Providing a rich, memorable, multi-sensory experience may assist students to memorise the information presented in the environment.

Many of these advantages apply generally to any form of computer based or simulator based training. If we focus on traditional expectations of virtual reality user interfaces, then the key differentiator is the ability to convey content which is heavily spatial, involving three dimensional models, visualisation, or navigation, and content which is multi-sensory or incorporates multiple media. New generations of input and output devices, particularly haptic devices, are applicable to training for dextrous tasks.

One field where all of these benefits stand to be realised is in medicine, and particular in training for surgery and similar interventional procedures. Surgery is clearly high risk, and danger to both patients and surgical trainees exists when skills are gained in the operating room. The traditional model of surgical training is similar to an apprenticeship, and the learning opportunities for trainees are limited by the case mix of the hospital where they are resident. Safe surgery requires an excellent knowledge of anatomy, which is fundamentally spatial and three dimensional, as well as the associated procedural and clinical knowledge, and manual dexterity. It is not surprising, then, that simulation and virtual reality have been recognised as likely to have an important role in surgical courses now and in the future (Champion & Gallagher, 2003). This paper comes from our experience in investigating the design of virtual environments for surgical and medical training, and the examples included are all from that field.

1.3 Practical design of virtual environments

The quality of a virtual reality experience is sometimes judged against the standard of human experience of the real world. Particularly where the environment is designed to provide training about some real world subject domain, such as surgery, it is natural to assume that a more realistic simulation will produce a higher quality result. In practice, there are two forces that counter the drive towards increasing realism.

- **Cost**. Realistic simulation is expensive, both in the human resources required to develop the models, algorithms, and other necessary components, and in the computing resources required to deploy them. There is currently no such thing as a completely realistic model of an entire human patient, and if there were, very few could afford the facilities required to operate it. It is important to realise the benefits that are obtainable from currently existing hardware and software techniques that produce low fidelity simulations
- **Benefit**. Many of the benefits described above for virtual reality training exist precisely because it is not real. If a simulation were as dangerous, expensive, complicated and unpredictable as real-world

> training, there would be no value in it. We assume that VR training will be a supplement to real world training, not a replacement. Thus the design of a VR training environment should concentrate on exactly those areas where the unreality of VR offers the most benefit in comparison to the real world.

Thus for each feature of a virtual reality training environment, there is a tension between attempting to copy the existing reality of, for example, the operating theatre, or designing an alternate representation that is cheaper but performs the same or improved function in terms of training.

Several authors have considered how the fidelity of virtual reality contributes to immersion, presence, and training transfer (for example, Steuer 1993; Mantovani & Castelnuovo 2003; Whitton 2003). Whereas they consider the environment and immersion in it as a whole experience, in this paper we approach the problem by considering the environment as an aggregation of components, and analyse these independently. This gives greater scope for decomposition of the problem, and more flexibility in making design tradeoffs.

1.4 Discussion of design decisions

Where design tensions and trade-offs exist, it is convenient to have a language or a framework to discuss the decisions that have been made in a particular system. This could be for analysing existing designs, recording design rationale, or as a prompt during the design process to open up the space of possibilities. This paper proposes a framework for discussing the design space of perceptual representations within virtual training environments. We start, in Section 2, by describing some examples of networked virtual environments for surgical training. In Section 3 we introduce the AFRADERVITE framework, and then in Section 4 give specific examples of perceptual representations used in the training environment, and discuss the design decisions we made when implementing them. We conclude the paper in Section 5 with a discussion of the results of our design approach.

2 Example Training Environments

2.1 Networked Haptic Workbenches

The hardware system we typically use to implement our virtual training environments is a pair of "hand-immersive" workbenches that allows two users to collaboratively explore 3D objects in a virtual space that is roughly spanned by the movement of their hands and forearms (Stevenson et.al., 1999; Gunn, Hutchins, Adcock and Hawkins, 2003). The two users, usually instructor and student, can be located in the same room, or separated geographically. They communicate through shared manipulation of objects in the scene, and if separated, through high quality audio and video links. The workbench frame incorporates a mirror to create a space where virtual objects, rendered in stereo using shutter glasses, are collocated with the actions of the user's hands. A Phantom haptic device (from Sensable Technologies, Inc.) is located in the space, providing tracking and force interactions, and is used as the main manipulation device (see Figure 1). There is a six degree of freedom mouse for secondary manipulations. The workbench incorporates stereo speakers for synthesized sound from the environment. A camera and small video monitor are mounted at the rear of the frame, providing a facial view of each user to the other.

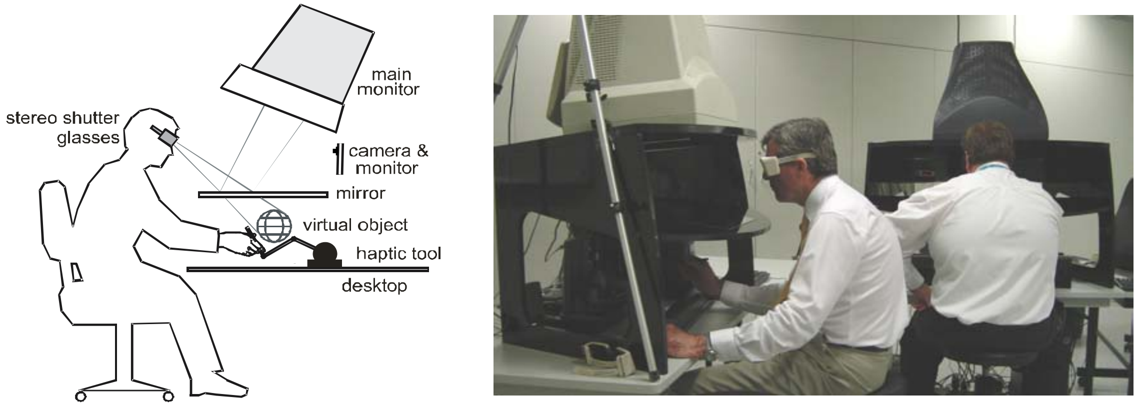

Figure 1: Diagram of the haptic workbench, and photograph showing networked adjacent haptic workbenches in operation.

In our typical collaborative examples, each workbench loads a copy of a particular model, such as a surgical training scenario, and changes and updates to the models are sent back and forth to keep the models on the two systems consistent. In terms of the taxonomy of networked virtual environments proposed by Macedonia and Zyda (1997) we use a replicated homogenous world database. Each user can see their own actions represented as the motion of tools in the scene, and also the actions of the other user, appearing in the same space. The communication between the two users is typically a mixture of speech, gesture, and shared collaborative action in the virtual environment.

2.2 Cholecystectomy training environment

A cholecystectomy is the surgical removal of the gallbladder. It is a fairly common operation, and one which surgical trainees encounter early in their training. In modern practice, cholecystectomy is usually performed laparoscopically, that is, using a surgical telescope and long instruments inserted through very small incisions. Laparoscopic surgery has many benefits for patients, but is a difficult skill to learn for surgeons. The increasing popularity of laparoscopic surgery (and similar technology-mediated interventions) is one of the driving forces behind the acceptance of structured, simulator-based skills training for surgeons (Cuschieri 1995; Champion & Gallagher 2003). We have developed a virtual environment for teaching the concepts (but not the manual skills) of laparoscopic cholecystectomy (Gunn, Hutchins, Adcock & Hawkins 2004; Gunn et. al., 2004). The environment includes a model of the organs of the abdomen, derived from segmented CT data, including the liver, gallbladder, stomach, pancreas and kidneys. The organs are deformable, and can be explored and manipulated. Several (but not all) of the key steps of the cholecystectomy procedure, including locating, clipping and cutting the cystic duct, can be performed.

2.3 Temporal bone surgery training environment

The temporal bone, located in the base of the skull, houses the delicate organs of the middle and inner ear that contribute to hearing and balance. Access to these organs, and other areas inside the skull, requires the bone to be removed by drilling. Safe surgical drilling in the region requires excellent knowledge of the complex 3D anatomy, confidence in the correct technique and approach to the procedure, and physical dexterity and tool handling skills. Training in temporal bone drilling is challenging; in particular, it is becoming more difficult to obtain access to the large number of human bone samples required to achieve competence. We have developed a virtual environment for teaching temporal bone surgery, using networked haptic workbenches (Hutchins, O'Leary, Stevenson, Gunn & Krumpholz 2005). The environment includes a drillable model of a temporal bone sample located within a simulated microscope view. The model includes several of the key anatomical landmarks that must be identified and safely dissected during the procedure. For the temporal bone environment, the haptic workbench hardware is augmented with an extra haptic device for the left hand, to control a sucker, and a foot pedal, to control the speed of the drill.

3 AFRADERVITE: A Framework for Describing Representations in Virtual Training Environments

3.1 Representations in virtual training environments

A virtual training environment is populated with computer generated objects that form part of the communication between the instructor and student. In many cases, these objects will be representations of objects that form part of the subject domain. For example, the cholecystectomy trainer contains a representation of a gallbladder, and the temporal bone surgery trainer contains a representation of a bone sample. The virtual environment also contains representations of events that occur as part of the subject domain, such as bleeding occurring from a vessel after a puncture. Also represented in the environment are actions, such as placing a clip on the cystic duct. If we take the viewpoint of one participant, there will be a representation of self, and another, possibly different, representation of others. We note that a typical virtual environment will also contain other content beyond the subject domain, including necessary control, navigation, and situational awareness interfaces for managing aspects of participation in the environment itself.

When analysing the design of the virtual environment, there are many questions we might ask of the representations. For the current discussion, we are mainly concerned with questions of the form:

- How is this object (or event or action) from the subject domain represented within the virtual environment?
- How are the participants (self and other) represented within the virtual environment?

In particular, we wish to consider these questions with respect to the cost and benefit of the representation in a practical environment, as discussed in Section 1.3. The following section presents a framework for addressing these questions. The framework is still in the early stages of conceptualisation, and is by no means a mature design tool. It will support discussion of design decisions, but it is not yet a methodology for optimising designs.

3.2 The AFRADERVITE framework

3.2.1 Symmetry and heterogeneity

In a multi-user virtual environment, different users could be provided with different representations of the same source material. For example, where two participants have different roles in the communication, such as instructor and student, we could expect them to have different representations of some objects, or different levels of control. The instructor, for instance, might have a unique representation that highlights unusual features that are yet to be discovered by the student, or that enables analysis and recording of information about the student's performance and errors. Thus we can classify representations as to whether they are symmetric or asymmetric across participants. The choice of an asymmetric representation expands the design question from the choice of a single representation to the choice of two (or more) representations, which can then be discussed individually within the following sections of the framework.

Similarly, we can consider heterogeneous representations across different classes of participant, for example early learners versus more advanced students. Some training environments will need to be delivered across multiple hardware platforms, which will have different capabilities in terms of what representations can be delivered. For example, different representations might need to be designed for platforms with and without haptics.

3.2.2 Multiplicity and augmentation

It is not necessary that a single object, event or action from the subject domain have a single representation to each participant in the virtual environment. Multiple simultaneous representations of the same thing can provide more complete coverage of the subject and overcome deficiencies in any one choice of representation. For example, a diagram, a photograph, and a 3D model of an object can convey different information. Different learners have different preferences in learning style and strategies, and so multiple representations might be required to support a broad learner base (Bransford, Brown & Cocking 1999; Groundwater-Smith, Cusworth & Dobbins, 1998).

One of the benefits of training in virtual environments is the ability to augment representations with features that go beyond what is available in the real world. So, when designing a representation for an object we need to consider not just how it is in reality, but how a somehow enhanced version of the object could improve its training value. Augmentation enhances the scope of the representation question to include not just the inherent features of the object, but also the added features.

3.2.3 Decomposition and multimodality

Objects from the subject domain usually have many different properties that could be represented in the virtual environment. From a perceptual point of view, we can consider how the object appears to each of the senses: the way it looks, sounds, feels, smells, tastes, etc. These can be further decomposed if necessary. For example, the way it looks might be decomposed into shape, colour, texture, and so on. When designing a representation for the object, each of these properties might be considered separately. For example, the way an object looks is usually represented in a virtual environment, but the way it smells is usually ignored. The types of decompositions that are considered will depend on the domain, and we do not prescribe them in the framework. Higher level properties of objects, such as topology, geometry, affordances and behaviours, will also need to be considered when designing representations.

Actions, interactions and communication will also need to be represented or facilitated in the virtual environment. If necessary, these too can be decomposed into multiple modes. For example, in considering the representation of another participant in the environment, we might separately consider how to facilitate communication using speech, facial expressions, and gesture. In considering an action that would be performed in the real world, such as controlling the speed of a surgical drill with a foot pedal, we would need to determine if the same action mode was necessary in the virtual environment, or whether it could be replaced by a different mode, such as a button press. (For further discussion on multimodal and multisensory representations see, for example, Bernsen, 1997, Martin 1995, Stanney, Mollaghasemi, Reeves, Breaux & Graeber, 2003; Nesbitt, 2004.)

The first three sections of this framework can be thought of as expanding the scope of the representation design question, and then breaking down the problem into components that can then be considered targets for representation. The remaining sections describe prioritising and filtering the resulting targets and then characterising chosen mappings.

3.2.4 Role and significance

The target items to be represented in the environment will have a variety of roles, and will vary in significance in respect to teaching in the subject domain. Some of the roles we have identified are:

- **Essential content**. These items are the core of the subject that the training experience is designed for. For example, models of human anatomy are essential content in most surgical training systems.
- **Context**. These items support the essential content by providing a context or background for the material. Examples: representations of the operating theatre; information on the simulated patient and the presenting symptoms.
- **Communication support**. These items support communication between participants. For example, we include a virtual whiteboard in our training environments to support discussions that go outside the boundaries of the explicitly represented content.
- **Situation awareness**. These items help the participants to remain aware of the state of the environment and the actions of the other participant. For example, we have operations such as tool selection provide audio feedback, so that both users are aware that the operation has occurred.
- **Engagement**. These items are included to enhance the richness and sensation of the experience, to support the sense of presence within the environment, or simply to make the experience more enjoyable.

In general, we would expect that any particular environment would include a balanced mix of representations fulfilling these roles.

The significance, importance or priority we assign to each target to be represented will depend on both its significance in the subject domain and the benefit that the virtual environment can offer over other forms of training. For example, surgeons learning temporal bone drilling must learn both the anatomy of the temporal bone, and how to hold the drill properly. Both are important to being a competent surgeon, but holding the drill can be taught easily outside the virtual environment with a real drill, whereas the anatomy presents more of a challenge, so its inclusion in the virtual environment is a higher priority. Other features of the subject domain, perhaps the colour and texture of the drill handle, are basically insignificant to the subject, and so can be given a low priority for accurate representation. On the other hand, leaving out too many seemingly insignificant details may reduce engagement or sense of presence within the environment.

3.2.5 Directness

Having established which targets are to be represented and their respective roles and significance, we can determine how they will be represented in the environment. We base this classification on the idea of directness of the representation, that is, how closely the representation matches the perception of the target in the subject domain.

- **Direct**. An object from the subject domain is directly included within the virtual environment, with little or no virtualisation. We further subdivide this class into:
  - **Actual inclusion**. A real object is actually included as part of the experience of the environment. For example, a tracked physical object such as a drill handle could be used as part of the user interface.
  - **Transmission**. A live video, audio, or other capture of the object is included in the environment. We use live transmission of video and audio to represent aspects of each participant to the other when they are not in the same room.
  - **Recording**. A pre-recorded video, audio, still image, or other capture is included in the environment. Although live and pre-recorded direct representations may appear the same, they have different properties with respect to timing and interactivity.
- **Mimetic modelled**. Objects are typically represented in virtual environments by constructing a model, and then synthesising a version of the object from the model that attempts to mimic reality. Most computer graphic representations of the visual properties of objects fall into this category. This class of representation can be further qualified in terms of some sort of fidelity scale. A basic approach

might be to sub-classify in terms of **high fidelity**, **medium fidelity** or **low fidelity** mimetic representations, although these terms are obviously not precisely defined. Usually, a higher fidelity representation will require more resources, and so must be justified in terms of the training value it delivers.

- **Substituted**. In this class, an indirect representation is used as a substitute for the original target. We further subdivide this class into:
  - **Indicative**. The existence of an object, event or action in the subject domain is indicated by the existence of a substitute object, event or action in the virtual environment. The substitute conveys few features beyond existence, but serves as a reminder that something is there. For example, in the temporal bone drilling simulation, indicative bone dust is generated by drilling. The virtual dust is not an accurate simulation of real dust, but reminds the surgeon that the sucker must be used during drilling to prevent dust from obscuring the view.
  - **Metaphorical**. An object or event from a different domain is used as a metaphor for the target. For example, in the bone drilling simulation a bleeding vessel is repaired with a sticking plaster. This is a metaphor for a much more complex procedure. In the cholecystectomy trainer, instantaneous healing of the patient (in order to have repeat attempts) is accompanied by the sound of angels singing.
  - **Mode replacement**. A representation in one mode replaces a different mode of the target. Examples: a visual signal indicates an event that would normally be detected by sound; a button press replaces a task that would be achieved by manual manipulation.
  - **Abstract**. We include a catch-all class for abstract representations that convey indirect information about the subject. Graphs, diagrams, and other visualisations would fall into this class.
- **Absent**. Not all of the targets for representation will be included in the virtual environment, many will remain absent. It is useful to consider explicitly which elements of the subject domain have been left out. As an example, we have chosen not to represent the smell of diathermy in our cholecystectomy trainer, because it would have little training benefit. More significantly, we have chosen not to represent irrigation of the bone in the temporal bone trainer. Although this would be of moderate training value, we have not found a suitable representation with a correspondingly moderate implementation cost.

4 Examples of Perceptual Representations in the Environments

4.1 Representations of the participants

4.1.1 Voice and face

We place a high priority on clear voice communication between instructor and student. Facial communication is important during early stages and when problems arise, but not during routine training. As described above, when we configure two workbenches in the same room, voice and face communication is “actually included” in the training environment without further effort. When the workbenches are separated, we rely on a high speed network to provide voice and face transmission (Adriaansen, Krumm-Heller & Gunn 2004).

4.1.2 Gesture

We suppose that the ability to communicate by gesture is also of high importance when teaching about a spatial subject. Within the virtual scene of our training environments, each user is primarily represented by tools that move in direct response to the motion of the Phantom haptic devices. Thus each user can clearly see their own gestures and those of the other user. The tool *positions* can be described as a direct transmission or a high fidelity mimetic model, and this means that pointing and other gestures are very natural within the environment.

4.1.3 Augmented gesture

We have augmented gesture with tools to enhance communication. A “vapour trail” mode makes a tool leave behind a gradually fading trail of images of itself as it moves. This facility is used to demonstrate actions in space and time. It is an abstract representation of motion. We have also implemented a “guiding hand” mode. This is a representation of the instructor’s ability to grab a student’s hand and guide it to a particular point in the scene. It is implemented by establishing a modelled haptic spring between the two tools. Visually, the guiding hand is indicative, haptically, it is a medium to low fidelity model, but the position information is high fidelity.

4.1.4 Viewpoint

In the temporal bone trainer, we have chosen to implement a shared viewpoint — both users always see the scene from the same position. In the cholecystectomy trainer, we experimented with a navigation interface. Users can move around the model independently. To support situation awareness between the users, the viewpoint of each is represented in the scene by a modelled pair of goggles similar to the shutter glasses that they are actually wearing to view the scene. This is an indicative visual representation of the viewpoint.

4.2 Representations of anatomy

4.2.1 Multiple representations

The anatomy in the surgical trainers is essential content, and we have chosen to represent it in multiple ways. The primary representation is a mimetic model, discussed in the following sections. In the cholecystectomy trainer we have included a virtual lightbox showing an intraoperative x-ray. In the temporal bone trainer, the virtual lightbox shows a slice-by-slice view of a CT scan of a bone (CT is computed tomography, a volumetric x-ray). In both systems, we have included video clips of surgical procedures being conducted. These recorded representations of the anatomy provide an additional source of information not provided in the models, and also serve as foci for anatomical and procedural discussions, and therefore potentially enhance communication.

4.2.2 Modelled geometry and appearance

In both systems, the geometry of the anatomy is modelled from segmented CT scans. This is a medium to high fidelity mimetic model, in terms of geometry. Ultimately, both models are represented as polygonal meshes (although the bone model is internally represented in a volumetric form) and displayed using standard computer graphics techniques, such as texture mapping and simulated lighting. The visual appearance is a low fidelity model, or in some cases just indicative. A more realistic view of the appearance of the anatomy can be gained from the recorded video clips.

The primary purpose of the temporal bone trainer is for the student to learn the location of, and correct approach technique for, significant anatomical landmarks embedded in the bone. In the real bone, differing regions are identified by a mix of colour, visual and haptic texture, and sound changes. We have instead used a false colouring scheme that broadly colours different regions of the bone according to their anatomical significance. For example, the area around the sigmoid sinus (a major vein) is coloured blue. Thus we have used a mode replacement representation instead of a high fidelity model across all these modes. We have also augmented the representation of the bone with the ability to become partially transparent to reveal the structures underneath. This simple capability, not imaginable in the real world, is frequently mentioned by students as a highlight of the learning experience.

4.2.3 Behaviours

In the cholecystectomy trainer we use a low fidelity model of the behaviour of the deformable organs under force. This is enough to be engaging in the scenario, but teaches little about the behaviour of real organs. We have included indicative representations of particular behaviours identified by the domain experts as being important to teaching about the scenario. For example if the gallbladder is stretched too hard, it will rupture, and a visual indication of bile fluid will leak into the abdomen. Similarly, the cystic duct will bleed if cut without correct clipping, and in the temporal bone trainer the major veins and arteries will indicate bleeding if they are punctured with a spinning drill.

4.3 Representations of surgical tools and actions

4.3.1 Tools and tool selection

Both training environments include a range of tools that can be used within the scene. Some of them are representations of surgical tools: probe, clipper, cutter, drill, sucker; others are parts of the environment: marker, eraser, guiding hand, navigator. Visually, the tools are indicative or low to medium fidelity models. We experimented with a representation of a surgical instrument tray for selecting tools. In practice, we found that a simple menu or keyboard selection was more usable. The information to be learned in the environment is which tool is to be used for which task, not how to select the tools from a tray (safe handling of tools in the operating room is important to learn, but can be practised in other ways).

4.3.2 Clipping and cutting the cystic duct

The operation of cholecystectomy requires locating the cystic duct, placing clips at either end, and then cutting between the clips. Our trainer includes indicative representations of these actions. For example, to clip the duct, the clipping tool is placed near the duct and the Phantom button is pressed. A visual marker is placed on the duct to indicate successful placement. These simple representations are enough to support a discussion about the correct procedure and technique, but not to train in manual skills.

4.3.3 The bone drill

The bone drill is visually represented by a medium fidelity mimetic model. We provide a selection of drill burrs of different sizes and properties. Part of the essential content of the trainer is learning when to use cutting burrs and when to use polishing burrs, and which sizes are appropriate. The drill speed is controlled by a foot pedal. We use a generic pedal as a substitute for the pedal control found in the operating theatre. A simple visual indication is used to show when the burr is spinning, and a low fidelity model of the sound of the drill motor gives feedback on the speed of the drill. We have experimented with adding an artificial haptic "weight" to the drill (a downward force), to more closely simulate the feel of a real surgical drill, which can be quite heavy.

The drill is used to remove bone from the bone model. The rate of bone removal is affected by the choice of drill size and type, the speed of the drill, and the type of bone being removed. Generally, the drilling operation is a low fidelity model, visually, haptically, and physically. The contact sound of the drill burr on the bone (in contrast to the sound of the drill motor) is not represented. However, the model is adequate to explore the surgical anatomy and approach, and to learn some of the techniques of drill handling. These were identified as the highest priority areas of essential content by our surgical collaborators.

5 Discussion

We use an iterative design process based on rapid prototyping and frequent communication with domain specialists. We believe this is necessary to effectively capture the educational requirements for training systems for highly specialised fields. In realising software components to meet these requirements, we have implemented a range of representations according to the training needs. The AFRADERVITE framework is essentially a distillation of our experience and approach in designing training systems.

Positive feedback from a public display of the cholecystectomy training environment has been reported in Gunn et. al. (2004). We are currently awaiting the results of a trial we have conducted to determine the efficacy of our temporal bone environment in terms of training transfer. As part of the trial, eleven surgeons in training were taught temporal bone drilling by an experienced instructor using our system, for one hour each. At the end of the trial we asked them to answer some subjective questions about the experience using a five point scale. In answer to the question "how would you rate your overall experience of this VR system?" 9 surgeons rated the experience 5 ("excellent"), and 2 rated it 4. To the question "would you recommend the VR training experience to a colleague?" all 11 answered 5 ("strongly"). In answer to the question "would you support formal inclusion of this type of training in the otology curriculum?" all 11 answered 5 ("strongly"). We are encouraged by these results, and observations of our systems in use, to think that our design approach is producing practical training environments that have real relevance to the target community. In further work, we can explore how different types of representations effect learning and engagement.

Acknowledgements

This work was funded under the CeNTIE project. The CeNTIE project is supported by the Australian Government through the Advanced Networks Program of the Department of Communications, Information Technology and the Arts. We would like to acknowledge the contributions made by the surgical experts to the design of the training environments discussed in this paper: for the cholecystectomy environment, Peter Cosman and Leroy Heinrichs; for the temporal bone drilling environment, Stephen O'Leary and Brian Pyman.

References

Adriaansen, T., Krumm-Heller, A. & Gunn, C. (2004). Enhancing human computer interaction in networked hapto-acoustic virtual reality environments on the CeNTIE network. *4th International Conference on Computational Science*, Springer Lecture Notes in Computer Science, Vol. 3037, 905-912.

Bernsen, N. O. (1997). Defining a taxonomy of output modalities from an HCI perspective. *Computer Standards and Interfaces*, 18 (6), 537-553.

Bransford, J. D., Brown, A. L. & Cocking, R. R. (1999). How People Learn: Brain, Mind, Experience, and School. Washington, D.C.: National Academy Press. (also from http://www.nap.edu/html/howpeople1)

Bricken, M. (1991). Virtual reality learning environments: potentials and challenges. Computer Graphics: SIGGRAPH 91, 25(3), 178-184.

Champion, H. R. & Gallagher, A. G. (2003). Surgical simulation — a 'good idea whose time has come'. *British Journal of Surgery*, 90, 767-768.

Cuschieri, A. (1995). Whither minimal access surgery: tribulations and expectations. *The American Journal of Surgery*, 169, 9-19.

Groundwater-Smith, S., Cusworth, R. & Dobbins, R. (1998) Teaching: Challenges and Dilemmas. Sydney: Harcourt.

Gunn, C., Hutchins, M., Adcock, M. & Hawkins, R. (2003). Trans-world haptic collaboration. *SIGGRAPH 2003 Sketches and Applications*, San Diego, July 27-31, 2003: ACM Press.

Gunn, C., Hutchins, M., Adcock, M. & Hawkins, R. (2004). Surgical training using haptics over long internet distances. *The 12th Annual Medicine Meets Virtual Reality Conference*, Newport Beach, January 14-17, 2004.

Gunn, C., Stevenson, D., Krumm-Heller, A., Srivastava, S., Youngblood, P., Heinrichs, L., & Dev, P. (2004). An interactive master class in remote surgery. *OZCHI 2004*, Wollongong, November 22-24, 2004

Hutchins, M., O'Leary, S., Stevenson, D., Gunn, C. & Krumpholz, A. (2005). A networked haptic virtual environment for teaching temporal bone surgery. *Medicine Meets Virtual Reality 13*, Long Beach, January 26-29, 2005: IOS Press.

Macedonia, M. R. and Zyda, M. J. (1997). A taxonomy for networked virtual environments. *IEEE Multimedia*, 4(1), 48-56.

Mantovani, F. & Castelnuovo, G. (2003). Sense of presence in virtual training: Enhancing skills acquisition and transfer of knowledge through learning experience in virtual environments. In G. Riva, F. Davide, W.A IJsselsteijn (Eds.), *Being There: Concepts, effects and measurement of user presence in synthetic environments*. Amsterdam: Ios Press.

Martin, J.-C. (1995). TYCOON: Theoretical framework and software tools for multimodal interfaces. In Lee, J. (Ed.), *Intelligence and Multimodality in Multimedia Interfaces: Research and Applications*. Menlo Park: AAAI Press.

Milgram, P. & Kishino, F. A. (1994) Taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems (Special Issue on Networked Reality)*, vol. E77-D, no. 12.

Nesbitt, K. (2004) MS-Taxonomy: A conceptual framework for designing multi-sensory displays. *Information Visualisation 2004*. 665-670.

Riva, G. (1999) Virtual reality as communication tool: A sociocognitive analysis. *Presence, Teleoperators, and Virtual Environments*, 8, 462-468.

Stanney, K. M., Mollaghasemi, M., Reeves, L., Breaux, R., & Graeber, D.A. (2003). Usability engineering of virtual environments (VEs): Identifying multiple criteria that drive effective VE system design. *International Journal of Human-Computer Studies*, 58, 447-481.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication*, 42(4), 73-93.

Stevenson, D., Smith, K., Veldkamp, P., McLaughlin, J., Gunn, C. & Dixon, M. (1999) Haptic Workbench: A multisensory virtual environment. *The Engineering Reality of Virtual Reality 1999, Electronic Imaging '99*, San Jose, January 23-29, 1999.

Whitton, M. C. (2003). Making virtual environments compelling. *Communications of the ACM*, 46(7), 40-46.