4th INTUITION International Conference on Virtual Reality and Virtual Environments proceedings
4-5 October 2007, Athens, Greece
ISBN-978-960-254-665-9
Institute of Communication and Computer Systems of the National Technical University of Athens, 2007

Collaborative Evaluation of a Haptic-based Medical Virtual Environment

Shamus P. Smith[1], Susan Todd[2]

(1) Department of Computer Science
Durham University, Durham, DH1 3LE, UNITED KINGDOM
E-mail: shamus.smith@durham.ac.uk

(2) UK Haptics Ltd.
Albert House, Rothbury, NE65 7SR, UNITED KINGDOM
E-mail: sue@ukhaptics.co.uk

Abstract

There is increasing use of virtual environments for medical training applications. Simulated medical environments allow clinical staff to practise medical procedures without endangering patients. Haptic devices are a key technology in many such systems adding a sense of touch to virtual interactions allowing healthcare practitioners to practise and test their clinical skills in life-like situations. However, evaluating haptic-based virtual environments can be difficult as gathering user feedback on haptic interaction by traditional methods, such as post-session interviews and questionnaires, can be error-prone if users misremember their experiences.

This paper describes the use of cooperative evaluation, a modified think aloud protocol, to evaluate the usability of a commercial haptic-based medical virtual environment called Virtual Veins. An evaluation was performed with 3rd year medical students using cooperative evaluation to evaluate Virtual Veins in a target deployment environment. Although the cooperative evaluation was successful in eliciting a number of usability issues, there were few haptic related verbalisations from participants. In haptic-based systems the evaluator does not share the sensory feedback with the participant. This can be problematic in questioning the participants about their haptic experience and encouraging haptic-specific comments.

Keywords: Virtual environments, Haptics, Evaluation, Medical training

1 Introduction

There is increasing use of virtual environments (VEs) for medical training applications (Satava and Jones, 2002), for example interaction with virtual patients (Raij et al, 2006) and training for medical procedures (Ganai et al, 2007; Hansen et al, 2004; Vidal et al, 2005). Simulated medical environments allow clinical staff to practise medical procedures without endangering patients. Haptic devices are a key technology in many such systems. Haptic technology adds a sense of touch to virtual interactions allowing healthcare practitioners to practise and test their clinical skills in life-like situations. However, evaluating haptic-based VEs can be difficult as gathering user feedback on haptic interaction by traditional methods, such as post-session interviews and questionnaires, can be error-prone if users misremember their experiences.

Clinical practitioners are a diverse user population with varying demands on system *usability*. Also 3D interaction is difficult (Bowman et al, 2005, pg6) and novice users of 3D user interfaces may require interaction to be simplified and easily learnable (Bowman et al, 2005, pg328). As traditional measures of haptic-based systems such as haptic rendering realism and real-time response rates are now default requirements, usability issues must drive system development and any associated evaluations. However it is unclear how usability can be successfully evaluated in the context of haptic-based systems.

This paper will describe the use of cooperative evaluation, a modified think aloud protocol, to evaluate the usability of a commercial haptic-based medical VE called Virtual Veins. An evaluation was performed with 3rd year medical students using cooperative evaluation to evaluate Virtual Veins in a target deployment environment to support usability analysis.

The remainder of this paper is as follows. Section 2 describes related work and Section 3 outlines the medical training simulator to be evaluated. Section 4 describes the evaluation study. The results are presented in Section 5, followed by conclusions in Section 6.

2 Related Work

An overview of usability evaluation for VEs is outside the scope of this paper (see Bowman et al, 2005; Hix and Gabbard, 2002). However, it is uncommon to find reports of usability evaluations specifically for haptic-based systems. Both medical and non-medical haptic-based applications are typically evaluated for performance metrics and the accuracy of the haptic renderings. Although both are important in supporting realistic haptic interaction, as haptic-based systems enter mainstream usage, user centric measures such as usability are increasingly important. Karaseitanidis et al. (2006) note, in general, that virtual reality technology "to date has focused on the highest possible technical standards, with less attention on the evaluation process involving end-users."

Satava and Jones (2002) observe that for medical applications the final determinate for successful use is the end user and the need for such systems to be user-friendly. One important aspect of usability evaluation is eliciting the user experience. However, gathering participant feedback of haptic interaction by traditional information gathering methods such as post-session interviews (see Raisamo et al, 2006) and questionnaires (see Baillie et al, 2005) can be error-prone if participants misremember their experiences. This has been a recurrent VE issue when measuring subjective experiences such as immersion and presence (Slater, 2004).

One alternative is to invite participants to verbalise their experience as they have it, i.e. using a *think aloud* protocol (Dix et al, 2004, pg343), during the session. Any verbalisations can be captured on audio recordings. This evaluation mechanism is not without its problems because the user may (i) forget to verbalise their actions, (ii) feel self conscious verbalising or (iii) miss reporting features that the evaluator is particularly interested in.

2.1 Cooperative Evaluation

Wright and Monk (1991) proposed an augmented form of the think aloud protocol called *cooperative evaluation*. This is a concurrent think aloud verbal protocol where users are encouraged to treat the evaluation as a shared experience with the evaluator and may ask questions at any time. Also subjects may be prompted to explain actions as the session progresses. Similar to an in-session interview, it can provide qualitative feedback on the motivations for the user's current behaviour. Marsh and Wright (1999) describe how the qualitative data from a cooperative evaluation session can be quantified. The think aloud verbalisations can be assessed according to their quantity and quality. Quantity can be determined by counting all the verbalisations. Quality is attained by judging the value of each verbalisation.

Low quality problems are judged to be of low importance or impact to the environment or task under analysis and are more likely to be overcome with more usage and time in the environment. *High quality* problems are judged to be of high importance to the designers, and something that is not likely to be reduced by prolonged use, i.e. something that may shock or startle a user, for example moving through or colliding with objects.

Cooperative evaluation has been successfully used for the evaluation of usability issues in visually-based VEs (Marsh and Wright, 1999; Patel et al, 2003; Smith and Hart, 2006). One drawback of this technique is that prompting the user to talk may interrupt their train of thought and be distracting. However, Marsh and Wright (1999) found that users encountered no significant addition to the time taken to complete their task whist thinking aloud as when not.

This paper describes the use of cooperative evaluation in the context of a haptic-based VE. It is hoped that cooperative evaluation can identify haptic-based usability issues and complement traditional evaluation tools such as performance metrics and user questionnaires. A commercial medical simulator for venepuncture training has been evaluated.

3 A Medical Simulator for Venepuncture Training

Venepuncture, or venipuncture, is the puncture of a vein, esp. with a hypodermic needle to withdraw blood or for intravenous injection (OED, 1989). Only clinical practitioners who are trained in venepuncture may carry out this procedure (Jordan, 2005). As with surgical education in general, extensive practice on patients with close faculty supervision may be an effective learning technique, but is a potentially cost-ineffective method (Prystowsky et al, 1999). Within this high faculty-to-student ratio setting, venepuncture instructors use “oranges, plastic arms or even each other for practising the skill” (Engum et al, 2003). There is continuing research and development of virtual reality technologies to simulate virtual patients on whom students can practice such procedures. Computer-based systems are typically not used as an alternative to traditional techniques but aim to move the student farther up the learning curve prior to a real patient contact (Engum et al, 2003). There is potential for providing students with a risk free environment where they can gain confidence and competence. In this paper, we focus on a specific commercial system for venepuncture training called *Virtual Veins*.

3.1 Virtual Veins

Virtual Veins, developed by UK Haptics Ltd., is a virtual reality training simulator to support healthcare practitioners in acquiring, developing and maintaining the skills necessary to perform venepuncture in a range of realistic scenarios within a safe controlled environment (see Figure 1). Virtual Veins allows healthcare practitioners across a range of disciplines, for example nursing, paramedics, blood transfusion staff and sports specialists, to practice skill-based procedures. In addition Virtual Veins provides metrics, for example needle angle, to support feedback during practice sessions and performance measurement for testing. After a practice or test session, the user can review their performance in an online report containing user and session details, an event log of the session i.e. skin/vein penetrations/retractions, a screenshot of the first skin penetration and metrics including skin insertion angle, skin retraction angle, bevel angle, vein insertion angle, vein retraction angle, vein diameter, vein penetration length and vein penetration depth.

Virtual Veins training sessions involve the user logging into the system via a desktop PC, performing venepuncture using a haptic workstation and then reviewing a report of their performance on the desktop PC.

4th INTUITION International Conference on Virtual Reality and Virtual Environments proceedings
4-5 October 2007, Athens, Greece
ISBN-978-960-254-665-9
Institute of Communication and Computer Systems of the National Technical University of Athens, 2007

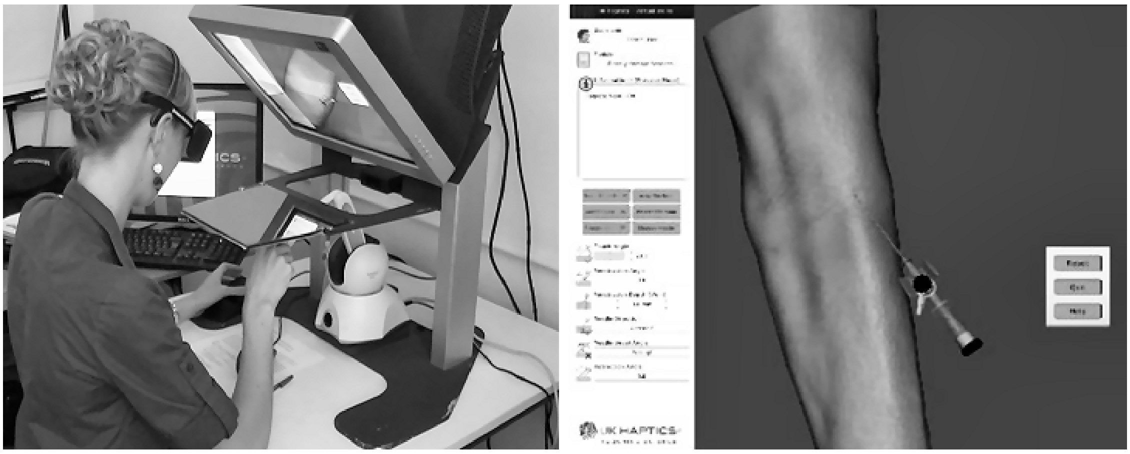

Figure 1: Virtual Veins in use and the user view of the venepuncture procedure.

3.2 Haptic Workstation

Virtual Veins uses a Reachin Display (Reachin, 2007) to enable the co-location of graphics and haptics to provide a high quality 3D experience. CrystalEyes 3D shutter glasses (Reald, 2007) are worn by the user to provide 3D stereoscopic views, a SpaceMouse (3Dconnexion, 2007) is used as a 6DOF device to manipulate the virtual arm/hand and a PHANTOM Omni haptic device (Sensable, 2007) is used to manipulate a virtual catheter, needle or syringe in the right hand (see Figure 1). The PHANTOM Omni provides a sense of touch so that users can feel the interaction between the virtual hand/arm model and the venepuncture tool, for example feeling a needle penetrate the skin and vein of a virtual arm.

4 Evaluation Study

A study using cooperative evaluation was conducted on Virtual Veins with first time users. A within-subject design was used for the study with all the participants receiving the same treatment (see Section 4.1). Eleven participants volunteered for the study but post-evaluation analysis found the audio was unusable with one participant. Therefore the research described here will focus on the ten participants with usable audio recordings.

The participants were all 3rd-year medical students (i.e. trainee doctors) at the Chantler Clinical Skills Centre, King's College London. The evaluation session had been previously advertised and the participants were all walk-ins between classes. All the participants were female, right-handed and within the 18-25 age range. All participants used computers on a daily basis but not virtual reality systems.

All participants used computers for educational activities, 5 participants used computers for work activities and 7 participants used computers for entertainment. Only one of the participants reported computer game usage and this was on a PlayStation Portable console. This was the first deployment of Virtual Veins at the Chantler Clinical Skills Centre therefore the participants had no prior experience with Virtual Veins. Participants were unpaid and all ten participants completed the evaluation sessions.

The evaluation was carried out in an empty room where Virtual Veins was deployed for a single day. The equipment consisted of a Dell Precision 670 workstation, Twin Xeon 3.00GHz CPU's, 2GB of RAM, Windows XP SP2, NVidia Quadro FX 3450/4000 graphics card, Dell 1907FP (Digital) LCD Monitor, a standard mouse and keyboard, 3D SpaceMouse, PHANTOM Omni, flat panel display and a pair of CrystalEyes 3D shutter glasses (see Figure 1). The evaluator (the first author) sat behind the participant with a video camera on a tripod to record video and audio from the evaluations.

4th INTUITION International Conference on Virtual Reality and Virtual Environments proceedings
4-5 October 2007, Athens, Greece
ISBN-978-960-254-665-9
Institute of Communication and Computer Systems of the National Technical University of Athens, 2007

4.1 Evaluation Procedure

Before entering the evaluation room, subjects were given an eligibility test. Virtual Veins utilises shutter glasses and requires participants to have 3D stereoscopic vision. It was therefore important to exclude any participants (i) with a history of photosensitive epilepsy and (ii) without 3D stereoscopic vision. Any participants with a history of epilepsy or photosensitive seizure, or lacking stereo vision were excluded from the study. A Frisby Stereotest (Frisby, 2007) was used to determine stereo vision in the participants. Demographic, computer usage and venepuncture training experience information was gathered in a pre-session questionnaire. A consent form was also signed by all participants. This included explicit permission for the use of video footage collected during the sessions.

On entering the study room, the evaluator (the first author) introduced the equipment and explained its function. The participants were given an information sheet outlining the tasks that they were about to attempt. Participants were to (i) log into the Virtual Veins system (using a standard 2D mouse, keyboard and desktop PC), (ii) perform venepuncture in a Virtual Veins practice session (using the haptic workstation), (iii) review the practice session report (on the desktop PC), (iv) perform venepuncture in a test session, (v) review the test session report and (vi) log out of the system. The use of a think aloud verbal protocol and the cooperative nature of the session were explained to the participants.

When using Virtual Veins, a practice session differs from a test session in two ways. Firstly in test mode the user receives no real-time feedback on performance metrics, i.e. needle angles. In practice mode, Virtual Veins provides feedback on the needle angle/bevel which is coloured green when an appropriate value is active and red when a poor value is active. Secondly, in practice mode, the user can change the transparency of the skin on the virtual arm model (see Figure 2). This allows the user to identify the vein and bone structure in the virtual arm model. This feature is not available in testing mode.

After completing the test session and logging out of the system, participants were thanked for their help and asked to complete a post-session questionnaire. This questionnaire contained questions about their satisfaction with the interface, realism of the training and engagement with the system. The questionnaire contained questions adapted from (Smith and Marsh, 2004). An analysis of the performance metrics and participant questionnaires is considered elsewhere (Smith and Todd, paper in preparation).

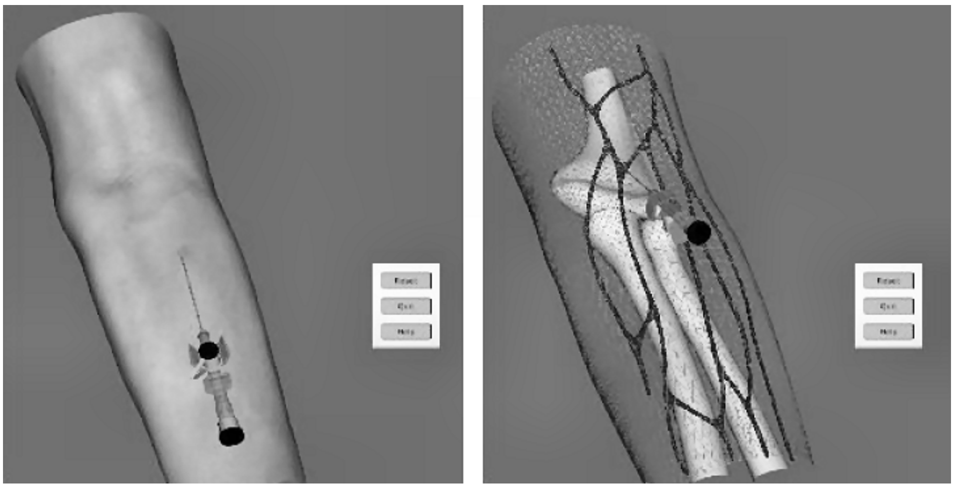

Figure 2. User view with and without the skin layer.

5 Results

The average time for an evaluation session was 11 minutes and 9 seconds. During this time the participants interacted with the desktop PC - with 2D mouse and standard keyboard - to start one practice and one test session and the haptic workstation - 3D mouse, PHANTOM Omni and shutter glasses - to perform the venepuncture procedure once per session.

5.1 Usability Analysis

Usability issues noted by the evaluator during the sessions were augmented with analysis of the video footage post-session. 81 usability issues were identified and grouped into 10 general areas namely; starting/stopping sessions (I1), use of 3D buttons (I2), pre-post venepuncture procedure (I3), injection feedback (I4), report feedback (I5), 3D interface (I6), 3D arm realism (I7), 3D mouse (I8), 2D mouse (I9) and screen confusion (I10), e.g. looking at the 2D screen with the shutter glasses on or looking at the 3D screen without the shutter glasses on. A summary of the usability issues can be seen in Table 1 where the user verbalisation/observations have been characterised as *L* for low quality and *H* for high quality (by the first author). Low quality issues are likely to be overcome with more usage and time with the system while high quality issues are not likely to be reduced by prolonged system use.

Table 1. Summary of usability issues by user (L = low quality, H = high quality).

Issue/ User	**I1**	**I2**	**I3**	**I4**	**I5**	**I6**	**I7**	**I8**	**I9**	**I10**	**Total**
User1	L	2L	L,H	-	H	3L	2L,H	L	L	L	15
User2	-	2L	L	2L	-	2L,2H	H	L	-	-	11
User3	L	6L	L	3L	2H	H	L	-	-	-	15
User4	-	L	-	L	H	-	-	-	-	-	3
User5	-	L	L	3L,H	L	-	-	-	-	-	7
User6	-	L,H	L	-	-	L	H	L	-	-	6
User7	-	L	2L	L,H	-	2L,2H	-	-	-	L	10
User8	-	L	L	L	H	L,H	-		-	-	6
User9	L	-	L,H	2L	-	-	-	L	-	-	6
User10	-	-	-	-	-	L	-	-	-	L	2
Total	3	16	11	15	6	16	6	4	1	3	81

There are three main features identified from the usability analysis. Firstly, the number of entries in Table 1 initially indicated that cooperative evaluation had been very successful in identifying 81 issues. The three most frequent issues identified were in the use of the haptic workstation, namely 3D buttons (I2), injection feedback (I4) and the 3D interface (I6). However on analysis of the raw data, none of the 47 issues related to haptic issues as they were concerned with visual feedback issues or misunderstandings in the way the venepuncture procedure had been represented by the system. None of the identified issues directly involved haptic properties of the interaction, i.e. with what the participants were *feeling* with the haptic device. The lowest usability issues were with interaction with the 2D desktop component of the system (I1, I9 and I10). This is not surprising as the participant's all used desktop systems on a daily basis.

The lack of haptic responses from the participants may be that both the participants and the evaluator lack a common vocabulary to describe haptic interaction. In addition, the evaluator does not feel the haptic feedback and can thus not easily identify when haptic related issues are relevant and further questioning is required. This is in contrast to the use of cooperative evaluation for visually-based VEs where the participant and evaluator can share the visual experience.

Secondly, of the 81 issues identified, only 19 were classed as high quality issues. Six high quality issues indicated 3D interface (I6) issues. On further analysis of the video recordings these issues were all to do with a visual-based technical problem on the system setup for the evaluations. Although this temporarily affected the co-location of visual and haptic feedback at one point during the venepuncture sessions, it was a quirk of the technology being used on the day of the evaluations and not representative of system interaction. Also the high quality issues for report feedback (I5) were on the desktop PC component of the system and not related to haptic interaction.
Finally, we are interested in the number of *issues per user* collected during the session. The first three users prompted the most issues with the system. The participants are listed chronologically in Table 1. One concern with cooperative evaluation is that the evaluations may be biased by the level of activity by the evaluator. As the evaluation sessions went on, in general fewer usability issues were identified by the participants. It may be that the evaluator was suffering fatigue by the final sessions and hence prompting the participants with fewer questions, or the evaluator may have biased the evaluations by incorporating knowledge from the initial evaluations into the later sessions. Also in the later sessions, the evaluator may have pre-empted usability issues, which then did not show in the usability analysis. In order to examine this, an analysis of transcripts from the evaluation sessions was performed.

5.2 Transcript Analysis

The ten evaluation sessions were transcribed in order to compare the session durations with the words spoken by the evaluator and the participants. A summary of the transcript data is shown in Figure 3. The evaluator word count has an average of 635 and as seen in Figure 3, there was no reduction in evaluator verbalisations in the later sessions. However, Figure 3 does indicate that the participants were less vocal in the later sessions.

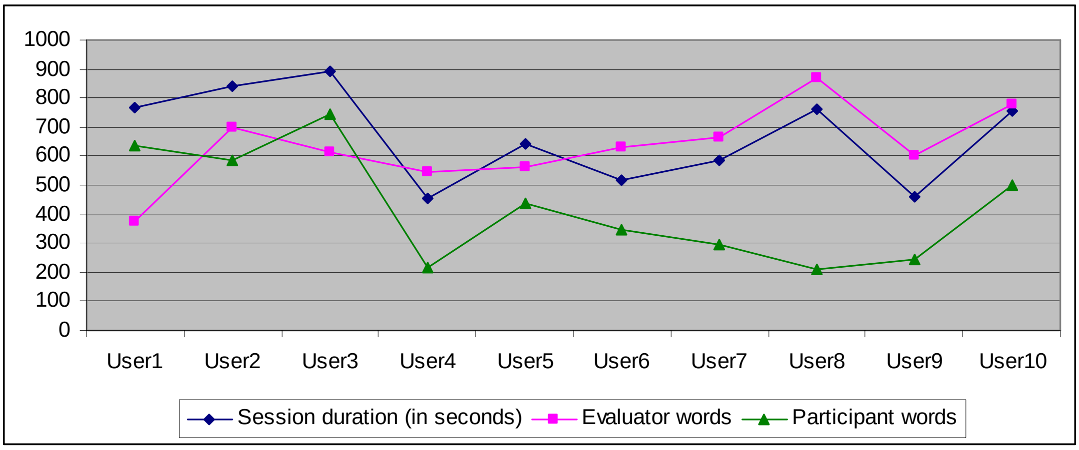

Figure 3. Summary of cooperative evaluation transcript data.

A Pearson r correlation test was performed on the transcript data and positive correlations were found for *participant words/time* (0.78, p>0.005), *participant words/usability issues* (0.71, p>0.025) and *time/usability issues* (0.58, p>0.05). The first and last correlations are unsurprising as it is expected that there would be more participant verbalisations and identified usability issues with extended session times. The second correlation supports cooperative evaluation usage as increasing participant verbalisations are eliciting more usability issues, i.e. the participant is not just chatting to the evaluator but expressing useful comments. Interestingly there is a negative correlation between evaluator words and usability issues (-0.45). However for this sample size this is not significant. Further analysis of the session transcripts indicate that, due to the novel nature of the deployed technology, tutorial verbalisations were required from the evaluator. Also as noted in Section 5.1, the evaluator may have pre-empted usability issues in the later sessions. Future studies will require explicit system tutorials before the cooperative evaluation in order to avoid skewing the verbalisation metrics.

6 Conclusions

Haptic-based systems are reaching more mainstream usage, particularly for healthcare professional training. When such systems are evaluated, traditional evaluation methods may not be suitable to capture an accurate view a user's experience. Cooperative evaluation, where user verbalise their experiences is a promising evaluation method for such systems.

This paper has described the evaluation of a commercial haptic-based medical VE. A user study was performed with 3rd-year medical students to support usability analysis. Although a number of usability issues were identified, the verbalisations from the evaluation sessions were not as descriptive as hoped. Few explicit haptic issues were identified. As evaluators lack haptic feedback it was found that this can be problematic in shaping questions during an evaluation session. Also the evaluator may have pre-empted usability issues in later evaluations sessions. Future work involves developing a framework of haptic-oriented verbalisations that can be used to both structure the evaluations to avoid evaluator bias and increase the quality of participant responses during haptic-based system evaluation.

7 Acknowledgements

The authors would like to thank Dr Sally Richardson (King's College London), the student participants from the Chantler Clinical Skills Centre (KCL), Gary Todd (UK Haptics Ltd.), Jonathan Lee (REDSS, Durham University), Dr Barbara Froner and Brendan Hodgson (Computer Science, Durham University) and Professor John Frisby (www.frisbystereotest.co.uk). This work was funded in part by a Durham University Initiating Business Collaboration Grant, UK Haptics Ltd. and the HEFCE Centre for Excellence in Teaching and Learning in Computing (Active Learning in Computing, Durham University).

8 References

3Dconnexion (2007): http://www.3dconnexion.com/ [last access 16/04/2007].

Baillie S., Mellor, D.J., Brewster S.A. and Reid, S.W.J. (2005): Integrating a Bovine Rectal Palpation Simulator into an Undergraduate Veterinary Curriculum. *Journal of Veterinary Medical Education 32*, 1, 79-85.

Bowman, D. A., Kruijff, E., LaViola Jr., J. J. and Poupyrev, I. (2005): *3D User interfaces: Theory and Practise*. Addison Wesley, USA.

Dix, A., Finlay, J., Abowd, G. D. and Beale, R. (2004): *Human-Computer Interaction*. Third edition. Pearson/Prentice Hall, Harlow, England.

Engum, S. A., Jeffries, P. and Fisher, L. (2003): Intravenous catheter training system: Computer-based education versus traditional learning methods. *American Journal of Surgery 186*, 67–74. Exerpta Medica.

Frisby, J. (2007): Frisby Stereotest, http://www.frisbystereotest.co.uk/ [last access 03/03/07].

Ganai, S., Donroe, J. A., St. Louis, M. R., Lewis, G. M. and Seymour, N. E. (2007): Virtual-reality training improves angled telescope skills in novice laparoscopists. *The American Journal of Surgery 193*, 206-265.

Hansen, K. V., Brix, L., Pedersen, C. F., Haase, J. P. and Larsen, O. V. (2004): Modelling of interaction between a spatula and a human brain. *Medical Image Analysis 8*, 23-33.

Hix, D. and Gabbard, J. L. (2002): Usability engineering of virtual environments. In K. M. Stanney (Ed): *Handbook of Virtual Environments: Design, Implementation and Applications*. 681– 699. Lawrence Erlbaum Associates, New Jersey.

Jordan, R. (2005): Policy and procedure for venepuncture. Northgate and Prudhoe NHS Trust, July. CP 18.

Karaseitanidis, I., Amditis, A., Patel, H., Sharples, S., Bekiaris, E., Bullinger, A. and Tromp, J. (2006): Evaluation of virtual reality products and applications from individual, organizational and societal perspectives The “VIEW” case study. *International Journal of Human-Computer Studies 64*, 251–266.

Marsh, T. and Wright, P. (1999): Co-operative evaluation of a desktop virtual reality system. In S. P. Smith and M. D. Harrison (Eds): *Workshop on User Centered Design and Implementation of Virtual Environments*, The University of York, 99–108.

OED (1989): *Oxford English Dictionary*, Second edition, Oxford University Press.

Patel, H., Nichols, S., Stedmon, A., Tromp, J., Griffiths, G., Bayon, V., Marshall, E., Johansson, E., Mourouzis, A., Karampelas, P., Stefani, O., Hoffmann, H., Karaseitanidis, G., Tzoumpas, A., Letourneur, S., Bullinger, A. and Mager, R. (2003): Pilot outline and results. Deliverable for the European Commission, *IST Project VIEW of the Future*. IST-2000-26089.

Prystowsky, J. B., Regehr, G., Rogers, D. A., Loan, J. P., Hiemenz, L. L., and Smith, K. M. (1999): A virtual reality module for intravenous catheter placement. *American Journal of Surgery 177* (February), 171–175.

Raij, A., Johnsen, K., Dickerson, R., Lok, B., Cohen, M., Bernald, T., Oxendine, C., Wagner, P. and Lind, D. S. (2006): Interpersonal scenarios: Virtual ≈ real? In *Proceedings of IEEE Virtual Reality Conference*, IEEE Computer Society, 59-66.

Raisamo, R., Hippula, A., Patomäki, S., Tuominen, E., Pasto, V. and Hasu, M. (2006): Testing usability of multimodal applications with visually impaired children. *IEEE Multimedia 13*, 3, 70-76.

Reachin (2007): http://www.reachin.se/products/Reachindisplay/ [last access 06/03/2007].

Reald (2007): http://www.reald-corporate.com/scientific/ [last access 16/04/2007].

Satava, R. M and Jones, S. B. (2002): Medical applications of virtual environments. In K. M. Stanney, (Ed): *Handbook of Virtual Environments: Design, Implementation and Applications*. 937–957. Lawrence Erlbaum Associates, New Jersey.

Sensable (2007): http://www.sensable.com [last access 16/04/2007].

Slater, M. (2004): How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence 13*, 4, 484–493.

Smith, S. P. and Hart, J. (2006): Evaluating distributed cognitive resources for wayfinding in a desktop virtual environment. In Y. Kitamura, D. Bowman, B. Fröhlich, and W. Stürzlinger, (Eds): *IEEE Symposium on 3D User Interfaces 2006*, IEEE, 3–10.

Smith, S. P. and Marsh, T. (2004): Evaluating design guidelines for reducing user disorientation in a desktop virtual environment. *Virtual Reality 8*, 55–62.

Vidal, F. P., Chalmers, N., Gould, D. A., Healey, A. E. and John, N. W. (2005): Developing a needle guidance virtual environment with patient-specific data and force feedback. *International Congress Series 1281*, 418–423.

Wright, P. C., and Monk, A. F. (1991): A cost-effective evaluation method for use by designers. *International Journal of Man-Machine Studies 35*, 6, 891–912. Academic Press.