Accepted for the International Journal of Clinical Skills
(Forthcoming 2008)

Usability evaluation of a haptic-based clinical skills training system

Dr Shamus P. Smith[1] BSc Hons PhD
Lecturer in Computer Science
Durham University

Susan Todd BSc, MBCS, MIoD
Technical Director
UK Haptics Ltd.

Abstract

Simulated medical environments allow clinical staff to practice clinical procedures for longer than traditional training methods without endangering patients, with resulting skill improvements. Haptic devices are a key technology in such systems allowing healthcare practitioners to practice and test their clinical skills in life-like situations. The potentially diverse population of users for such clinical training systems have varying demands on *usability* but it is unclear how usability can be successfully evaluated in the context of haptic-based systems. In order to justify deployment costs and any pedagogical aims, such systems require appropriate evaluation.

This paper describes the evaluation of a commercial haptic-based clinical training system. A user study has been performed with 3rd-year medical students using three complementary evaluation methods. The aim is to evaluate haptic technology using first time users from a target deployment environment and gather both qualitative and quantitative data to support usability analysis. In contrast to evaluating visually-based systems, with haptic-based systems the evaluator does not share the haptic feedback with the participant. It was found that this can be problematic in shaping questions during an evaluation session and interpreting the collected data.

KEYWORDS: Medical training, clinical skills, venepuncture, usability evaluation, virtual environment

Introduction

Virtual environments are a growth area for clinical skills training [1], for example interaction with virtual patients [2] and training for medical procedures [3, 4, 5]. Simulated medical environments allow clinical staff to practice medical procedures for longer than traditional training methods without endangering patients, with resulting skills improvement. Also staff can be assessed more consistently and effectively by teaching staff. A key technology to support such systems is the use of haptic devices. Haptic technology adds a sense of touch to virtual interactions allowing healthcare practitioners to practice and test their clinical skills in life-like situations. The potentially diverse population of users have varying demands on the *usability* of such systems. Satava and Jones [1] observe that for medical applications the final determinate for successful use is the end user and the need for systems to be user-friendly. In addition to traditional evaluation measures such as performance metrics, usability issues are increasingly important in a system's development

---

[1] Corresponding author: shamus.smith@durham.ac.uk, tel +44 (0)191 3344284, fax +44 (0)191 3341701

and evaluation. However it is unclear how usability can be successfully evaluated in the context of haptic-based systems.

One important aspect of usability evaluation is eliciting the user experience. Gathering participant feedback of haptic interaction by traditional information gathering methods such as post-session interviews and questionnaires can be error-prone if participants misremember their experiences. One alternative is to ask users to verbalise their experiences when using a system. *Cooperative evaluation* is a modified think-aloud verbal protocol that has been successful in evaluating usability in visually-based virtual environments [6, 7].

This paper describes the evaluation of a commercial haptic-based clinical skills training system. Three complementary evaluation techniques have been used. A cooperative evaluation was conducted and the results augmented with performance metrics and user questionnaires. The aim of the study is to evaluate a clinical skills simulator using first time users of the system from a target deployment environment and gather both qualitative and quantitative data to support usability analysis.

A medical simulator for venepuncture training

Venepuncture is the puncture of a vein, esp. with a hypodermic needle to withdraw blood or for intravenous injection[2]. Failed attempts at venepuncture can "increase patient discomfort, delay necessary therapeutics or testing, create hostility between the patient and staff, or result in stress to staff members performing the procedure"[8]. Therefore only clinical practitioners who are trained in venepuncture may carry out this procedure [9]. Hewitt and Roberts [10] observe that "students require opportunity to develop dexterity for new skills before they practice on patients for the first time, familiarising themselves with the equipment and steps for achieving the skill."

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 [11]. Within this high faculty-to-student ratio setting, venepuncture instructors use "oranges, plastic arms or even each other for practising the skill" [12]. 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 [12]. 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*. The haptic-based experience provided by Virtual Veins is supported by a front-end system called the *Clinical Skills Trainer*.

Clinical Skills Trainer

UK Haptics Limited[3] has developed a learning environment to complement existing medical procedure training methods. This environment currently consists of a training management component called the Clinical Skills Trainer and a training solution called Virtual Veins for practising the venepuncture procedure. Users log into personalised accounts in the Clinical Skills Trainer to run Virtual Veins practice and testing sessions. The system provides each user with an account which documents their progress through training scenarios. Between sessions, students,

---

[2] *Oxford English Dictionary*, Second edition, 1989, Oxford University Press.
[3] See http://www.ukhaptics.co.uk [last access 05/12/2007].

and teaching staff, can review student progress and reflect on feedback provided by the system. Access to the Clinical Skills Trainer is through a desktop environment using a PC and standard mouse/keyboard built with off-the-shelf components while Virtual Veins supports venepuncture training with a haptic environment, both components being provided via a single haptic-enabled workstation.

Virtual Veins

Virtual Veins 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). The system 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 a report is generated which contains 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. The research reported in this paper focuses on skin/vein insertion[4]/retraction angles for performance metrics.

Figure 1 - Venepuncture training using Virtual Veins

Haptic workstation

Virtual Veins uses a Reachin Display[5] to enable the co-location of graphics and haptics to provide a high quality 3D experience. CrystalEyes 3D shutter glasses[6] are worn by the user to provide 3D stereoscopic views, a SpaceMouse[7] is used in the left hand as a 6DOF device to manipulate the virtual arm/hand and a PHANTOM Omni[8] haptic device is used in the right hand to manipulate a virtual catheter, needle or syringe (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 of a virtual arm.

---

[4] Insertion and penetration are used interchangeably in this paper.

[5] See http://www.reachin.se/products/Reachindisplay/ [last access 05/12/2007].

[6] See http://www.reald-corporate.com/scientific/ [last access 05/12/2007].

[7] See http://www.3dconnexion.com/ [last access 05/12/2007].

[8] See http://www.sensable.com [last access 05/12/2007].

Evaluation study

A user study of the Clinical Skills Trainer and Virtual Veins was conducted. Ethics approval was obtained from the Ethics Committee of the Department of Computer Science, Durham University. A within-subject design was used for the study with all the participants receiving the same treatment. Seven participants were all 3rd-year medical students at the Chantler Clinical Skills Centre, Kings College London[9]. 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, 4 participants used computers for work activities and 4 participants used computers for entertainment. None of the participants reported frequent computer game usage. This was the first deployment of Virtual Veins at the Chantler Clinical Skills Centre therefore the participants had no prior experience of Virtual Veins.

The evaluation was carried out in an empty room where Virtual Veins was deployed for the single day of evaluation. The evaluator sat behind the participant with a video camera on a tripod to record video and audio from the evaluations. A modified think aloud verbal protocol called cooperative evaluation was used to elicit usability issues from the participants.

Cooperative evaluation [13] 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. Subjects may also 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 [6] 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 as either *low* or *high* quality. *Low quality* issues 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* issues 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.

Evaluation procedure

Before entering the evaluation room, subjects were given an eligibility test. The evaluation involved use of shutter glasses and required 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. Potential participants were asked an explicit question on any history of epilepsy or photosensitive seizure. This was also repeated in a consent form for the study. A Frisby Stereotest[10] 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

---

[9] See http://www.kcl.ac.uk/schools/medicine/learning/clinicalskills/ [last access 05/12/2007].

[10] See http://www.frisbystereotest.co.uk/ [last access 05/12/2007].

about to attempt. Participants were to (i) login to the Clinical Skills Trainer, (ii) perform venepuncture in a Virtual Veins practice session, (iii) review the session report, (iv) perform venepuncture in a test session, (v) review the session report and (vi) logout of the Clinical Skills Trainer. The use of a think aloud verbal protocol was explained, and this was also on the information sheet, as was the cooperative nature of the session.

When using Virtual Veins, a practice session differs from a test session in two main ways. Firstly, in practice sessions, real-time feedback on needle angle and bevel are provided to the user. Secondly, 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. These features are not available in testing sessions.

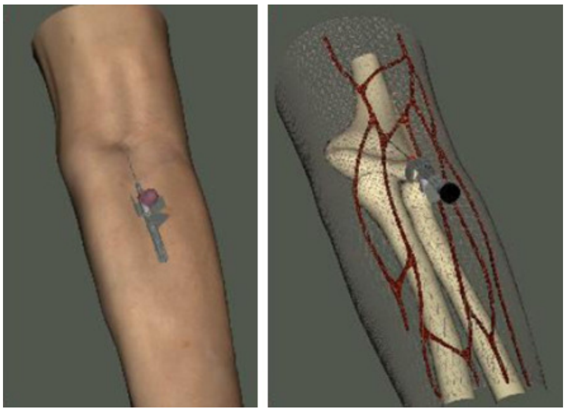

Figure 2 - Virtual arm with and without skin layers

After a practice or test session a report on the users performance using Virtual Veins is generated in the Clinical Skills Trainer. The report allows the user to reflect on their performance after a practice session and allows the user, and teaching staff, to review student performance after testing.

After completing the test session and logging out of the Clinical Skills Trainer, 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.

Results

The average time for an evaluation session was 12 minutes and 15 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. In this section, the results of the usability study will be presented in the following categories: performance metrics from skin/vein penetration/retraction, subjective questionnaire results and usability analysis from the cooperative evaluation. Statistical analysis was performed using SPSS 14.0 for Windows, release 14.0.2.

Performance metrics

Virtual Veins collects data on the first skin penetration and first vein penetration during training sessions. However skin/vein pairs can not be directly compared as only the first penetration event is logged and the first vein penetration may have been from a second or later skin penetration. Figure 3 shows average penetration/retraction angles for skin and vein in the practice and test tasks. Typical good practice for venepuncture angles are ~30° [9] but can vary from between 15° and 30° [14]. Both the average practice and test results are within this range. A small angle difference between penetration and retraction can indicate minimal trauma to the entry/exit area. Figure 4 shows the average absolute difference for penetration and retraction angles. This difference was on average below four degree for participants for both practice and test sessions and therefore minor.

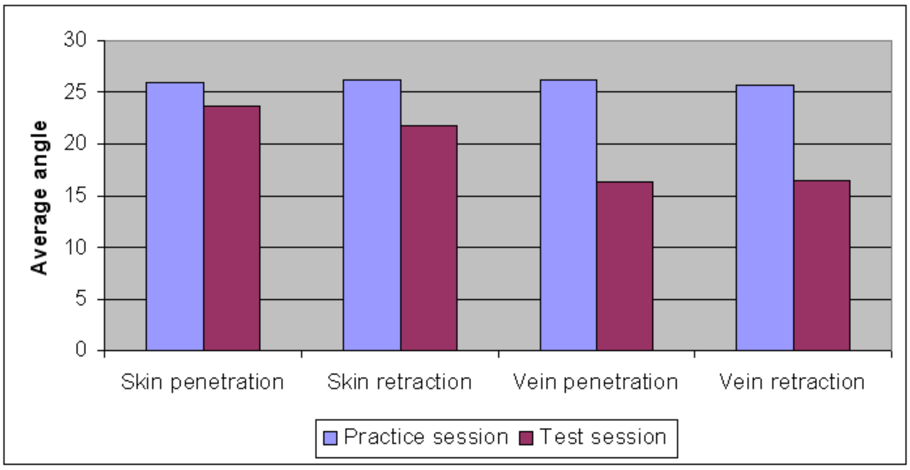

Figure 3 - Average angles (in degrees) of penetration and retraction during practice and test sessions

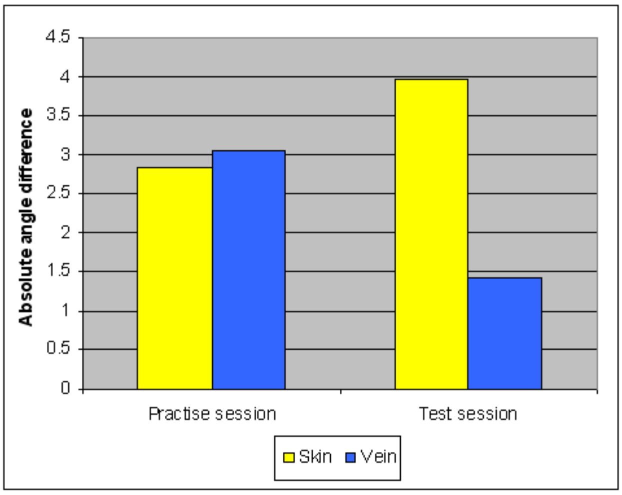

Figure 4 - Angle difference (in degrees) between penetration and retraction

Pearson's r correlation (2-tailed) was performed on the needle penetration/retraction angles and strong correlations were found between skin and vein parings over practice and test sessions; practice skin penetration/retraction (0.860, $p<0.05$), practice vein penetration/retraction (0.990, $p<0.001$), test skin penetration/retraction (0.870, $p<0.05$) and test vein penetration/retraction (0.979, $p<0.005$). This is encouraging as it indicates realistic performance by the participants by not varying the needles position between penetration and retraction. This is good practice for venepuncture and indicative of the realism of the Virtual Veins experience.

Questionnaires

Participants completed a post-session questionnaire including ten 7-point Likert scale questions (see Table 1). The Likert scale questions in Table 1 have been ranked in descending order by average score. The responses were all very positive with the average over the ten questions being 5.58. It is interesting to note that the best two scores (Q1 and Q2) are from use of the desktop/2D mouse component of the system. Two of the lowest scores (Q8 and Q9) are from use of the haptic workstation. This may be due to the novelty of the haptic workstation for new users. The participant's perception of system usefulness (Q6) was high (5.33) as was the perceived realism of the 3D arm image (Q5, 5.57).

Table 1- Likert scale questionnaire results

Question (7 point Likert scale)	**Average**	**Standard deviation**
Q1. Please rate the level of difficulty in selecting a practice or testing session *(1 Difficult - 7 Easy)*	6.86	0.38
Q2. Please rate the level of difficulty in logging into Virtual Veins *(1 Difficult - 7 Easy)*	6.57	0.79
Q3. How attentive/focused did you feel during the training session? *(1 Not focused - 7 Totally focused)*	6.29	0.49
Q4. How comfortable were you in viewing the 3D image? *(1 Uncomfortable - 7 Very comfortable)*	6	0.82
Q5. How realistic did you find the image of the virtual human arm? *(1 Unrealistic - 7 Very realistic)*	5.57	1.13
Q6. Do you think Virtual Veins would be useful in your training? *(1 Not helpful - 7 Essential)*	5.33	1.21
Q7. Please rate the level of difficulty in using the mouse to select Virtual Vein options *(1 Difficult - 7 Easy)*	5.29	1.89
Q8. How realistic did it feel to conduct the venepuncture procedure? *(1 Unrealistic - 7 Very realistic)*	4.86	1.21
Q9. Please rate the level of difficulty in using the Omni for venepuncture *(1 Difficult - 7 Easy)*	4.83	1.72
Q10. How did you find the feedback provided by the system? *(1 Redundant - 7 Informative)*	4.17	0.75

Spearman's rho correlation (2-tailed) was performed on the questionnaire answers. The strongest correlation was between the ease of use of the Omni device for venepuncture (Q9) and the comfort of viewing the 3D image (Q4) (0.939, p<0.01). This indicates the successful integration of visual and haptic feedback provided by the haptic workstation. The need for high quality 3D renderings is indicated in the correlation between 3D arm realism (Q5) and whether Virtual Veins would be useful for training (Q6) (0.909, p<0.05). A correlation was found between mouse use (Q7) and logging into Virtual Veins (Q2) (0.840, p<0.05). This is not surprising as all the participants had regular use of desktop computers, i.e. familiarity with using a standard mouse and keyboard.

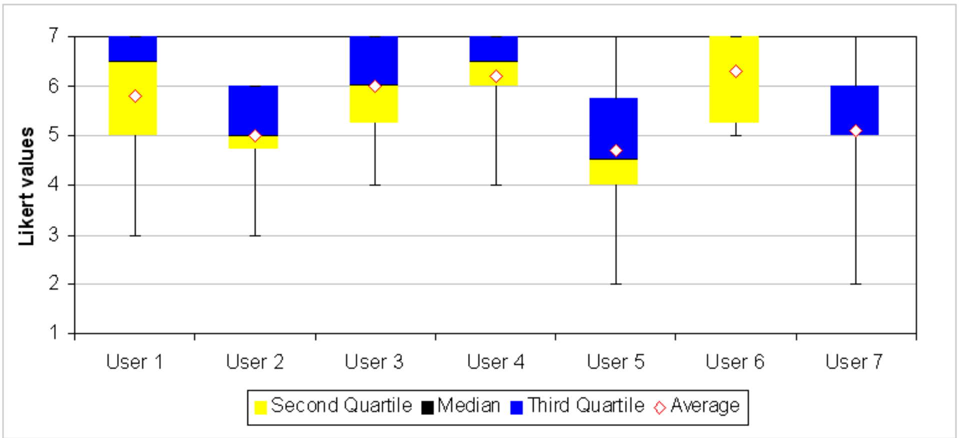

Figure 5 - Likert questionnaire summary by participant

The questionnaires provided a positive view of the subjective participant experience of using the system. On average the participants rated the system highly (see Figure 5).

Usability analysis

Usability issues noted by the evaluator (the first author) during the sessions were augmented with analysis of the video footage post session. 62 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 2 where the user verbalisation/observations have been characterised as *L* for low quality and *H* for high quality. 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 2: 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	0	H	3L	2L,H	L	L	L	15
User2	0	2L	L	2L	0	2L,2H	H	L	0	0	11
User3	L	6L	L	3L	2H	H	L	0	0	0	15
User4	0	L	L	3L,H	L	0	0	0	0	0	7
User5	0	L,H	L	0	0	L	H	L	0	0	6
User6	L	0	L,H	2L	0	0	0	L	0	0	6
User7	0	0	0	0	0	L	0	0	0	L	2
Total	3	13	8	11	4	10	6	4	1	2	62

One of the aims of the study was to determine if cooperative evaluation is an appropriate technique to evaluate haptic interaction. There are two main features identified from the usability analysis. Firstly, the number of entries in Table 2 initially indicated that cooperative evaluation had been very successful in identifying issues during use of the haptic workstation. The three most frequent issues were in the use of the 3D buttons (I2), injection feedback (I4) and the 3D interface (I6). However on review of the recorded verbalisations, none of the 34 issues related to haptic issues as most 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). Similarly to the questionnaire results, this is not surprising as the participant's all used PC 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 virtual environments where the participant and evaluator can share the visual experience. Also the video capture of the session only records the visual component and not the haptic feedback.

Secondly, we are interested in the number of *issues per user* collected during the session. The users with the most issues with the system were *User1* and *User3*. However, both these participants had high average scores in the questionnaires (see Figure 5). This may be indicative that these participants were more comfortable with the think aloud protocol, thus generating more verbalisations.

All but *User7* expressed at least one high quality issue. However, in general these issues are spread over all the usability categories so are not indicative of any specific problems with the system. Only one issue had three high values, namely I7 (*3D arm realism*). On review of the verbalisation for I7 it was found that two of the high quality issues where about the visual appearance of the virtual arm. These participants rated the realism in the questionnaire as 6 and 5 respectively thus is not helpful for haptic analysis. The third high quality issue was one of the only haptic-based comments from the evaluations namely that when inserting a needle that the participant felt they were "pushing in with some force and then it just goes in. I mean in real life it's not like that". This participant rated the venepuncture experience 3 in the questionnaire so indicates a good match between actual and reported experience. Overall there were a minimum number of high quality verbalisations. Further analysis here is limited due to the small sample size.

The participants are ordered chronologically in Table 2. It is interesting to note that as the evaluation sessions went on, fewer usability issues were identified by the participants. Cooperative evaluation allows the evaluator to actively discuss the evaluation as an in-session interview. 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 earlier evaluations into the later sessions. It is possible that in the later sessions, the evaluator pre-empted usability issues, which then did not show in the usability analysis. A review of the video and audio transcripts is discussed elsewhere [15].

Conclusions

In order to justify deployment costs and any pedagogical aims, clinical skill training systems require appropriate evaluation. When haptic-based systems are evaluated, traditional evaluation methods may not be suitable to capture an accurate view of a user's experience.

This paper has described the evaluation of a commercial haptic-based clinical skills training system. A user study was performed with 3rd-year medical students in order to gather both qualitative and quantitative data to support usability analysis. The use of data logging, user questionnaires and a modified think-aloud verbal protocol was investigated. The verbalisations from the evaluation session were not as descriptive as hoped. With haptic-based virtual environment evaluation the evaluator does not share the haptic feedback with the participant, only the visual feedback. It was found that this can be problematic in shaping questions during an evaluation session and interpreting the collected data.

One limitation of the user study is the small number of participants. With such a sample size, individual variations in participants will weaken any objective results. However, the usability analysis did elicit a number of useful comments, although not as haptic-oriented as hoped for. Future work involves developing a framework of haptic-oriented verbalisations that can be used to increase the quality of participant responses during haptic-based system evaluation and completing the evaluations with a larger number of participants.

Acknowledgements

The authors would like to thank Dr Sally Richardson (Kings College London), the student participants from the Chantler Clinical Skills Centre, Gary Todd (UK Haptics), Jonathan Lee (REDSS, Durham University) and John Frisby (www.frisbystereotest.co.uk). This work was funded in part by a Durham University *Initiating Business Collaboration* Grant and UK Haptics.

References

[1] Satava, R. M., and Jones, S. B. 2002. Medical applications of virtual environments. In Handbook of Virtual Environments: Design, Implementation and Applications, K. M. Stanney, Ed. Lawrence Erlbaum Associates, ch. 47, 937–957.

[2] Raij, A., Johnsen, K., Dickerson, R., Lok, B., Cohen, M., Bernard, T., Oxendine, C., Wagner, P., and Lind, D. S. 2006. Interpersonal scenarios: Virtual ≈ real? In IEEE Virtual Reality Conference (VR 2006), IEEE Computer Society, 59–66.

[3] 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.

[4] 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.

[5] 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.

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

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

[8] Roberge, R. J. 2004. Venodilatation techniques to enhance venepuncture and intravenous cannulation. Journal of Emergency Medicine **27** (1): 69–73. Elsevier.

[9] Jordan, R. 2005. Policy and procedure for venepuncture. Northgate and Prudhoe NHS Trust, July. CP 18.

[10] Hewitt, J., and Roberts, C. M. 2004. Evaluating the practice outcomes of a venepuncture and intravenous cannulation. Nurse Education in Practise **4**: 77–79. Elsevier.

[11] 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.

[12] 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.

[13] 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 Ltd.

[14] Toson, R. 2005. Clinical guidelines for venepuncture. South Gloucestershire NHS Primary Care Trust, January, 2005.

[15] Smith, S. P. and Todd. 2007. Cooperative evaluation of a haptic-based medical virtual environment. In 4th INTUITION International Conference on Virtual Reality and Virtual Environments proceedings. Institute of Communication and Computer Systems of the National Technical University of Athens.