

Formative Evaluation of a Virtual Reality Telerehabilitation System for the Lower Extremity

Elizabeth WHITWORTH¹, Jeffrey A. LEWIS², Rares BOIAN³, Marilyn TREMAINE¹,
Grigore BURDEA³ and Judith E. DEUTSCH²

Corresponding Author: deutsch@umdnj.edu

¹Dept Of Information Systems, NJIT

²RIVERS Lab, Doctoral Program in Physical Therapy, UMDNJ

³VR LAB, CAIP, Rutgers

Abstract

Usability studies are an essential and iterative component of technology development and ease its transfer from the laboratory to the clinic. Although such studies are standard methodology in today's graphical user interface/windows world of applications, new technologies strain the current methods and require experimentation in what can be transferred to the new user interaction situations. In this paper five integrated interfaces with three simultaneous users and a remote monitoring setup are evaluated via a usability study that uses both traditional and new methods to assess the viability of the interface design. This formative evaluation was carried out by an expert domain user on the Rutgers Ankle Rehabilitation System (RARS) and its telerehabilitation sub-system. The tester and developer's observations, along with the session videotapes and therapist-user questionnaires were triangulated to identify user problems and suggest design changes that will increase the usability of the system. Changes that resulted from the analysis are described and recommendations for future usability studies are proposed. The validity of usability studies in the development and refinement of rehabilitation technology is highlighted.

Keywords

Virtual reality, telerehabilitation, usability assessment, Rutgers Ankle, remote monitoring

INTRODUCTION

The design of any rehabilitation technology requires an assessment and understanding of how the technology will be used and deployed as well as the interaction of individuals from different disciplines. Evaluating the technology's ease of use facilitates its transfer from the laboratory into the clinic. This is achieved by the application of a set of specific methodologies called usability engineering. A major usability engineering methodology used to evaluate how easily the user understands and employs the application is called a usability study. The iterative application of usability studies to the design process serves as an integral component to the generation of easier-to-use interfaces [1,2].

Standard methods for conducting usability studies are well-documented and described in basic human-computer interaction textbooks [3]. Typical usability studies involve giving a real user a real task to perform while videotaping the process [4]. Literature on usability study operation has focused on the evaluation of web page usage [5] and the evaluation of single-user interfaces [6]. Assessing usability of interfaces involving multiple users, remote connections and integrated processes still lacks standard approaches.

This paper describes how usability testing and software design iteration were performed collaboratively by a group of engineers and clinician scientists. Key issues in this description are the adaptation of standard usability methodology to assess the ease of use of a virtual reality and telerehabilitation system. Findings from a formative evaluation, an early stage usability evaluation, on an expert domain user and the modifications to the system are reported here.

BRIEF DESCRIPTION AND APPLICATION OF THE TECHNOLOGY

The Rutgers Ankle Rehabilitation System (RARS), consists of local hardware and software components, as well as a remote monitoring sub-system. The key hardware element is a small parallel-kinematics robot in which the patient places his foot. The main local software component is a library of virtual reality simulations written in

WorldToolKit [7]. The remote monitoring station is a standard PC with a video, audio and Ethernet connection with a local PC driving the simulation seen by the patient. The RARS is outfitted with a pan-tilt-zoom (PTZ) camera and software used for remote monitoring. The remote monitoring screen runs a Java 3D [8] simplified graphics applet, which follows in real time the local VR exercise. Detailed descriptions of the RARS and telerehabilitation sub-system can be found in the following references [9,10,11,12,13]. The system is currently designed for use in sitting for individuals who have lower extremity dysfunction that interferes with functional mobility. It has been used with individuals who have had ankle sprains or fractures [9,14], and in the rehabilitation of individuals post-stroke to address neuromuscular deficits that have interfered with ambulation [11,15]. The local RARS system elements are presented in Figure 1.

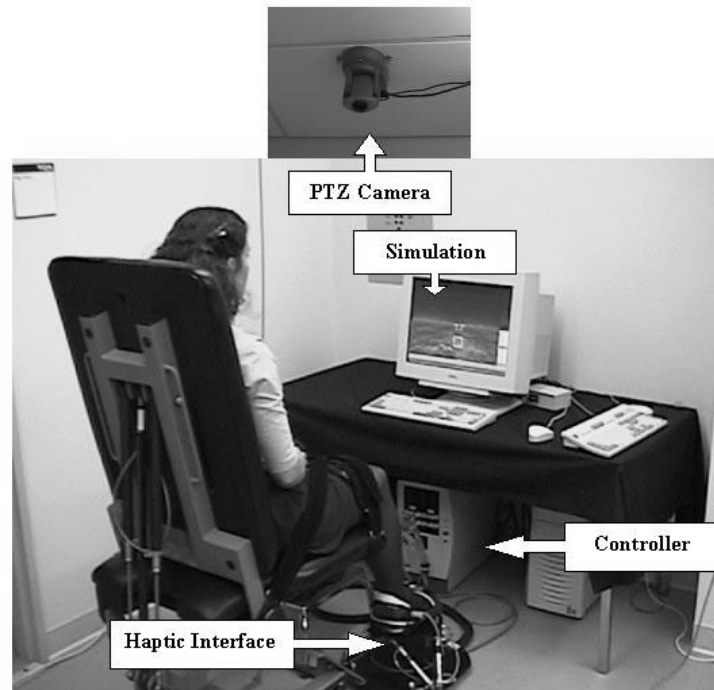


Figure 1. RARS System Overview (patient's station). © UMDNJ, Rutgers 2003

The system was developed collaboratively between engineers (GB, MG, RB, JAL) and a clinician scientist (JED). It was designed for use by physical therapists to allow them to apply principles of exercise in designing a treatment session. Thus, with basic knowledge of the system and the parameters that can be modified to create exercises of varying difficulty levels, any physical therapist can use the system with a variety of patients. The software evaluated in this study consists of four interfaces used sequentially; to set a patient baseline (Fig 2a), configure an exercise session (Fig 2b), supervise and modify the exercise task in the virtual environment (Fig 2c), and monitor the patient remotely (Fig 2d). The figures of each interface are listed below as they appeared in the therapist's training manual, which was used and also evaluated in this study.

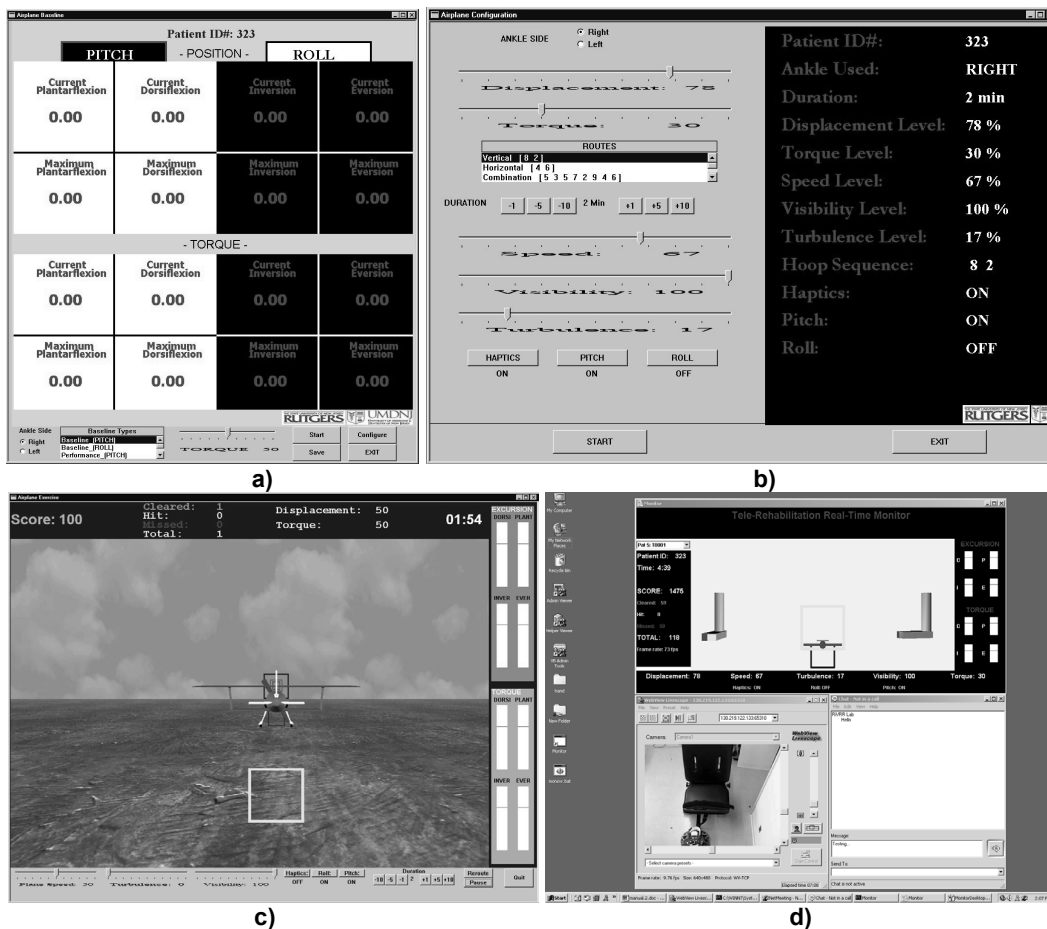


Figure 2. The RARS software components: a) Baseline screen, b) Configuration screen, c) VR Exercise Simulation seen by the patient and local therapist, d) screen view of the Remote Monitoring station. © UMDNJ, Rutgers 2003.

TESTING USABILITY

Specific Challenges Faced in Testing Usability

Typically a usability study involves a user who is asked to interact with software and perform a set of tasks that they would normally carry out using the system [1]. A record is made of the user interaction either via tester coding, audio recording, transaction logging or videotaping. Often, more than one such data collection is used. In the present study the same procedures were used except that testing was performed simultaneously or in sequence on five integrated user interfaces (a training manual, baseline testing, exercise configuration, exercise testing and remote monitoring) and with multiple users (a local and a remote therapist and a patient). Normally, each of the interfaces (except for the user manual) is tested in an unrelated usability study. However, given the manner in which rehabilitation is practiced, it was important to maintain the tight coupling between each of the interfaces and the patient's therapy. We therefore had to design a set of integrated tasks that related each of the user interfaces to each other. This created a problem of separating the effects of one interface study on the next interface evaluation. Because of this, methods were needed to sort out the interactions between the integrated user interfaces.

To run a realistic usability study, one needs to use real patients. Normally, usability engineers conduct usability studies, but such individuals would not have experience in patient therapy, and so, would not be able to make corrective suggestions if the therapist-user in the study generated an error due to a system interface misunderstanding. Having a therapist and designer of the application present solved this problem but is not accepted in usability evaluations because of the possible unintended effect on the user in the study.

A challenge, also unique to this study, was how to evaluate a remote monitoring interface. Specifically, it was important to ascertain whether the remote therapist had enough information about the patient to accurately direct the session. Part of the difficulty in evaluating the effectiveness and usefulness of the patient information provided through remote monitoring was that the user was new to a complex system. Here, the effect of being a novice RARS user is confounded with the usefulness of the remote information provided. The data collected on this user may or may not reflect the experiences of more experienced users.

A final challenge was the number of users involved in the interaction. The local sessions have two users: a therapist-user and a patient-user. The remote monitoring involves three users, the therapist-user at a remote location, a local therapist (or therapist assistant) and the patient-user. Although multi-user evaluation sessions encourage users to talk to each other about the problems they are having with the interface, data is lost as the number of users increase, since one user may solve a problem before it could be observed with the other two users [16].

Methods Used in the Present Study

A training manual was written by selected system co-designers (JAL and JED), in which the RARS system and its telerehabilitation component are described. The manual was designed as a teaching tool to be used over two sessions. In the first session the therapist-user completed the following sections: System Overview, Haptic Interface, Patient Set Up, Session Configuration, and Exercise Training, by reading the manual and practicing on the system. This was followed by an opportunity to practice using the system by setting up and running exercises for an experienced “patient” (JAL). At the end of the session the user completed two questionnaires, a standard usability questionnaire [17], and an instrument specific questionnaire that tested knowledge about the system.

In the second session, the therapist-user reviewed the training manual for fifteen minutes and was allowed to ask questions of the experimenter about the manual content. The therapist-user then proceeded to position and test an experienced patient. Guided by the manual, the therapist-user then followed the tutorial on the remote monitoring system and had an opportunity to direct a session remotely. A usability questionnaire was administered for the remote monitoring component of the system.

During both sessions a usability tester (EW) interacted with the patient while collecting video data from the screens and making notes of her observations; one of the system developers (JED) remotely monitored the sessions, also taking notes and serving as resource to the tester if questions about the system and its use arose during testing. Throughout both sessions, the user was required to talk aloud about what she was experiencing within the system. Talking aloud is a well-established method in usability testing that gives experimenters a better understanding of the user’s thought processes. This method allows the capture of cognitive difficulties that occur in real time that may be forgotten, or even pass unnoticed by the user.

The videotape was reviewed for a gestalt impression and data from the tester’s notes, developer’s notes, and user questionnaires were triangulated to identify the strengths and limitations of the five integrated user interfaces.

How This Study’s Methods Differ From Existing Usability Studies

Although common usability methods were used in this study, as indicated earlier, the study involved several difficulties because of its multi-user, multi-interface, and remote testing issues. The common methods were the use of videotaping, experimenter coding during the study, and the use of user satisfaction questionnaires following the study. In addition to this approach the following methods were applied: First the entire set of user studies was embedded into a single system manual, which was used as a guide for the therapist-user.

Second, the non-remote studies were run with one of the interface designers (JED) remotely monitoring the subject. Although remote monitoring has been conducted in other usability work, this is usually done when no experimenter is in the room. A key problem with application designers being present during usability studies is the possibility for their non-verbal and sub-verbal behavior to influence the user being tested. Nevertheless, in a study with the complexity of the one conducted, it is useful to have the advice of the therapist-designer, who has a deeper understanding of the therapy than the usability engineer. Having the remote observation capability

permitted us to have the therapist-designer present without unduly influencing the subject in the study. It also gave us the added advantage of having a second, remote note-taker.

Third, a set of true-false questionnaires, to assess how much the therapist in the study understood about the operations of the system, was administered to the user. Although a verbal report of the problems with the system was recorded on video and through experimenter coding, this report did not ascertain what problems were due to a misunderstanding of the interface by the user or the cognitive complexity of the task at hand, i.e., handling a patient, and in the remote session, an assistant therapist. Thus, although cognitive comprehension questionnaires are used in usability work, this set was used specifically to evaluate the depth of problems found in the video and coded data capture.

Finally, the therapist-user was required to explain the usage of the system to both the patient and to the therapist assistant in the study. This is the most unique of the new approaches that were taken in this study. Having a user explain what this person has just learned can give a usability engineer much more detail about what is misunderstood, omitted, or misrepresented by the user. The data captured with this technique was especially useful for the evaluation of the remote session. Because the therapist-user not only explained to the therapist assistant what steps to take in adjusting therapy, but also gave reasons for the system adjustments, it could be determined how well the therapist-user understood the patient-user problems via the remote monitoring.

RESULTS

User Questionnaires

The therapist-user generally had positive ratings of the RARS system rating ‘comfort of use’, ‘ease of learning’, ‘ability to effectively complete work’, ‘organization of information on the system’, ‘enjoyment and pleasantness of interfaces’ and ‘overall satisfaction,’ receiving scores between 6 and 7, with 7 being the highest rating on the scale. Items that received lower ratings were the ‘ability to complete work efficiently and quickly’, ‘the systems ability to signal the user when they have made an error’, the ‘information provided for the system being easy to understand,’ which received scores between 3 and 4.

The therapist-user’s response to open-ended questions about the positive aspects of the system included “enjoyment of the exercises”, “flexibility of the system that allowed customizing the exercises for the patient” and “the ease with which [she] learned the system”. Accuracy for responding to the RARS questionnaire score was 87%. The results for the telemonitoring component were slightly lower than those of the RARS as a whole, with 6 being the highest rating. The therapist-user rated the organization of the screen, her liking the system and the functionality of the system as a 5. Positive comments were made about the ‘opportunity to direct treatment from a remote location’ and the ‘ability to monitor the patient from various viewpoints’, including focusing on the ankle. In addition, the therapist stated, “With some clarification, the system is easy to figure out, especially with a prior training session.” Specific comments on the negative aspects of the systems were regarding the time lag of the voice feedback (microphones), the inconsistency of buttons between the local and the remote interfaces and the difficulty of multitasking; especially during remote monitoring where the therapist-user would be running a chat, verbal interactions with the local therapist, and monitoring the patient performance at the same time.

Tester and Developer Observations Resulting in Interface Changes

In this section are described the observations of the tester and developer, as well as patient-user comments made on video. These observations were used to make changes to the therapist training manual and the interfaces. The manual lacked clarity and its construction produced cognitive overload for the user. To address this issue the format was changed reducing the information on each page and labeling and numbering figures to make it easier for the reader to follow sequential information. (See Figure 3a, 4, 5b)

Terminology was inconsistent across four of the interfaces and was not the preferred terminology of a physical therapist. For example ‘Displacement’ was used when the term ‘Range of Motion’ was preferable, being both less ambiguous, and more self-explanatory (Figure 3a, 4, 5ab). Inconsistent placement of the same terms in different

interfaces caused confusion. For example, the term pitch preceded the term roll on two of the screens and then order was reversed. This proved to be distracting and was corrected (Figure 5c).

A problem involving visibility and a high cognitive load, or the amount of information that needed to be remembered, was present in the baseline interface. The original design involved four buttons, “Start,” “Save,” “Configure,” and “Exit” in a square as shown in Figure 3b. This interface design was confusing to the therapist / user in that it was not clear what they were supposed to do within the interface, or what would happen once they pressed the buttons.

Making procedures more apparent within the interface by reorganizing the presentation of the functions solved this command structure problem. The new design involved three large buttons “Start,” “Save,” and “Next” in a line, and a small X button to quit the screen (see Figure 3b). The order and direction of the buttons in the revised interface correspond to the actions that the therapist-user is supposed to take to complete a patient baseline. The “Configure” button, which would take the user to the next screen to configure an exercise, was changed to “Next” being an easily understood term, and a more obvious indicator that the user would be taken to a new screen. The button that would exit the user from the screen was made smaller and separated from the other buttons, making it less likely that users would click on the exit button by mistake.

Another problem involved the toggle switch in the exercise screen, where the system started off paused so that the patients could orient themselves with the plane they were supposed to pilot, and be ready for the exercise. In the initial design, the user would have to press the pause button again in order to start the system, which is not very intuitive. In order to make the process more intuitive, the previously large quit button was reduced to a small X at the bottom right of the screen, and the space created used to include a large start button (see Figure 5d).

Observations of the remote monitoring component of RARS highlighted several limitations. There was a reliability issue with the audio communication. The order of the elements on the remote screen did not match that on the local interfaces making it more difficult for the user to learn. The therapist-user did not consistently read the messages presented on the chat window, which is attributed to having to attend to three elements; the VR exercise simulation, chat window, and camera view of the patient. As a result of these difficulties, the terms used on the remote monitoring station and their order were altered.

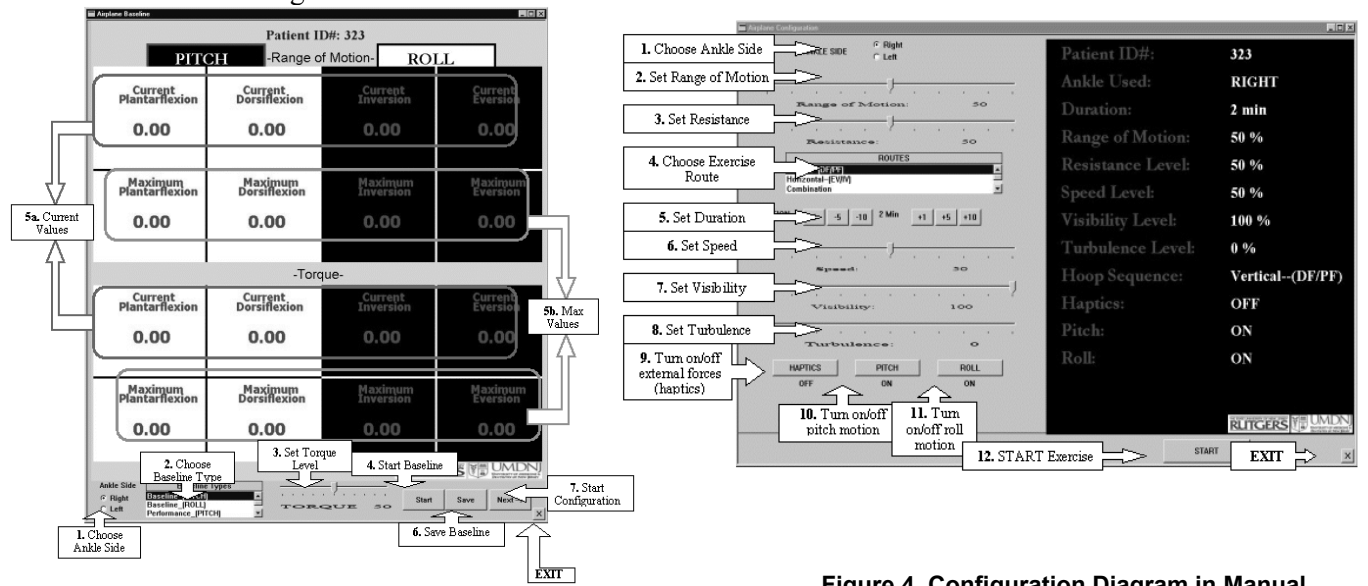


Figure 4. Configuration Diagram in Manual ©UMDNJ, Rutgers 2003

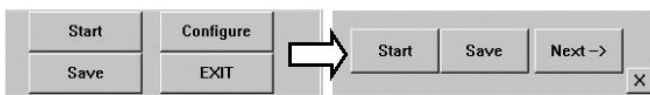


Figure 3. a) (above) Baseline Diagram in Manual b) (below) Command Modification ©UMDNJ, Rutgers 2003

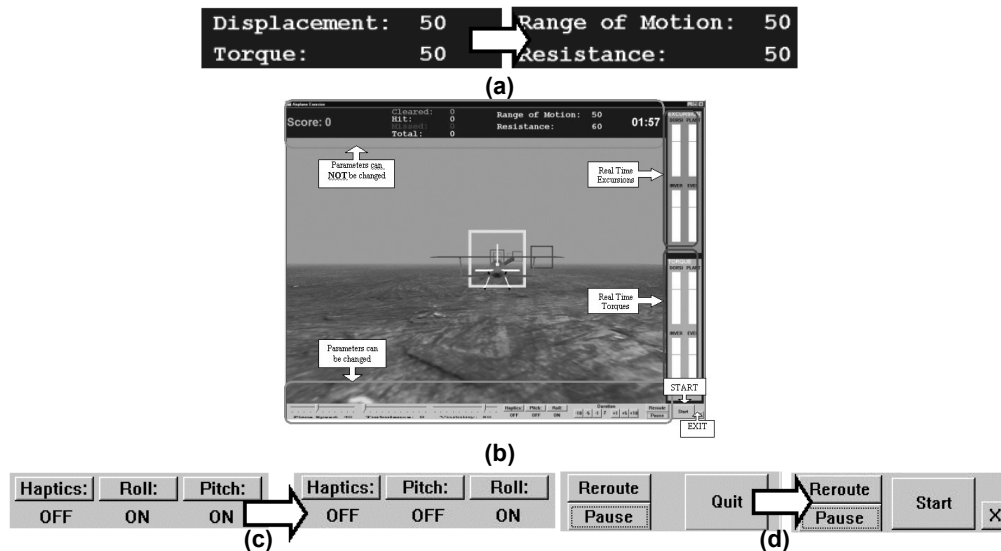


Figure 5. Interface changes following the study: a) Terminology Modification, b) Simulation Diagram, c) Ordering Modification, d) Command Modification © UMDNJ, Rutgers 2003

DISCUSSION

A formative evaluation of the RARS was conducted adapting existing usability study methodology. The findings from the tester's notes, videos, user questionnaires and developer's notes found key problems, which required modifications. Changes were made to the user interfaces to be consistent with the therapist's language and use [3]. Small order changes on the screen were made to reduce the difficulty the user had in finding information, therefore consistency in lay out was implemented [18]. The command structure was altered to better reflect the clinical decision making process. The toggle switch which started the machine in a state that the user did not specify and then required a toggle switch to alternate between states was also modified [19]. The most substantial changes were made to the therapist-training manual to reduce the cognitive load experienced by the reader.

While all these changes will serve to refine and ease the use of the system, it is noteworthy that the therapist rated her experience as highly satisfying, with the 'comfort of use', 'ease of learning', 'enjoyment of the interfaces' and 'overall satisfaction. The aspects of the system that received lower rating were related to the training manual and its use. Importantly, the therapist was able to correctly apply principles of exercise to create an exercise session for an experience patient-user and her knowledge retention of the system was 87% after using the system only once. It appears that the telerehabilitation aspect of the system will require additional training.

Limitations of the Current Study

This study was a pilot study in which usability was tested using a single expert domain user. In addition, every feature of the system was not tested exhaustively because of the need to create a realistic testing scenario. While video information was gathered on the therapist-user experience from the remote monitoring point of view, the resources did not exist to also collect video data from the point of view of the local therapist and patient-user from the same interaction. Collecting data from the local therapist would involve collecting video data and having a second camera person present on the local therapist's side of the interaction. Collecting data from the patient-user within the same interaction would prove even more difficult. As described earlier, it is important to hear a user talk through their ideas. However with both the patient-user and the local therapist in the same room, having both talk aloud would be cumbersome, if not entirely unfeasible, and as a result, important real-time information would be lost. Lack of time and other resources, as well as a lack in trained users to serve in secondary roles in the study, e.g., the assistant therapist, restricted some of the control of the study and the information that could be captured

CONCLUSIONS AND FUTURE WORK

The five integrated interfaces of the RARS and its telerehabilitation component were evaluated by an expert domain user. Based on this therapist's report, the RARS and telerehabilitation sub-system were easy to use and consistent with her style of therapy practice. Many modifications were prompted by this formative evaluation

validating usability studies as an integral component of design and implementation of rehabilitation technology. In addition, the modifications made to traditional usability methods aided in capturing data necessary for the interface evaluation and helped to explain the meaning of data that was confounded by the complexity of the study.

Ten therapists from different practice areas and varying levels of experience will be tested in a subsequent study. In anticipation of this study the training manual has been modified for clarity. Volunteer “patients” will be naïve system users, who will be asked to complete questionnaires about their experience with the system and the therapist. The immediate planned study will be on novice therapist-users, however, additional testing involving experienced users would also be useful. For example, testing experienced users as local therapists as well as remote therapists would allow us to evaluate if the information provided remotely was deficient in any way.

ACKNOWLEDGEMENTS

Research reported here was supported by grants from the National Science Foundation (BES-0201687) and the New Jersey Commission on Science and Technology (R&D Excellence Grant).

REFERENCES

- [1] J. Nielsen, *Usability Engineering*, New York: Academic Press, Inc, 1993.
- [2] D. J. Mayhew, *Principles and Guidelines in Software User Interface Design*, Englewood Cliffs: Prentice-Hall, 1992.
- [3] J. Preece, Y. Rogers and H. Sharp, *Interaction Design: Beyond Human-Computer Interaction*, New York: John Wiley & Sons, 2002.
- [4] M. Mantei and T. J. Teorey, “Cost/Benefit Analysis for Incorporating Human Factors in the Software Lifecycle,” *Communications of the ACM*, 31, 4, April 1988, 428-439. Paper republished in *Software Engineering*. B. Boehm (Ed.) Los Angeles: The Computer Society Press, 1989.
- [5] J. Cato, *User-Centered Web Design*, London: Addison-Wesley, 2001.
- [6] D. J. Mayhew, *The Usability Engineering Lifecycle: A Practitioner’s Handbook for User Interface Design*, San Francisco: Morgan Kaufman Publishers, 2002.
- [7] Sense8 Co., *WorldToolKit Release 8 Technical Overview*, February 1998.
- [8] H. Sowizral, K. Rothforth and M. Deering, *Java3d API Specification*, AddisonWesley, 1998.
- [9] M. Girone, G. Burdea, M. Bouzit, V. G. Popescu and J. Deutsch, "Orthopedic Rehabilitation using the 'Rutgers Ankle' Interface," in *Proceedings of Virtual Reality Meets Medicine 2000*, January 2000, pp. 89-95.
- [10] M. Girone, G. Burdea, M. Bouzit, V. G. Popescu, and J. Deutsch, "A Stewart Platform-based System for Ankle Telerehabilitation," *Autonomous Robots*, Vol. 10, pp. 203-212, March 2001.
- [11] R. F. Boian, C. S. Lee, J. E. Deutsch, G. Burdea and J. A. Lewis, "Virtual Reality-based System for Ankle Rehabilitation Post Stroke", in *Proceedings of 1st International Workshop on Virtual Reality Rehabilitation (Mental Health, Neurological, Physical, Vocational)*, Lausanne, Switzerland, November 2002, pp. 77-86.
- [12] R. F. Boian, J. E. Deutsch, C. S. Lee, G. C. Burdea and J. A. Lewis, “Haptic Effects for Virtual Reality-based Post-Stroke Rehabilitation”, in *Proceedings of the Eleventh Symposium on Haptic Interfaces For Virtual Environment And Teleoperator Systems*, Los Angeles, CA, March 2003, pp. 247-253.
- [13] J. A. Lewis, R. F. Boian, G. C. Burdea and J. E. Deutsch, "Real-time Web-based Telerehabilitation Monitoring", in *Proceedings of Medicine Meets Virtual Reality 11*, Newport Beach, CA, January 2003, pp. 190-192.
- [14] J. E. Deutsch, J. Latonio, G. Burdea, and R. Boian, “Rehabilitation of Musculoskeletal Injuries Using the Rutgers Ankle Haptic Interface: Three Case Reports,” in *Proceedings of Eurohaptics*, 2001, pp.11-16.
- [15] J. E. Deutsch, J. Latonio, G. Burdea and R. Boian, "Post-Stroke Rehabilitation with the Rutgers Ankle System - A case study," *Presence*, MIT Press, vol. 10(4), pp. 416-430, August 2001.
- [16] P. Sullivan, “Multiple Methods and the Usability of Interface Prototypes: The Complementarity of Laboratory and Focus Groups,” in *Proceedings of the ACM Conference on Systems Documentation*, New York, NY, 1991, pp. 106-118.
- [17] IBM Ease of Use Survey, 2002, http://www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/2023.
- [18] R. C. Teitelbaum and R. E. Granda, “The Effects of Positional Constancy on Searching Menus for Information,” in *Proceedings of the ACM CHI’83 Conference on Human Factors in Computing Systems*, New York, NY, 1983, pp. 150-153.
- [19] D. A. Norman, *The Psychology of Everyday Things*. New York: Basic Books, Inc, 1988.