Application of Usability Analysis Techniques to the Design of Rehabilitation Equipment

Meghan Hegarty[1], David Kaber[2], and Edward Grant[1,3]

[1]UNC/NCSU Joint Department of Biomedical Engineering, North Carolina State University; [2]Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University; [3]Department of Electrical and Computer Engineering, North Carolina State University

The rehabilitation device industry stands to benefit from the use of iterative design techniques in the product development process. This paper describes a three-pronged evaluation method for the analysis of rehabilitation devices. A case study comparing the usability of a re-designed interface for an isokinetic testing and training system is used to illustrate the technique. Simulations of the current and re-designed interfaces were created, and an informal usability evaluation was conducted. A composite *usability score* was computed by combining both quantitative (i.e., error and HELP reference counts) and qualitative (i.e., SUS score) means of assessment. This allowed for more comprehensive comparison between the current and re-designed interfaces, demonstrating the later to be superior in terms of general usability. Additionally, the nature of errors and HELP references was recorded and explored in order to ensure complete coverage of potential problems. Finally, a questionnaire targeting user opinion on newly added features severed to justify their inclusion/exclusion in future versions of the interface. By using this three-pronged evaluation method, we were able to objectively evaluate the re-designed interface in terms of general usability, as well as gain direct feedback for future improvements.

INTRODUCTION

Over 700,000 people suffer a stroke each year in the United States. Of these, two-thirds require some form of rehabilitation (National Institutes of Health, 2007). Currently, the demand for rehabilitation services exceeds the number of physical therapists available (Farmer, 2004). This situation illustrates the need for the development of a more effective means of delivering therapy. Robotic rehabilitation platforms have been developed to fill this niche.

The GENTLE/S platform was the first robotic rehabilitation system to fully embrace the concept of participatory design (Hawkins et al., 2002). The designers initially conducted an evaluation of the users' needs. Existing products and means of delivering therapy were reviewed, and therapists and patients were interviewed. Sketches and informal prototypes were used to explain system concepts. The information gained from this preliminary research was incorporated into a design specification. A rapid prototype of the system was created and a pilot study conducted. The results of this study proved valuable in the final design of the system. Suggestions such as the provision for wheelchair access to the robotic platform, improving the setup process, and reduction in overall size were incorporated into later prototypes (Hawkins et al., 2002).

Building on the work of Hawkins et al. (2002), a more formalized approach to the evaluation of rehabilitation equipment is presented in this paper. A case study is used to illustrate application of this technique to an existing system.

Case Study: Re-Design of an Isokinetic Testing and Training System

The system evaluated in this work is a widely used tool for isokinetic testing and training. Despite its diverse range of functionality, the system has lost market acceptance due to its poorly designed interface. A contextual task analysis was completed on a bimanual isokinetic knee evaluation. Based upon the findings of this analysis, as well as discussions with physical therapists, usability goals for the interface re-design were established.

The decision was made to assess this system from a 'walk-up-and-use' perspective. Therefore, the first usability goal dealt with increasing ease-of-use through improved consistency, error prevention/recovery, and system-state feedback. The second usability goal was aimed at increasing rates of learnability through the formation of a more accurate/complete mental model, increased system-state visibility, and incorporation of appropriate diagrams.

In order to realize these goals, the following major revisions were made: (1) re-ordering of the dialog structure to more closely approximate the user's internal model of the system, (2) inclusion of a 'Control Panel' to promote system-state visibility and allow for manipulation of variables, (3) inclusion of feedback dialogs, (4) inclusion of graphics for clarification, and (5) inclusion of a 'Quit' confirmation dialog. The original and re-designed interfaces were both prototyped using Visual Basic 2005 .NET. Example screenshots of the current and re-designed interface are presented in Figure 1.

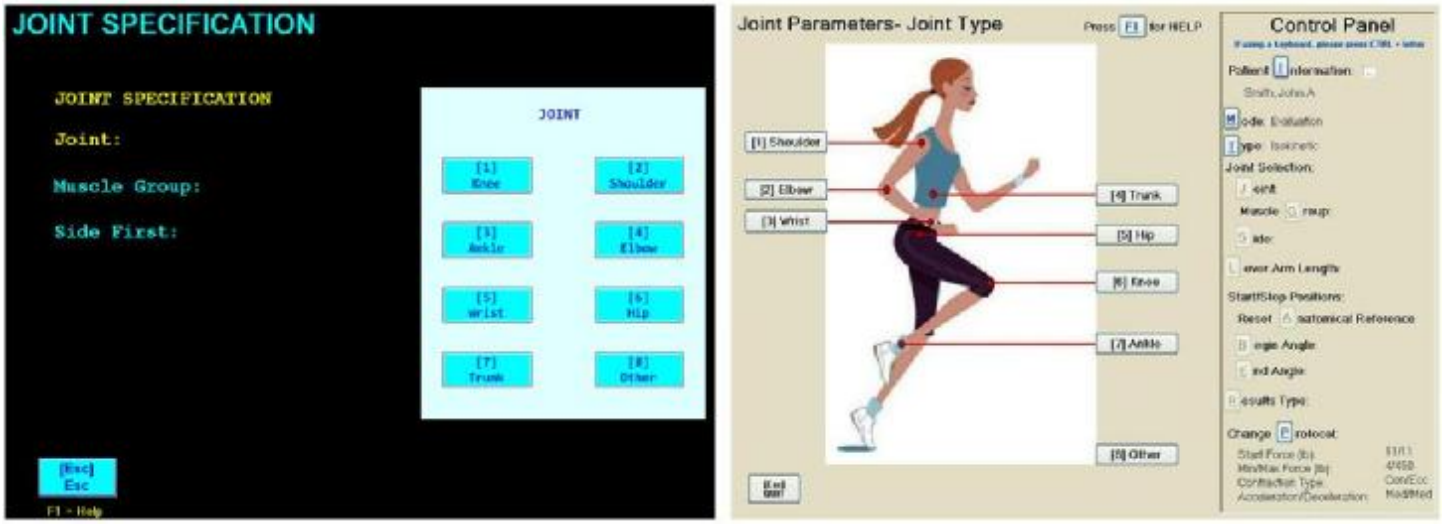

Figure 1: Example screenshot from current (left) and re-designed (right) simulated interface.
(Picture of female jogger from http://lediet.m6.fr/conseils_sport.html?unique_id=f4b134a24d8870c42aa87b042e931a60)

METHODS

Participants

Five graduate students (PhD-level) with knowledge in usability engineering and interface design were recruited for this study. Of these, one was considered to be an expert user and the remaining four were considered to be novice users. None of the participants had been introduced to the system being studied and/or similar equipment as patients. The expert user had worked with the system in other research.

Test Materials

Pre-trial questionnaire. This questionnaire was used to assess experience with the system being studied and/or similar rehabilitation equipment. It was also aimed at determining if the updated dialog more closely matched the user's internal model of the system. The major dialog events were listed in random order, and participants were asked to arrange them in the most logical sequence for the task.

Post-trial questionnaire, Part 1. In order to assess the current and re-designed systems in terms of general usability, a modified version of the System Usability Scale (SUS) developed by Brooke (1996) was used. As is common practice with the SUS, modifications were made to the original questionnaire in order to reflect the nature of this study.

Post-trial questionnaire, Part 2. The second part of the post-trial questionnaire was aimed at determining the influence of all updates on the interaction process. Additionally, any comments/suggestions concerning their helpfulness and/or implementation were sought.

Protocol

Each participant was given 5 min to read through a task description outlining a bimanual isokinetic knee evaluation. The pre-trial questionnaire was also completed at this time. Each participant was then asked to configure the system in the manner specified in the task description. As ease-of-use was the primary goal of this study, no task training with the software was provided in advance of testing. Participants were allowed to refer to an activity flow chart of the procedure throughout the experiment. Three participants completed the task first using a simulation of the current interface, and then using a simulation of the re-designed interface. The order of testing was reversed for the remaining two participants. For each system, 3 min was allotted to complete the task. Time to task completion was measured using a stop watch. The stop watch was started when the participant indicated that (s)he was ready to begin the task, and stopped when the participant reached the 'Results' screen. The selections made by the participant and the number and nature of HELP references were tracked by an experimenter on a form. At the end of 3 min, participants were asked to complete a questionnaire regarding the general usability of the relevant system (i.e., Post-Trial Questionnaire, Part 1). When testing was complete for both systems, a final questionnaire targeting specific system updates was administered (i.e., Post-Trial Questionnaire, Part 2).

Data Analysis

Participant expertise. Participants were classified as a novice, transfer, or expert user depending upon whether they had used the system being evaluated and/or similar equipment as an evaluator. If the participant had used the system and/or similar equipment as a patient, but not an evaluator, then he/she was considered to be a novice user; this selection was offered in order to determine if the participant might be biased towards a particular system.

SUS score. The SUS scores were totaled, and the average and standard deviation computed for each interface.

User observation data. The task completion rate was computed for each interface. The total task time, error count, and HELP reference count were recorded for each participant. These were averaged, and the standard deviation computed. Additionally, the nature of all errors and HELP references was consolidated into a master list.

Usability score. A composite usability score was created by combining both objective (i.e., error and HELP reference counts) and subjective (i.e., SUS score) measures. This holistic approach to assessment represents a more robust form of analysis insofar as statistical methods can be combined with measures of perceived usability. Such practices have recently become more popular and have shown agreement with more traditional approaches (Lee & Kaber, 2008).

Penalties for the number of errors and HELP references made were selected based upon Brooke's SUS scoring system (0 points for an extreme negative score and 4 points for an extreme positive score). Using this scale, a critical error (a total loss of information) was assigned a maximum penalty of -4 points (i.e., counter-balancing an extreme positive score). A non-critical error (not resulting in a total loss of information) was assigned a penalty of -3 points (counter-balancing a moderate positive score). Finally, a HELP reference was assigned a penalty of -2 points (counter-balancing a neutral score).

The usability score was computed for each interface and each usability expert using the following formula:

$$\textit{Usability Score} = 2.5*[(\#\ \textit{Critical Errors}* -4) + (\#\ \textit{Non-Critical Errors}* -3) + (\#\ \textit{References}* -2)] + \textit{SUS Score}$$

These scores were averaged, and the standard deviation computed.

Dialog order. The results of the second question on the pre-trial questionnaire were compiled and an overall dialog order generated. This was accomplished by determining the percentage of participants placing a given procedure before or at a particular step in the dialog process. The procedure was presumed to occur at a given step if the highest percentage of users placed it at or before that step (i.e., after all assigned procedures had been eliminated). If two or more unassigned procedures received identical 'high scores' for a particular step, then the procedure assigned to that step was selected by determining what percentage of users placed a given selection before the competing selections. After an order was determined, the degree of overlap between the participant-generated order and current and re-designed system order was assessed.

Assessment of re-design updates. Responses for each statement in Part 2 of the post-trial questionnaire were averaged, and the standard deviation computed. Additionally, all comments/suggestions were consolidated into a master list.

RESULTS

A learning effect was noted during the course of this experiment. In all cases, the task was completed in a significantly shorter period of time using the second system compared to the first system ($t(4) = 5.13;\ p = 0.01$). The learning effect was attributed to user inexperience. Additionally, because the elapsed time between presentations was minimal and the available interface options were similar, participants may have remembered their previous selections.

General Usability

The re-designed system interface was found to be superior in terms of general usability, as evidenced by significantly higher usability scores ($t(8) = -3.11;\ p = 0.01$). The re-designed system received an average score of 80 ± 13 as compared to the current system, which received an average score of 32 ± 32. This represents a 152% increase in terms of overall usability.

Usability was assessed using both quantitative and qualitative measures. Due to the significant learning effect, it is difficult to compare task times, error counts, and HELP references across systems. However, the task was completed 100% of the time using the re-designed system and only 40% of the time using the current system. The task was only completed successfully when the re-designed interface was presented first. Additionally, fewer errors and HELP references were made using the re-designed system.

In all cases, the re-designed system received a higher SUS score compared to the current system; this difference was significant ($t(8) = -3.43;\ p = 0.01$). The average SUS score for the redesigned system was 85 ± 14, compared to 39 ± 27 for the current system (i.e., a 118% improvement).

Ease-of-Use

Ease-of-use was gauged by comparing task-completion. The task was never successfully completed with the current system when it was presented first. On the other hand, the task was completed 100% of the time with the re-designed system (i.e., the order of presentation did not affect the participant's ability to complete the task with respect to the re-designed system)

Learnability

The re-designed system was also found to be superior in terms of learnability. Learnability was quantified by counting the total number of errors and HELP references made throughout the task. Again, caution must be taken when comparing these items across interfaces because of the observed learning effect. In terms of the total number of errors, two were encountered using the current system versus one with the re-designed system; this difference was not determined to be significant. Of these, the error made with the re-designed system was identical to one of the errors made with the current system (i.e., confusion over which results/feedback format to select). These errors were not deemed critical because all system information was still presented, albeit in a less intuitive format. The other error

made with the current system was deemed critical because it resulted in a system exit and a total loss of information. The particular error would have occurred despite the order of presentation. There were also fewer HELP references made using the re-designed versus current system. A total of two references were made using the re-designed system, compared to four made using the current system (i.e., a 50% reduction); this difference was not determined to be significant.

With respect to the current system, half of the HELP references made were concerned with the meaning of 'Overlay' and 'Continuous' in the 'Feedback/Results Format' screen. This problem is unique to the current system because pictures and explanations are provided in the re-designed system. There was also confusion over how to proceed from a 'Lever Arm' screen (there was no 'Accept' button as there is in other screens; instead the user is instructed to press [Enter]). Again, this is unique to the current system because there is an 'Accept' button provided in the re-designed system. The HELP references made using the re-designed system resulted from confusion in a 'Joint Parameters-Muscle Group Selection' screen. The first question related to the use of the '/' character to represent 'and' versus 'or' (i.e., for the 'Extensor/Flexor' button); this question was considered to be platform-independent because the same convention was used in both versions of the system. Upon examination of the underlying cause of confusion, it was discovered that this is an unconventional use of the '/' character. Based on this finding, it was recommended that the '/' character should be changed to an '&' character in future versions of the system. The other question concerned the proper selection for testing the knee in extension. This question is considered to be platform-dependent because the current system uses the abbreviation 'EXT' to represent the extensor muscle group ('EXT' could also be mistaken for extension by a novice user). Although this confusion likely resulted from the fact that the participant did not have domain expertise in rehabilitation/biomechanics, even an 'expert' user would most likely think about performing an exercise pattern (i.e., extension) versus testing with a particular muscle group (i.e., extensor). The validity of this statement may, however, be the result of the selected task (i.e., a ligament, not a muscle group, was the focus of the test).

User-Opinion Concerning Re-Designed System

Control Panel. This study revealed that the Control Panel was the weakest/least useful of the updates made to the interface. Although the participants agreed that the Control Panel provided excellent feedback (a preference score of 4.6 ± 0.5 out of 5, with 5 indicating the strongest preference) and had the potential to be useful as a navigation aid (a preference score of 3.4 ± 0.9 out of 5), there were concerns raised related to its purpose and prominence. Participants felt they did not need to use the Control Panel to navigate through the system and/or did not notice/understand its purpose. Although the participants did not suggest any ways to improve the Control Panel, re-naming it could help to better impart its function. Additionally, using a different background color and/or border style may help to increase its salience.

Dialog order. In terms of dialog order, this study revealed the new dialog order more closely matched the user's mental model of the system. All participants noticed a difference in the dialog structure, and preferred the order presented in the re-designed system to that presented in the current system (a preference score of 4.6 ± 0.5 out of 5). This is further supported by the fact that the participant-generated dialog order more closely matched that of the re-designed system versus the current system (see Table 1). In fact, correlation to the participant-generated order improved 300% from the current to re-designed system.

Despite this substantial increase in overlap, differences still existed between the re-designed and patient-generated order. The first of these discrepancies suggests that the participants thought of 'evaluation' as being a sub-set of the 'isokinetic' movement pattern (versus 'isokinetic' being an available option when conducting an 'evaluation'). In order to determine if the participant's view of the system was an artifact of the selected task (i.e., the task description instructed the participants to perform an "isokinetic evaluation"), the available options in both screens were examined. Based on this analysis, it did not appear as if either system-view was more accurate. The second discrepancy between the participant-generated and re-designed order suggested that the anatomical reference calibration and start/stop angle setting be performed before the lever arm calibration. The participant-generated order was actually more consistent with the previous step in the dialog process (i.e., setting the joint parameters), as both were concerned with the joint/limb. The lever arm calibration, on the other hand, is unrelated to any other step in the dialog process. Additionally, setting the anatomical reference and start/stop angle was not dependent on the results of the lever arm calibration, so the order could easily be reversed. Future revisions should incorporate this change in order to allow for a more fluid dialog.

Diagrams. All participants agreed that the inclusion of graphics greatly enhanced the interface (a preference score of 5.0 ± 0.0 out of 5). One participant noted that the diagrams made the re-designed system appear "much less intimidating than the original".

Feedback dialogs. All participants agreed that the feedback dialogs aided in system navigation (a preference score of 5.0 ± 0.0 out of 5). Additionally, the participants did not feel that the dialogs were over-used, which was a concern of the researchers. One participant did note during the task, however, that the appearance of the dialog might be too similar to that of an error message.

DISCUSSION

Hawkins et al. (2002) previously identified participatory design as an important element in the process of designing rehabilitation equipment. Extending this work, a detailed

PARTICIPANT-GENERATED	CURRENT	RE-DESIGNED
Patient Information	Mode Selection	*Patient Information*
Evaluation Type Selection	Patient Information	Mode Selection
Mode Selection	Joint Parameters	Evaluation Type Selection
Joint Parameters	Lever Arm Measurement	*Joint Parameters*
Anatomical Reference & Start/Stop Angle	Evaluation Type Selection	Lever Arm Measurement
Lever Arm Measurement	Results/Feedback Format	Anatomical Reference & Start/Stop Angle
Results/Feedback Format	Anatomical Reference & Start/Stop Angle	*Results/Feedback Format*

Table 1: Comparison of dialog order.

three-pronged approach to assessing usability was presented in this paper.

To begin with, usability was evaluated using both quantitative (time-to-task completion, number of errors, number of HELP references, and task completion rate) and qualitative (SUS Score) measures. In order to provide a robust indicator of usability, both objective and subjective measures were combined. This metric allowed the systems to be easily compared in terms of overall usability. In the future, this technique could be extended to include the evaluation of iterative prototypes and/or comparison of a new system to a similar, existing system.

Second, the nature of all errors and HELP references was recorded. This information, which is not often collected in system usability evaluation and design processes, was used to ensure coverage of all potential problems. Additionally, by exploring the nature of user confusion, suggestions for improvements were made.

Finally, a separate questionnaire was used to measure the influence that the new updates had on the interaction process. By quantifying the nature of these updates, the root cause of any improvements and/or failures was ascertained. Additionally, this process served to justify the inclusion/exclusion of certain features and suggest further design improvements.

The inclusion of the interface 'dialog ordering section' (Question 2) in the pre-trial questionnaire also proved valuable in the evaluation process. This section served to capture the user's view of how the setup process should proceed, thus suggesting the most natural dialog order. With regard to the case study presented in this paper, this tool served to uncover a point of inconsistency in the dialog process. The inconsistency was not immediately obvious to the researchers. This technique could be extended to ensure that instructions are worded appropriately and/or that novel features are intuitive to use in other types of systems.

CONCLUSION

In summary, iterative design techniques may serve to greatly enhance the rehabilitation industry by making devices safer and more user-friendly. In order to evaluate such systems, a three-pronged assessment strategy should be used. First, usability should be evaluated using both quantitative and qualitative metrics. Second, the nature of all errors and HELP references should be explored in order to ensure coverage and suggest further interface improvements. Third, new features should be assessed in order to determine their effect on usability. In addition to the results of this analysis, user feedback should be incorporated into future versions of the system.

ACKNOWLEDGEMENTS

Meghan Hegarty's work on this study was supported under a National Science Foundation Graduate Research Fellowship. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

We would like to thank the members of the physical therapy department at North Carolina State University who demonstrated use of the system used in the case study.

REFERENCES

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester & A. L. McClelland (Eds.) *Usability Evaluation in Industry*. London: Taylor and Francis.

Farmer, Juliet (2004, September). Hiring Staff in Private Practice: Practice Makes Perfect [Electronic version]. *PT Magazine*, 12.

Hawkins, P., Smith, J., Alcock, S., Topping, M., Harwin, W., Loureiro, R., Amirabdollahian, F., Brooker, J. Coote, S., Stokes, E., Johnson, G., Mak, P., Collin, C., & Drissen, B (2002). GENTLE/S Project: Design and Ergonomics of a Stroke Rehabilitation System. *Proceedings of the 1st Cambridge Workshop on Universal Access and Assistive Technology (CWUAAT)*; University of Cambridge.

Lee, L. and Kaber, D. (2008). Assessing Interactive System Effectiveness with Usability Heuristics and Markov Models of User Behavior. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 52(16), 1127-1131.

National Institutes of Health (2007, December 11). *Post-Stroke Rehabilitation Facts Sheet*. Retrieved on December 17, 2007, from the World Wide Web: http://www.ninds.nih.gov/isorders/stroke/poststrokerehab.htm.