Usability and Performance: Comparative Evaluation of Surgeon Interfaces in Laser Phonomicrosurgery

Giacinto Barresi[1], Nikhil Deshpande[1], Leonardo S. Mattos[1],
Luca Guastini[2], Giorgio Peretti[2], and Darwin G. Caldwell[1]
[1]Department of Advanced Robotics, Istituto Italiano di Tecnologia, Genova, Italy
[2]Department of Otorhinolaryngology, Università degli Studi di Genova, Italy
Email: {giacinto.barresi, nikhil.deshpande, leonardo.demattos, darwin.caldwell}@iit.it,
luca@guastini.eu, g.peretti@tin.it

***Index Terms*—Surgeon performance, ergonomic evaluation, unified rating, usability, SUS.**

I. Introduction

State-of-the-art Laser Phonomicrosurgery (LP) is a set of complex otolaryngological surgical techniques used in the treatment of laryngeal abnormalities, including cysts and tumors. The complexity of this surgical procedure influences its own outcome because of its peculiar issues: (i) the surgical site access can be difficult, (ii) the setting can cause surgeon discomfort, (iii) the outcome quality is totally dependent on the skills of the surgeon and is prone to errors, (iv) the operation methods require extensive training, and (v) the equipment characteristics themselves affect all the previous aspects [8]. The importance of the role of robot-assisted surgical systems is increasing in the operating room [7]. They provide significant advantages in effectiveness, efficiency and safety of surgery through, for instance, increased task precision, reduced tremor in gestures, and timely execution of repetitive tasks. Nevertheless, such recent advances still require an ergonomic approach for a user-centered design and evaluation of surgeon-machine interfaces [12]. The cognitive limits of human beings can become critical factors in excessive mental workload (the amount of cognitive resources used to accomplish a task) [10] and in human error [11]. Thus, human factor-oriented approach is required in user interface design and evaluation to increase both safety (for the surgeon and the patient) and performance of the system. It permits the analysis of the issues of user interface design in terms of human-machine interaction constructs like usability (a complex set of behavioural concepts, like execution time, user satisfaction and ease of learning, [1]) and the objective performance of the user (according to quantitative data automated collection, [4]). Such constructs are considered in this paper for the comparison of 2 user interfaces for LP, in the context of the European project - μRALP. At the Istituto Italiano di Tecnologia (IIT), this research has already resulted in new and improved computer-assisted LP systems. Mattos et al. [8] presented a novel surgeon interface, the Virtual Scalpel system, which replaces the traditional manual micromanipulator interface [6] with a motorized micromanipulator, controlled through a graphics stylus and a touch-screen tablet with live video of the surgical area. The immediate advantage offered by the new interface is that both the aiming and the activation of the laser are controlled by the graphics stylus. The analysis of this novel solution indicated that this system offers greater precision along with better ergonomics over the traditional laser micromanipulator [3], [8], [9]. Through a series of field trials involving a sample of potential users, the new Virtual Scalpel laser control interface [8] was compared with the AcuBlade [6], the traditional, state-of-the-art laser control interface with the manual micromanipulator for LP. This study defines the perceived usability of each interface and the performance of the users in surgery-like tasks. The subjective evaluation of the usability of the interface has been measured by means of questionnaires, and the objective evaluation of the performance of the interface by means of an imaging-based feature extraction method [4].

II. Evaluation Methodology

The experiments were designed to evaluate the effects on usability perception and user performance in two conditions of laser control in LP: (i) the AcuBlade condition (traditional user interface, consisting of microscope for visualization, a micromanipulator for laser aiming, a foot-pedal for surgical laser activation), and (ii) the Virtual Scalpel condition (novel user interface, consisting of a touch-screen laptop for visualization, and a stylus pen for both laser-aiming and activation). The trials were performed at the San Martino Hospital (Genova, Italy), involving a sample of potential end users of the surgical devices. The dependent variables were the scores of the System Usability Scale questionnaire (SUS) [2] and the objective values of performance which were processed through the imaging based metrics algorithm developed in Deshpande et al. [4].

Two dedicated rooms in the otolaryngology division were used for the trials, with complete control of the environmental and social context in order to avoid any disturbance and any effect of latent variables. The bigger room was used for the trials, and it contained a condition-specific chair for the subject (surgical intervention chair for the AcuBlade condition; an

office chair for the Virtual Scalpel condition), a table to support the experiment materials (plaster targets in both conditions, touch-screen laptop in Virtual Scalpel condition), the laser system column, the video column, a camera for video data recording (its orientation permitting the capture of relevant data on the subject's movements and posture). The smaller room consisted of a table and a chair, and it was used to collect the informed consent form and the questionnaires from the subjects.

24 medical students (undergraduate and graduate) with surgery training and experience as assistants in operation room were chosen as subjects for the trials. The subjects were divided into two groups, one per condition, AcuBlade group vs. Virtual Scalpel group, and each subject performed the trials in only one condition. The groups were composed in order to balance the study level (undergraduate and graduate) and the gender of the subjects.

The experimental task consisted of sets of trajectory following exercises wherein the subject performed maneuvers with the laser to follow preset random black traces on red plaster blocks. The traces were shaped as straight lines, C-curves, and S-curves, representative of real surgical actions. They were stamped on small plaster blocks, with each target block having 12 shapes, featuring one of five different randomized sequences of shapes and shape orientations. The subjects performed the trials in two sessions of two plaster blocks each, with a break of 10 minutes between the two, at least.

Before each experimental session, demographic and subject background data were collected by means of a information form. After the trials, the subject filled out an Italian version of the SUS questionnaire in order to collect information about his/her perception of the tool's usability. A question regarding the subjective evaluation of the time spent performing the tasks was also included.

The metrics introduced in [4] permitted to assess the quantitative performance of the subjects. Here, the metrics obtained from the image processing of the trials provided a quantitative basis for the assessment of the two conditions, comparing the trial times and the data of the laser-traced shapes against that of the desired shape along different dimensions (e.g. area, perimeter, orientation).

III. Data Analysis and Discussion

A t-test analysis was used in order to find the differences between the scores of the two groups from the SUS questionnaire and from the objective performance scores [4]. These analyses verify whether the users perceive a statistically significant difference in the level of usability for each interface (AcuBlade vs Virtual Scalpel) and if there is a statistically significant difference of quantitative performance between the two conditions.

The t-test analysis demonstrated the different level of usability of each interface, with a significant difference: the global score of subjective usability of Virtual Scalpel is higher than the score for Acublade. Further explorative analysis highlights the differences on 5 specific sub-scales of the SUS between the two interfaces. The users of the Virtual Scalpel system feel more confident than the users of AcuBlade during the tasks. The Virtual Scalpel interface is easier to use and easier to learn than the AcuBlade interface. The users of Virtual Scalpel would require to learn less processes than the users of AcuBlade as well. The sub-scale analysis clearly points to the ease-of-use and comfort of the Virtual Scalpel interface, over the AcuBlade interface.

In the objective performance assessment, the *unified rating* is used, a weighted sum of the individual metrics. This takes into account the natural performance variation in any human-operated equipment. Each individual laser-traced shaped contributes to the performance assessment analysis. Using equal weighting for each metric, the ideal value for the unified rating in this case is 60.

The two conditions again show a statistically significant difference on the unified rating for quantitative performance. The Virtual Scalpel condition permits a significantly better performance than the AcuBlade condition. Three individual metrics show a significant difference in performance between the two conditions: Perimeter Ratio (under the Virtual Scalpel condition, the laser-traced shapes conform to the length and thickness of the desired shapes, significantly better than in the AcuBlade condition), Orientation Measure (under the Virtual Scalpel condition, the laser-traced shapes are better aligned with the desired shapes than in the AcuBlade condition), Path Following Error (the Virtual Scalpel condition permits easier trajectory following with the laser than the AcuBlade condition).

The comparisons of SUS scores and of unified ratings between the two conditions is represented in Table I.

A t-test was also performed to understand the existence of significant differences between the perceived and actual times for trial completion effect in the two conditions. This result allows to infer the level of mental workload required in each condition [5]. The actual time spent in trials for

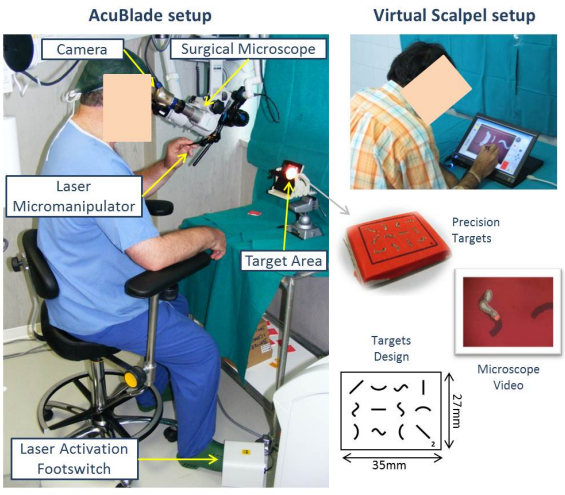

Figure 1. Experimental setup, video snapshot from a trial, and details of the precision targets stamped on a plaster block.

Table I
COMPARISON OF AVERAGE VALUES OF THE METRICS AND UNIFIED RATING SCORES

	AcuBlade condition	Virtual Scalpel condition		
	Means	Means	t	p
SUS Global Score	**65.56**	**83.06**	**4.83**	**0.009**
Unified Rating	**51.37**	**55.96**	**4.26**	**4e-4**

The *SUS Global scores* (subjective assessment of usability) and the *Unified Rating* (objective assessment of performance) for the two conditions. '100' and '60' are the respective maximum values.

Virtual Scalpel is significantly greater than the actual time in AcuBlade, while the estimated time taken for the trials with it is significantly lesser. Therefore, the Virtual Scalpel condition seems to induce less mental workload than the AcuBlade condition. This is advantageous for the acceptance and usability of the Virtual Scalpel condition since it is perceived as a low mental workload interface.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, an integrative ergonomic evaluation of surgeon interfaces in LP was used to compare subjective usability and objective performance of a traditional system (AcuBlade) and a novel system (Virtual Scalpel). The results from the trials can be summarized as follows. (i) The usability of the Virtual Scalpel interface is demonstrated through the global usability value and the 5 different sub-scales of the SUS scale (e.g. the system is "easy to use" and "easy to learn"). (ii) The objective evaluation points to the clear advantage of the Virtual Scalpel interface over the AcuBlade interface. (iii) Both the interfaces show the same actual time for task performance, but the Virtual Scalpel interface is perceived to take significantly less time than the AcuBlade interface, implying that the Virtual Scalpel interface requires a much lower mental workload on part of the user. The t-test analysis showed the significant superiority of Virtual Scalpel in terms of the efficiency and safety aspects of its performance. (iv) Summing up, the subjective and objective evaluations classify the Virtual Scalpel interface as having superior usability and performance capability as a surgeon interface for more effective, efficient and safer laser microsurgeries. In the extension of this research, further studies are planned to better understand the relationships between subjective usability and objective performance in this specific interactive context.

ACKNOWLEDGMENT

The research leading to these results has received funding from the European Union Seventh Framework Programme FP7/2007-2013 – Challenge 2 – Cognitive Systems, Interaction, Robotics – under grant agreement μRALP - n°288233.

REFERENCES

[1] A. Abran, A. Khelifi, W. Suryn, and A. Seffah, "Usability Meanings and Interpretations in ISO standards," *Software Quality Journal*, vol. 11, no. 4, pp. 325–338, 2003.

[2] J. Brooke, "SUS: A "Quick and Dirty" Usability Scale," *Usability Evaluation in Industry*, 1995.

[3] G. Dagnino, L. S. Mattos, and D. G. Caldwell, "New Software Tools for Enhanced Precision in Robot-Assisted Laser Phonomicrosurgery," in *Proc. 34^{th} Intl. Conf. of IEEE Engineering in Medicine and Biology Society, (EMBC 2012)*, 2012, pp. 2804–2807.

[4] N. Deshpande, L. S. Mattos, G. Barresi, A. Brogni, G. Dagnino, L. Guastini, G. Peretti, and D. G. Caldwell, "Imaging based Metrics for Performance Assessment in Laser Phonomicrosurgery," in *Proc. Intl. Conf. on Robotics and Automation, (ICRA 2013)*, 2013.

[5] M. Lind and H. Sundvall, "Time estimation as a measure of mental workload," in *Engineering Psychology and Cognitive Ergonomics*. Springer, 2007, pp. 359–365.

[6] Digital AcuBlade System. Lumenis Inc. Israel. [Online] Available: http://www.surgical.lumenis.com. Accessed on 6-Sept-2012.

[7] S. Martelli, L. Nofrini, P. Vendruscolo, and A. Visani, "Criteria of Interface Evaluation for Computer assisted Surgery Systems," *International Journal of Medical Informatics*, vol. 72, no. 1-3, pp. 35–45, 2003.

[8] L. S. Mattos, G. Dagnino, G. Becattini, M. Dellepiane, and D. G. Caldwell, "A Virtual Scalpel System for Computer-assisted Laser Microsurgery," in *Proc. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems, (IROS 2011)*, Sep. 2011, pp. 1359–1365.

[9] L. S. Mattos, M. Dellepiane, and D. G. Caldwell, "Next-generation Micromanipulator for Computer-assisted Laser Phonomicrosurgery," in *Proc. 33^{rd} Intl. Conf. of IEEE Engineering in Medicine and Biology Society, (EMBC 2011)*, Sep. 2011, pp. 4555–4559.

[10] T. E. Nygren, "Psychometric Properties of Subjective Workload Measurement Techniques: Implications for their Use in the Assessment of Perceived Mental Workload," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 33, no. 1, pp. 17–33, 1991.

[11] J. Reason, "Human error: models and management," *BMJ: British Medical Journal*, vol. 320, no. 7237, p. 768, 2000.

[12] S. Serefoglou, W. Lauer, A. Perneczky, T. Lutze, and K. Radermacher, "Multimodal User Interface for a Semi-Robotic Visual Assistance System for Image Guided Neurosurgery," in *Proc. Computer Aided Radiology and Surgery, (CARS 2005)*, vol. 1281, 2005, pp. 624–629.