Haptic display improves training and skill assessment performance in a virtual paracentesis simulator: A pilot evaluation study

C. S. Tzafestas*, D. Christopoulos
National Technical University of Athens,
School of Electrical and Computer Engineering

K. Birbas†
University of Athens,
Faculty of Nursing, Department of Surgery

ABSTRACT

We have developed a prototype virtual reality simulator platform for training in a class of paracentesis operations. A pilot study was conducted to obtain an initial proof-of-concept on the effectiveness of such a platform, and evaluate its performance as tool: (a) to enhance clinical training and (b) to provide objective assessment of surgical dexterity and skills. The case study that is presented in this paper focused on the emulation of a specific surgical technique used to perform paracentesis of the subclavian vein, which constitutes one of the most commonly applied invasive procedures in the treatment of emergency cases. Two user groups participated in the study, namely: (I) novice users and (II) experienced surgeons. The system automatically provides quantitative assessment scores on the user's performance, applying a set of objective metrics that involve 'optimality' of needle track and measures of manipulation errors. To implement these assessment metrics, the prototype VR simulator incorporates 3D models of all the anatomic areas of interest involved in this procedure, namely: skin, bones, lungs and arterial/venal tree. Furthermore, a haptic interaction system is integrated in the platform, providing the user with a sense of touch during simulated needle insertion. Our main goal in this study was to evaluate the impact of haptic display in the performance of the simulator, regarding: (a) the learning curve for novice users (trainees), and (b) the correlation of the system-generated assessment scores with the actual experience and skill of the user. Experimental findings provide evidence about the significance of a haptic dimension in such a surgical simulation system, especially regarding the ability of the system to correctly and reliably assess the associated surgical dexterity and skill level of the user.

Keywords: haptic display, virtual reality, paracentesis simulator, surgical simulation and training, objective skill assessment

1 INTRODUCTION

Providing efficient training of health-care professionals particularly with respect to invasive critical operations practiced in special clinical environments like an emergency room (ER), is undoubtedly a very difficult process. In such clinical settings, the ultimate degree of dexterity together with "real-time" decision-making skills are needed to safely perform the required procedures. The timely and persistent adaptation of the trainee medical doctor from the theoretical education field to the clinical "hands-on" practice on the real hospital environment, constitutes a great challenge and a primary educational objective in the Health Sciences. Furthermore, performing objective assessment of these clinical dexterity and skills, is in practice extremely difficult. The use of virtual reality (VR) simulators could help bypass these drawbacks by emulating the acquisition of hands-on experience and skills on a 'virtual patient'.

In this direction, we have developed a prototype e-learning platform aiming to provide training on the procedures and operations involved in an ER environment. Operating in ER requires complex clinical skills involving: from one hand (a) rapid decision-making, and on the other hand (b) manual dexterity and sensorimotor abilities, both (a) and (b) under very strict time constraints. The work presented in this paper focuses particularly on the experimental evaluation of a system called DEX (Clinical Dexterity Enhancement and Assessment), which constitutes a generic VR based platform with haptic display, performing realistic interactive simulation for a class of emergency-room invasive procedures, with a particular focus on paracentesis operations.

This paper presents the first pilot study conducted with the DEX platform, to systematically evaluate the performance of the system. A specific scenario was implemented during this first pilot study, concerning a particular surgical technique for obtaining central access on a deep vessel, and more precisely paracentesis of the subclavian vein. Subclavian vein paracentesis is one of the most common procedures in clinical practice and is often involved in the treatment of patients in ER. This operation is particularly difficult to learn, and requires a combination of visual and haptic skills in order to identify the needle insertion point, and to control the needle position and orientation during penetration. To incorporate the "sense of touch" in the simulator platform, a *haptic interaction* system has been developed coupling the VR-based simulation engine with a typical desktop force-feedback device. The challenges addressed at this stage are twofold: (i) from a technical point of view, to achieve a trade-off between realism (accuracy of the physically based simulation) and real-time performance necessary particularly for a stable haptic interaction; (ii) from an educational and human-factors point of view, to conduct experimental evaluation studies in order to identify the critical factors affecting the performance of the system in terms of clinical skill training and assessment.

There exist a few research efforts reported in the literature aiming to develop simulators for specific paracentesis procedures, such as a lumbar puncture simulator [7], an amniocentesis simulator [16], or an abdominal paracentesis simulator [5]. Most of these works focus on the technological challenges that have to be addressed to develop efficient surgical VR simulators. A few other research efforts focus on developing models for 'needle insertion into soft tissue' procedures, such as the force model presented in [12], or the relevant soft tissue deformation models based on finite element methods [4][11].

Such models constitute a necessary step towards the development of efficient simulators for needle insertion operations, particularly when haptic display is to be involved. The goal of effectively integrating such sensorimotor functionalities and skills within a VR system, particularly an interactive VR-based surgical simulator, poses extraordinary challenges for researchers and engineers in the field. The study and integration of haptic interaction modalities in VR systems consists in providing the user of a virtual environment with all the important sensations involved when touching, exploring or manipulating virtual objects [13]. The main objective of the work presented in this paper is to evaluate the effects that such

*e-mail: ktzaf@softlab.ntua.gr
†e-mail: birbas@otenet.gr

haptic display components have on the performance of a VR-based surgical simulator.

The performance of the simulator system is evaluated in a twofold way: (i) as a *training tool*, aiming to help users safely acquire and enhance specific basic surgical skills and dexterities, required to successfully perform such common but critical invasive procedures, prior to any real hands-on clinical practice, and (ii) as a *skill assessment* (and potentially accreditation) tool, on the basis of objective metrics and computer-generated scores (with no subjective human intervention). In the research work described in this paper, we aim to evaluate the extent to which such a system can effectively accomplish these goals. Furthermore, we aim to obtain quantitative evaluation indicators regarding the importance of haptic display in the simulator, and the effect that the presence (or absence) of such an interactive, sensorimotor simulation component has on the performance of the system.

The paper is organized as follows: Section 2 presents a brief overview of the simulator and the haptic display features of the system. This paper focuses on the pilot evaluation study conducted, which is described in Section 3, including the methodology and experimental protocol employed, the objective metrics used for skill assessment and the experimental results obtained. The results are analyzed and discussed in respect to our research hypotheses regarding training and assessment, as well as to the effect of haptic display in the performance of the system. Section 4 provides conclusive remarks and future work directions.

2 VR Paracentesis Simulator: Experimental Setup

2.1 Overview

We have developed a prototype VR paracentesis simulator platform, which was originally presented in [14]. The core of the first prototype simulator system is comprised of a *virtual patient* with "visible" anatomic landmarks and "touchable" anatomic areas of interest. The user can: (a) select the puncture site with respect to specific visible anatomic landmarks, (b) orient the puncture device on a 3D anatomic space, (c) perceive the tissue deformation (on the human skin) caused by the device, and (d) feel the haptic equivalent produced during the insertion through the anatomic layers.

In its current configuration, the system can simulate most of the potential complications in accordance to the situations that may be encountered in the real clinical setting. For this reason, the system incorporates 3D models of all the human anatomical structures of interest involved in the simulated paracentesis procedure, namely: skin, bones (clavicle and thorax), lungs and arterial/venal tree. It also includes 3D models of surgical instruments (in our case the needle parts). Stereoscopic 3D graphics rendering routines are implemented using OpenGL, and collision detection is implemented using the ColDet library. Details about skin deformation and force computation modeling used at this stage are presented in [15].

Figure 1 shows the experimental setup of the simulator platform employed throughout the evaluation trials. A pair of 3D stereo glasses was used for stereoscopic display, to enhance the visual realism of the simulation. To provide the user with a feeling of contact (touch) forces applied during the performed paracentesis operation, the system was coupled to a haptic device, a Phantom® Desktop™ force feedback device (from SensAble Technologies: http://www.sensable.com/), as discussed in the sequel.

2.2 Haptic Display

The most important issue in a haptic display system is undoubtedly that of achieving an 'optimal' tradeoff between two, often contradictory, requirements: transparency and stability. The first one is related to the realism of the simulation and the physically-based

Figure 1: *Overview of the simulator setup used during experiments.*

accuracy of the forces displayed to the user, which calls for complex and usually computationally expensive models. The second requirement calls for efficient calculations and real-time fast control loops to ensure a stable force-reflecting interaction. The main bottleneck in this respect is the visualization loop, performing all the necessary computations regarding collision detection, contact and deformation modelling within the 3D graphics environment. These procedures all together usually run at a 20-30 Hz frequency, which is definitely a very slow update rate with respect to the real-time control requirements. The force feedback control loop usually operates in a separate thread at 1 KHz or more. However, this fast control rate is in fact not exploited, if all critical updates for force-feedback computation (contact location, deformation data, etc.) are directly coming from a slow graphics loop. Similarly to the effect of a zero-order hold in any control system, this delay in obtaining new critical information updates may cause undesirable chattering or even instability, particularly when simulation of hard contact is involved.

All these problems are now very well known in the haptics research community, and several solutions have been proposed, most of them related to the application of some type of a 'virtual coupling' (instead of a 'hard' direct force/position interconnection) aiming to improve stability of interaction between haptic master and virtual simulation environment (e.g. [3][8]). Needless to say, though, that such a 'soft coupling' between master and slave system may deteriorate the transparency of the system and, in our case, significantly decrease the realism of the simulation with respect to the haptic skills involved in the specific simulated procedure.

In this respect, we have decided to employ, instead, a local geometry approximation technique, decoupling the haptic display computations from the visualization engine, for the particular case of a needle in contact with specific anatomical structures (such as the 'flat-type' human skin or the 'cylindrical-type' clavicle bone). Feeling these constraints through the sense of touch (particularly the hard haptic constraint invoked by the presence of the clavicle bone above the needle track), by means of the forces exerted on the manipulated needle, is considered as the most important part of the manual skill involved in performing paracentesis of the subclavian vein. Ensuring efficient computations and fast updates was considered as of primary importance, even if it is in the disadvantage of the accuracy of the geometrical representations used in force feedback. This was subsequently validated during experimental trials through qualitative assessments performed by experienced surgeons.

3 PILOT STUDY: METHODOLOGY AND RESULTS

Paracentesis of a deep vessel (such as the subclavian vein) is one of the most common procedures used in clinical practice and is often involved in the treatment of patients in an ER environment. Such invasive procedures are particularly difficult to learn, and "hands-on" training may eventually include substantial risk for the patient. As already mentioned, the VR-based haptic simulator presented in this paper aims to help clinicians acquire some basic skills and dexterity prior to any real clinical operation. The effectiveness of such a VR simulator as a training tool needs to be evaluated in terms of its capacity to enhance the surgical skills of the user, which requires long-term clinical studies to validate. However, validating the reliability of such a system as a tool for objectively assessing the clinical skill and dexterity of the user, can constitute an initial 'proof-of-concept' and provide an indication about the effectiveness of such simulation techniques in practical use. The objective assessment of the psychomotor skills required to perform an invasive procedure in a real clinical setting, is a very difficult task due to the inherent measurement difficulties associated with evaluating skills and dexterities in vivo. Developing procedures and techniques to reliably perform such objective and real-time skill assessment is now becoming a priority in the medical sector. In this context, there is now an emerging body of evidence to establish the validity of simulation techniques for assessing surgical skills [2][1][10], such as for instance studies demonstrating the effectiveness of the MIST-VR laparoscopic simulator [6][9].

In the pilot study presented in this paper, we aim to evaluate the extent to which such a system can effectively perform: (a) realistic simulation of the considered invasive procedures, in this case, paracentesis of the subclavian vein; (b) measurement and reliable assessment of the user's "level of dexterity and skills" in performing the considered clinical technique. In the rest of the paper, we describe the metrics defined to compute user assessment scores, as well as the experimental protocol, the procedures employed and the results obtained during the pilot study presented in this paper.

3.1 Metrics for clinical dexterity and skills assessment

The prototype paracentesis simulator system is designed to automatically generate a global quantitative assessment score of the user's performance in executing the simulated clinical procedure. This assessment score is based on evaluating a set of objective performance measures, which include the following:

(i) indicators of major, safety-critical manipulation errors that may lead to serious complications or risk of injury to vital anatomical areas, within the virtual surgical space

(ii) deviations from an 'optimal needle path' including: selecting the needle insertion point, as well as needle orientation in the venal access track

(iii) 'economy of movements' measures, and indicator on the 'smoothness' of needle motion

(iv) the total time needed to successfully complete the simulated procedure

The system continuously records, in real-time, all the above mentioned performance indicators, automatically generating individual values as "penalty-marks" for each one of these score categories, at the completion of each experimental trial. All these performance measures are then weighted together to obtain an overall (total) score indicator for each performed procedure. For instance, regarding time-to-complete, a value e.g. of 30,000 (that corresponds to 30,000 msecs, i.e. 30 secs) is weighted by a factor of 0,01 contributing to the final score the value of 300. All other penalty values are normalized accordingly with a corresponding weight factor, regarding their considered significance to the successful outcome of the procedure. In other words, an error corresponding to a major complication risk (such as, for instance, needle touching a lung, with the risk of pneumothorax) is weighted more than other false maneuvers considered by doctors as minor in the course of the procedure. It must be highlighted again that, since score values correspond in fact to penalty marks (resulting from: (i) time elapsed to successfully complete the procedure, and (ii) errors committed, false maneuvers performed, or generally suboptimal surgical manipulation), a higher obtained score is an indicator of worse performance in the simulated procedure.

With respect to these metrics, an important issue concerns calibrating the system to obtain reliable objective score for each user. For this reason, we conducted a set of "calibration sessions", where the system was set on a "teach mode" and used by a group of experienced surgeons to demonstrate the characteristics of the "average optimal path" and "acceptable deviations". In other words, during these sessions, the experienced surgeons in fact "teach" the simulator how the considered paracentesis technique should be performed in an optimal way (i.e. 'by the book'). These data have then been processed to extract parametric representations of the 'optimal needle-insertion track', based on which the real-time evaluation of the respective objective measures will proceed. It must be noted, of course, that the surgeons who have 'calibrated' the skill-assessment module of the system, did not participate in any way in the subsequent experimental evaluation study.

3.2 Experimental protocol and evaluation procedures

The pilot study presented in this paper was based on a specially designed experimental protocol that was consistently employed throughout the trials. According to this protocol, users from two pre-defined groups participated in the experimental sessions:

• Group-I (novices): consisting of nursing students or graduates. It is important to note that all users in this group had the theoretical (human anatomy etc.) knowledge of the considered clinical procedure (most of them had also been physically present, several times in the past, in such operations), but had no real prior 'hands-on' experience and definitely no specialized clinical training; in other words, a-priori they did not possess the surgical dexterity and skills required to perform such operations.

• Group-II (experienced): consisted of trained surgeons, with significant experience in performing the considered paracentesis procedure. All users in this group had performed over 100 operations prior to their participation in this experiment, some of them being also in charge of teaching trainees surgeons in such deep-vessel paracentesis procedures.

Each group included 10 users. The total number of the sample was, thus, 20 (N) subjects.

In this pilot study, the system was experimentally tested in a specific surgical technique employed for the paracentesis of the subclavian vein. According to the specific technique used in this pilot study, a large-caliber needle, attached to a 12-ml syringe is introduced 1 cm below the junction of the middle and the medial thirds of the clavicle. After the skin has been punctured, with the bevel of the needle upward, the needle and the syringe are held parallel to the frontal plane. Then, the needle is directed medially, slightly cephalad, and posteriorly behind the clavicle toward the posterior, superior angle to the sternal end of the clavicle (toward finger placed in the suprasternal notch). The needle is slowly advanced while the plunger of the syninge is genlty withdrawing until a free flow of blood appears in the syringe.

Particular attention was payed as to conduct the experimental procedures in a very systematic way, consistently for all users throughout this pilot study, in order to eliminate from the obtained results any undesirable biasing effect that may be due to reasons other than the controlled parameters' variation. For this reason, we have also decided to disable, during these controlled experiments, any interactive virtual navigation features within the virtual patient's simulated anatomical space. Instead, the simulator

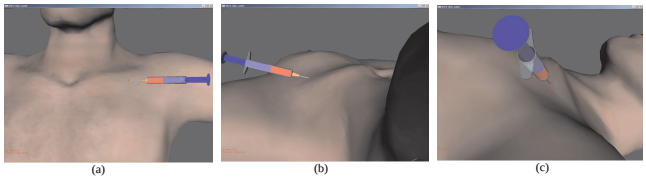

Figure 2: *Orthographic 3D views of the virtual patient used during the simulated paracentesis operation.*

interface allowed the user to selectively switch between three pre-specified fixed orthographic views of the virtual patient, in order to also assist the user to accurately position the needle and orient it according to the appropriate visual guide-points (Fig. 2).

Each user participated in one *experimental session*, consisting of ten consecutive *experimental trials*. One *familiarization trial* preceded each experimental session, where the trainer (experimenter) also exposed in a predefined way the basic requirements of the procedure (that is, basically, the theoretical, anatomical and technical aspects of the specific paracentesis technique the user was asked to perform, and the means of interaction with the simulator system). Each experimental session lasted approximately 30 minutes.

The values of the controlled parameters were selected at the beginning of each experimental session in a randomized manner. In this pilot study case, this controlled variation concerned only the activation (or not) of haptic display, thus leading to two distinct experimental conditions to be comparatively evaluated:

• condition (a): NO, i.e. paracentesis simulation without force-feedback (active force-feedback disabled)

• condition (b): FF, i.e. simulation with active force-feedback enabled.

A total of of five users from each group performed the experiments under each experimental condition. Besides this controlled variation, each user performed the same series of experimental trials with the simulator, with exactly the same pre-defined level of difficulty and interaction features. In the future, we plan to evaluate the effect of other parameters in the performance of the system as a surgical skill training and assessment tool, involving both the visual channel (stereo visual display and active VR navigation features) and the haptic display channel (particularly, the effect of bimanual haptic interaction).

3.3 Evaluation Results and Discussion

As already mentioned, our main goal in this pilot study is to evaluate the degree to which the presence (or absence) of a haptic (in this case, kinesthetic) display during the VR-simulated paracentesis procedures can affect the performance of the system. This can be evaluated in terms of: (1) the learning curve of novice vs. experienced users, and (2) the capacity of the system to accurately 'predict' the clinical skill level of the user, and thus perform efficiently as an objective skill assessment tool. Such findings should give us an indication about: (i) how well the related sensorimotor skills are emulated by the prototype VR platform, and thus consequently (ii) how well such a simulator system could be potentially used as a tool for teaching basic surgical skills.

In line with these objectives, our interest during this pilot study focused on the evaluation and potential experimental validation of two main research hypotheses:

Hypothesis-1 (Learning curve of novice users): *The users belonging to group-I (novices) are expected to show a significant improvement of their performance in the course of each experimental session, that is from the first experimental trial (trial 1) to the last one (trial 10).*

GROUP-I (a, NO)	TIME (secs)	TOTAL SCORE
Mean (Trials 1-2)	71.5	1918.99
STD (Trials 1-2)	27.1	1163.55
Mean (Trials 9-10)	55.2	1223.76
STD (Trials 9-10)	42.3	767.60
Learning rate (%)	22.84%	36.22%

Table 1: Learning Rate for Group-I users, under experimental condition (a): NO force-feedback

Group-I (b, with FF)	TIME (secs)	TOTAL SCORE
Mean (Trials 1-2)	89.9	1636.37
STD (Trials 1-2)	41.4	1447.34
Mean (Trials 9-10)	46.3	859.32
STD (Trials 9-10)	28.3	346.26
Learning rate (%)	48.42%	47.48%

Table 2: Learning Rate for Group-I users, under experimental condition (b): FF, with active force-feedback

This performance improvement (that is, the learning rate for the novice group-I users) could be considered to constitute an initial indicator about the potential of the system as a *basic skills acquisition and training tool*. Of course, to draw such a conclusion we need to correlate the simulator-generated performance scores with the actual clinical skills of the users, which constitutes our 2nd research hypothesis to be evaluated in the sequel.

The validation of our first experimental hypothesis is demonstrated in Table 1, where the results (mean value and standard deviation) of group-I scores ('time' and 'total score') are presented, for experimental condition (a) (i.e. NO force-feedback). The "learning rate" (and consequently the "skill-transfer" potential) of the system in this case can be deduced by comparing the mean score values between the first two trials (trials 1-2) and the last two ones (trials 9-10). For the 'Time', the learning rate is approximately 22%, while for the 'Total' score this improvement is 36%.

On the contrary, as opposed to group-I users, group-II (experienced) users show no apparent learning gain in the course of each experimental session (trial-1 through trial-10). The performance of group-II users remains practically constant throughout the experiment, without any apparent improvement rate in the learning curve, and this irrespective of the activation or not of the haptic display features. This is definitely a "positive sign" for the system, revealing that the experienced users adapt easily to the requirements and the functionality of the simulation, which could also mean that they are able to intuitively perform the required procedures in the simulator, employing their normal skills and dexterities.

The Effect of Haptic Display on the Learning Curve

A research question that needs to be addressed at this point is the following: *Is the learning curve, particularly for group-I (novice) users, affected by the experimental condition, that is, by the activation (or not) of the haptic display in the simulation?*

An answer to this question can be deduced by analyzing the results presented in Table 2, which concern group-I users for the experimental condition (b) (FF, i.e. with active force-feedback). Comparing the results in this table with the ones presented above in Table 1, we can observe a significant performance improvement of approximately 48% for 'Time' (with respect to the value of 22%, in Table 1 above), and 47% for the 'Total' score (as compared to 36%, for experimental condition (a) above). This improvement provides an initial demonstration of the positive impact the activation of haptic display (experimental condition (b)) has on the performance of the system, as evaluated here with respect to the learning rate for

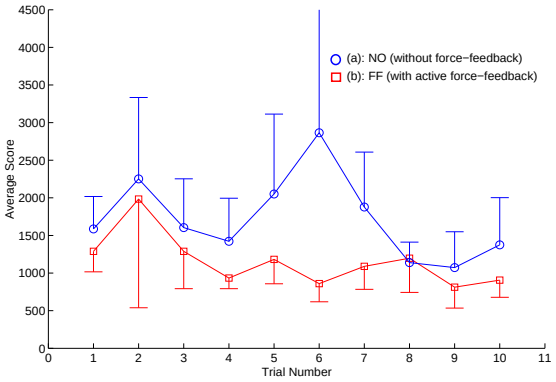

Figure 3: *Experimental results for group-I users: Average score over all users for each experimental trial, comparatively for the two experimental conditions: (a) without (NO), and (b) with (FF) active force-feedback.*

novice users.

These results concerning the learning curve for group-I users, comparatively for the two experimental conditions, are also depicted in Figure 3 for the average total score. This figure shows the evolution of the simulator-generated performance scores in the course of the experimental sessions (i.e. from trial-1 to trial-10), with the numerical data averaged over all users performing under each experimental condition, and depicted separately for each one of the ten trials. The improvement observed in the case of experimental condition (b), that is when haptic display is active, concerns both: (i) average absolute values for the scores as well as (ii) the reliability of these measures as related to the variance of these values (see standard deviation STD values for trials 9-10 in table 2 being significantly reduced as compared to the corresponding values in table 1). This provides a clear indication about the benefit that can be anticipated when realistic haptic display features are integrated in such a simulation and training system, with respect to a "faster and more reliable learning curve", a particularly interesting finding if it is interpreted as referring to a more *efficient skill acquisition process* for novice users (trainees). Of course, these simulator-generated score values need to be correlated with the actual skills that are needed to perform the clinical procedure in real, which is the scope of our second experimental hypothesis.

Hypothesis-2 (Skill assessment): *The system can reliably and objectively assess the dexterity and skill level of the user, by means of the performance scores automatically computed after each experimental session.*

To explore the validity of this hypothesis, the experimental results were analyzed comparatively for the two user groups (group-I: novice users, vs. group-II: experienced surgeons). The overall results obtained for group-II users (for the time-to-complete and total score values) are given in table 3. Comparing these results with the ones of group-I above (even with best group-I performance of trials 9-10 in table 2), one can observe that the overall performance of group-II (i.e. experienced) users becomes significantly better in the case of experimental condition (b), that is, with the haptic display active:

- ▷ Average Time-to-complete : 32.7 secs, vs. 46.3 secs before (std = 14.7 vs. 28.3 before)
- ▷ Mean Total Score : 753 vs. 859 before.

Therefore, these results demonstrate a performance improvement for group-II (experienced) users with respect to group-I (novice) users. However, this conclusion is eminently accurate only when haptic display is active (exp. condition (b), FF), as further explained in the sequel.

Group-II	TIME (secs)	Total SCORE
Mean value, (a): NO	53.4	1580
STD, (a): NO	35.4	900
Mean value, (b): FF	32.7	753
STD, (b): FF	14.7	430

Table 3: Overall results for Group-II, for the two experimental conditions (a) without (NO) , and (b) with active force-feedback (FF).

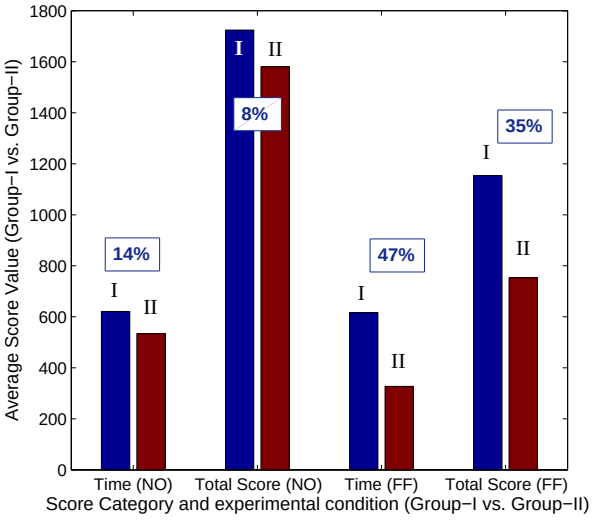

Figure 4: *Overall comparative results between group-I and group-II, for the two experimental conditions: (a) without (NO) and (b) with force-feedback (FF).*

The Effect of Haptic Display on Skill Assessment

Figure 4 illustrates the overall comparative results between group-I and group-II, for the two experimental conditions (that is, with and without haptic display). As shown in this graph, the performance improvement, in terms of the scores obtained by group-II (experienced) users as compared to users of group-I (novices), varies from approximately 8% (total score) to 14% (time to complete) in the case of experimental condition (NO), that is, when haptic display is disabled. This performance improvement for the group-II is apparently much more important in the case of experimental condition (FF), that is, when haptic display (in our case force-feedback) is activated. The performance gain in this case varies between approximately 35% (total score) and 47% (time to complete).

Furthermore, a statistical analysis of the results reveals that this apparent performance improvement for group-II users is statistically significant only in the case of experimental condition (FF), that is, only when haptic display is activated in the simulation (t-test results: $p < 0.01$ for the total score, and $p < 0.001$ for the time-to-complete, in the case of experimental condition (FF)).

We can thus conclude that there is a statistically significant correlation between the objective performance scores, automatically generated by the system, and the level of experience / skill of the user, as related to the group to which each user belongs. Furthermore, what is even more interesting, is that this important correlation seems to appear only when haptic display is active. This

finding supports our experimental hypothesis about the significant importance of haptic display in such interactive surgical simulator and training system. Without this 'haptic component' the experienced users seem to be unable and fail to apply on the simulator their skill and dexterity acquired over long-time training and hands-on clinical practice. This is eminent when one observes the 'time-to-complete' score, which shows a particularly significant improvement ($p < 0.001$) when activating haptic display in the simulator, meaning that in this case experienced surgeons seem to become 'familiar' with the experimental setting and perform with 'confidence' and success the simulated surgical procedure. In this case, one can conclude that haptic display adds an irreplaceable component for the realism of the simulation, contributing significantly to an efficient emulation of the dexterity and skills required to perform the considered surgical procedures.

4 Conclusion

This paper presented a pilot evaluation study conducted to explore the performance of VR paracentesis simulator with respect to two principal goals: (a) training for the acquisition and enhancement of basic clinical skills and dexterities, (b) objective assessment of the user's skill level and dexterity related to specific surgical procedures. Our main objective in this study was to obtain quantitative evaluation results regarding the importance of haptic display to the user, and the effect that such an interactive component has on the performance of the system.

The experimental results obtained during this pilot study are presented and analyzed in this paper. The conclusions drawn support our basic hypotheses, regarding the learning curve for novice users, and the skill assessment functionality of the system. Particularly, with respect to the beneficial effect of haptic display on system's performance, our conclusions can be summarized in the following statements: (1) the presence of haptic display considerably improved the learning rate for novice users, leading to a steeper and more consistent learning curve, and (2) the activation of haptic display had a statistically significant effect on skill assessment performance of the system, with respect to a comparative analysis of the scores obtained between novice and experienced users.

These experimental findings, thus, seem to support the conclusion that the effect of active haptic display in such an interactive simulator system is predominant, adding an irreplaceable component for the realism and efficiency of the simulation, in terms of both: (a) contributing significantly to an efficient emulation of the dexterity and skills required to perform the considered surgical procedures, and (b) improving the performance of the system with respect to the potential for skill training and assessment. It must be noted nevertheless that larger-scale, long-term clinical studies are required before safe conclusions can be drawn regarding the use of such a simulator as a training tool. Feedback collected from the users (nursing students or graduates, MDs, experienced surgeons) participating in the pilot study presented in this paper is more than encouraging (sometimes even enthusiastic), clearly demonstrating the potential of this VR and haptics technology to provide breakthrough solutions and tools that meet (with a yet unattained performance) specific, particularly challenging requirements imposed in the medical education and training sector.

Acknowledgments

This work was partially supported by the Greek General Secretariat for Research and Technology and the European Commission, under grant EL4-2003 (Project Name: "VRES"). This research is also conducted in the frames of "Intuition", EU FP6 Network of Excellence (WG 2.10 on Haptics).

References

[1] E. Acosta and Bhart Temkin. Haptic Laparoscopic Skills Trainer with Practical User Evaluation Metrics. *Studies in Health Technology and Informatics*, vol. 111, pp. 8-11. In: Medicine Meets Virtual Reality 13 (ISBN: 1-58603-498-7), 2005.

[2] J. A. Aucar, et al. A Review of Surgical Simulation With Attention to Validation. *Surgical Laparoscopy, Endoscopy & Percutaneous Techniques*, 15(2): 82-89, 2005.

[3] J. E. Colgate, M. C. Stanley, and J. M. Brown. Issues in the haptic display of tool use. In: *IEEE/RSJ International Conference on Intelligent Robots and Systems* (IROS'95), pp. 140-145, 1995.

[4] S. P. DiMaio and S. E. Salcudean. Needle Insertion Modelling and Simulation. *IEEE Transactions on Robotics and Automation*: Special Issue on Medical Robotics, 19(5):864-875, 2003.

[5] A. Frisoli, L. F. Borelli, C. Bianchi, E. Ruffaldi, G. Di Pietro, and M. Bergamasco. Simulation of real-time deformable soft tissues for computer assisted surgery. *The International Journal of Medical Robotics & Computer Assisted Surgey*, vol. 1, issue 1, Jan. 2004 (available at: http://www.roboticpublications.com/).

[6] A. G. Gallagher, N. McClure, J. McGuigan, I. Crothers, and J. Browning. Virtual reality training in laparoscopic surgery: a preliminary assessment of minimally invasive surgical trainer virtual reality (MIST VR). *Endoscopy*, 31(4):310-3, 1999.

[7] P. Gorman, T. Krummel, R. Webster, M. Smith, and D. Hutchens. A Prototype Haptic Lumbar Puncture Simulator. In: *Medicine Meets Virtual Reality 2000*, J.D. Westwood et al. (Eds), pages 106-109, IOS Press 2000.

[8] M. Mahvash and V. Hayward. High-fidelity passive force-reflecting virtual environments. *IEEE Transactions on Robotics*, vol. 21, no. 1, pp. 38-46, 2005.

[9] S. S. McNatt and C. D. Smith. A computer-based laparoscopic skills assessment device differentiates experienced from novice laparoscopic surgeons. *Surg. Endosc.*, 15: 1085-1089, 2001.

[10] L. Moody, C. Baber, T. Arvanitis, and M. Elliot. Objective metrics for the evaluation of simple surgical skills in real and virtual domains. *Presence: Teleoperators and Virtual Environments*, 12(2): 207-221, 2003.

[11] A. Han-Wen Nienhuys and Frank van der Stappen. Interactive needle insertions in 3D nonlinear material. Technical Report UU-CS-2003-019, Institute of Information and Computer Sciences, Utrecht University, 2004.

[12] A. M. Okamura, C. Simone, and M. D. O'Leary. Force Modeling for Needle Insertion Into Soft Tissue. *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 10, October 2004.

[13] C. S. Tzafestas. Whole-hand kinesthetic feedback and haptic perception in dexterous virtual manipulation. *IEEE Transactions on Systems, Man and Cybernetics* (Part A), 33(1):100-113, Jan. 2003.

[14] C. S. Tzafestas, Y. Koumpouros, and K. Birbas. Haptic interaction in VR-based paracentesis simulation for dexterity enhancement and assessment. In: Proceedings of the *18th International Congress and Exhibition on Computer Assisted Radiology and Surgery - CARS 2004*, Chicago, USA, June 23-26, 2004.

[15] C. S. Tzafestas, Y. Koumpouros, and K. Birbas. Paracentesis modeling and VR-based interactive simulation with haptic display for clinical skill training and assessment. In: *Intern. Conf. on Integrated Modeling & Analysis in Applied Control & Automation* (IMAACA'2004, Part of I3M: International Mediterranean Modeling Multiconference), Special Session on Virtual Reality and Haptics for Modeling, Design and Interactive Simulation, Genoa, Italy, October 28-31, 2004.

[16] J. Zhang, S. Payandeh, and J. Dill. Haptic Aided Design: A Case Study. In: *Proceedings of the 11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems* (HAPTICS'03), 2003.