

Costas S. Tzafestas*

Division of Signals, Control and
Robotics
School of Electrical and Computer
Engineering
National Technical University of
Athens
Zographou Campus
15773 Athens, Greece

Kostas Birbas

Faculty of Nursing
Department of Surgery
University of Athens

Yiannis Koumpouros

Faculty of Technological
Applications
Department of Informatics
Technological Education Institute of
Athens

Dimitris Christopoulos

Virtual Reality Department
Foundation of the Hellenic World
Athens 11851 Greece

Pilot Evaluation Study of a Virtual Paracentesis Simulator for Skill Training and Assessment: The Beneficial Effect of Haptic Display

Abstract

Effective, real-time training of health care professionals in invasive procedures is a challenging task. Furthermore, assessing in practice the acquisition of the dexterity and skills required to safely perform such operations is particularly difficult to perform objectively and reliably. The development of virtual reality (VR) simulators offers great potential toward these objectives, and can help bypass some of the difficulties associated with classical surgical training and assessment procedures. In this context, we have developed a prototype VR simulator platform for training in a class of invasive procedures, such as accessing central vessels. This paper focuses more particularly on a pilot study treating the specific application case of subclavian vein paracentesis. The simulation incorporates 3D models of all the human anatomy structures involved in this procedure, where collision detection and response algorithms are implemented to simulate most of the potential complications in accordance with the situations encountered in real clinical practice. Furthermore, haptic display is integrated using a typical force feedback device providing the user with a sense of touch during the simulated operations. Our main objective in this study was to obtain quantitative evaluation results regarding the effect of haptic display on performance. Two user groups participated in the study: (I) novice users and (II) experienced surgeons. The system automatically provides quantitative assessment scores of users' performance, applying a set of objective measures that also involve the optimality of the needle insertion path and indicators of maneuvering errors. Training and skill assessment performance of the system is evaluated in a twofold manner, regarding respectively: (a) the learning curve of novice users, and (b) the correlation of the system-generated scores with the actual surgical experience of the user. These performance indicators are assessed with respect to the activation of the haptic display and to whether this has any beneficial effect (or not). The experimental findings of this first pilot study provide quantitative evidence about the significance of haptic display, not only as a means to enhance the realism of the surgical simulation, but especially as an irreplaceable component for achieving objective and reliable skill assessment. Further larger-scale and long-term clinical studies are needed to validate the effectiveness of such platforms for actual training and

dexterity enhancement, particularly when more complex sensorimotor skills are involved.

I Introduction

The barrier between theoretical knowledge and clinical performance is probably the biggest problem as far as the education of medical doctors and nursing professionals is concerned. Efficient clinical training of health care professionals constitutes undoubtedly a very difficult task, which is nowadays primarily based on the close supervision and monitoring by a specialist trainer, on the patient's consent, but also under suitable general conditions; a combination of factors which is not always attainable since the main and primary concern always remains, that of treating the patient in the best possible way. All these remarks highlight the inherent difficulties of providing efficient training for practicing in specific clinical environments, such as the Emergency Room or the Intensive Care Unit, where the ultimate degree of dexterity together with real-time clinical skills are needed. The timely and persistent adaptation of the trainee medical doctor from the theoretical education field to the clinical hands-on practice in the real hospital environment, constitutes undoubtedly a great challenge and a primary educational objective in the Health Sciences. Training on cadavers, which is still used today after over a century of practice, is not always adequate to cover other than special needs, not always of primary importance in surgical education and training. The use of animals constituted an attractive solution for decades, but for obvious reasons related to human ethics and evolving legislation issues, as well as due to the need for important investments in terms of the related infrastructure, this solution has been practically abandoned. The experience obtained from the application of flight-simulation systems for the training of pilots, in an application environment that resembles in a way the field of clinical practice (in terms of the need for rapid complex data evaluation and critical decision making) has more recently increased interest in the use of virtual reality (VR) technologies in the medical education field (Burdea & Coiffet, 2003).

The research work described in this paper is carried out in the frame of a larger-scale research project aiming to develop a computer-based educational platform to provide training on the procedures and operations involved in an emergency room (ER) environment. Operating in ER requires complex clinical skills involving: (a) rapid decision making, and (b) manual dexterity and sensorimotor abilities, both (a) and (b) under hard time constraints. In this context, we have developed a prototype VR-based simulation platform for a class of invasive operations that are very often needed in ER clinical settings.

The work presented in this paper focuses particularly on the first pilot study conducted using this system. To systematically evaluate the performance of the system, a specific clinical case has been implemented during this study, emulating a particular surgical technique used to perform paracentesis and obtain central access on a deep vessel, namely 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 an ER situation. 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, a haptic interaction system has been developed coupling the VR-based simulation engine with a typical desktop force-feedback device (a PHANToM Desktop 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 dynamic simulation) and real-time performance necessary particularly for a stable haptic interaction with the system; (ii) from an educational and human-factors point of view, to conduct evaluation studies in order to identify the critical factors affecting the performance of the system in terms of clinical skill training and assessment. The latter constitutes the focus of the work presented in this paper.

I.1 Background

Over the last ten years, a number of virtual reality based medical simulators have been developed for train-

ing in specific surgical procedures. Attractive application fields include minimally invasive surgery (MIS, e.g., Kühnapfel, Çakmak, & Maasz, 2000) and endoscopy procedures (e.g., Baur, Guzzoni, & Georg, 1998), also called keyhole procedures. Laparoscopy trainers (such as the MIST-VR trainer, e.g., Sutton et al., 1997) and arthroscopy simulators (e.g., Heng et al., 2004) constitute typical application scenarios. Real-time simulation of open surgical procedures poses significant technical challenges, and only some simple application examples have recently begun to appear, involving specific surgical procedures, such as anastomosis and wound suturing (e.g., Webster, Zimmerman, Mohler, Melkonian, & Haluck, 2001). Developing complete open surgery simulators still remains a distant target.

In this context, our research work reported in this paper explores the application of VR and, particularly, haptics technologies in a specific class of invasive operations that involve paracentesis and catheterization of deep vessels to obtain central vascular access. There exist a few research efforts reported in the literature aiming to develop simulators for specific paracentesis procedures, such as a lumbar puncture simulator (Gorman, Krummel, Webster, Smith, & Hutchens, 2000), an amniocentesis simulator (Zhang, Payandeh, & Dill, 2003), or an abdominal paracentesis simulator (Frisoli et al., 2004). 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 Okamura, Simone, and O’Leary (2004), or the relevant soft tissue deformation models based on finite element methods (DiMaio & Salcudean, 2003; Nienhuys & van der Stappen, 2003).

Such models constitute a necessary step toward the development of efficient simulators for needle insertion operations, particularly when interactive simulation with haptic display is to be involved. The term “haptic” derives from the Greek word “ $\alpha\varphi\eta$ ” and refers generally to the sense of touch. This is usually divided into two main sensory modalities (Burdea, 1996): (a) kinesthesia, which includes perception of muscular effort as well as proprioception, and (b) tactile sense, which provides

cutaneous information related to contact between the skin and the external environment, thus enabling the perception of physical properties such as the surface characteristics of touched objects (texture, temperature, etc.). The integration of haptic interaction modalities in virtual reality systems means providing the user of a virtual environment with all (or part of) these sensations that are involved when touching, exploring, or manipulating virtual objects (Tzafestas, 2003). Integrating such sensorimotor functionalities and skills, particularly within medical (and especially surgical) VR-based simulation systems, poses extraordinary challenges for researchers and engineers in the field.

1.2 Research Objectives—Aim

The main objective of the work presented in this paper is to study the effect that a haptic display component can have on the performance of a VR-based surgical simulator. Performance of such a system needs to be assessed in a twofold way: (i) as a *training tool*, aiming to help users safely acquire and enhance specific surgical skills and dexterities, required to successfully perform 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 (i.e., with no subjective human intervention).

To evaluate the performance of a VR simulator as a training tool, long-term clinical studies are needed to assess any actual improvement of surgical skills in real operations. However, an initial proof-of-concept on the efficiency of such VR simulation techniques can be established by evaluating the performance of the system as a skill assessment tool. Objective assessment of psychomotor skills related to an invasive procedure is undoubtedly a very difficult task, due to the inherent difficulties associated with a qualitative and/or quantitative measurement of surgical 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. It is consid-

ered that VR-based simulation technologies can play a significant role in this direction.

Most of the existing research efforts (such as the ones referenced in the previous section) focus on the technological challenges that have to be addressed to develop efficient VR surgical simulators. However, there exist very few results reported in the literature that provide objective and measurable evaluation data on the performance of such systems, particularly with respect to the effect of haptic display as a sensorimotor, interactive simulation component. Most of these works focus on specific tasks such as controlling surface puncture during a simple one degree-of-freedom haptic needle insertion task, like for instance in Gerovich, Marayong, and Okamura (2004). Our research work is substantially different in that it covers a more complete surgical procedure, an invasive technique for specialized paracentesis operations, which requires composite surgical skills and manual dexterity to perform (as opposed to, e.g., a simple needle insertion task, which in fact constitutes the very first step of the initial phase of a complete paracentesis operation). Our goal was to study such a skill as a whole, and assess the beneficial effect (if any) of haptic display in the simulator. The assessment metrics that we used cover a variety of performance indicators, including both speed and surgical precision, as well as safety of operation in terms of the risk for potential complications, in accordance to real operating conditions.

The paper is organized as follows: Section 2 presents an overview of the application requirements and the technical aspects of the various system components. Some of the implementation techniques employed to develop the prototype simulator are briefly discussed, including visualization and haptic display features. The pilot evaluation study is described in detail in Section 3. This includes the methodology employed, the objective metrics used, and the results obtained. These results are analyzed with respect to our research hypotheses regarding training and skill assessment, as well as in relation to the effect of haptic display on the performance of the system. Section 4 provides conclusive remarks and future work directions, both from a technical and experimental evaluation perspective.

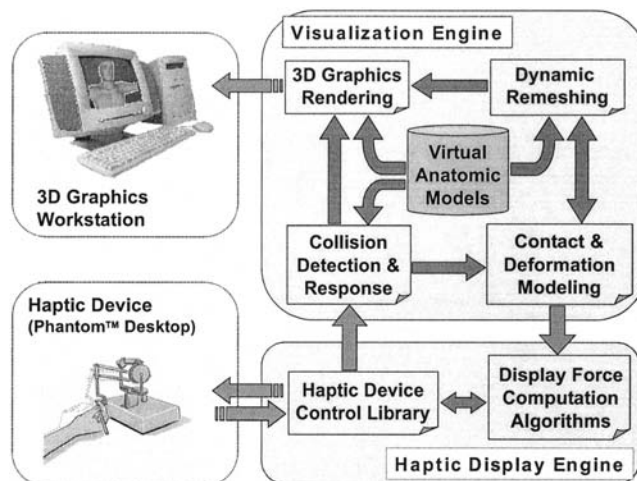


Figure 1. Overall architecture of the virtual simulation platform integrating haptic display.

2 The VR Paracentesis Simulation System with Haptic Display

2.1 General Description—Platform Architecture

The core of the first prototype simulation system is composed of a *virtual patient* with “visible” anatomic landmarks and “touchable” anatomic areas of interest. The user can: (a) orient the puncture device on a 3D anatomic space and select the puncture site, and (b) feel the haptic equivalent produced during the insertion through the anatomic layers, as well as during contact with clavicle and thorax.

Figure 1 depicts the overall architecture of the prototype platform, showing the two main subsystems with their functional modules, which are:

1. The *visualization engine*, which includes all software routines for stereoscopic visualization and interaction with the virtual patient 3D models. Its main modules include:
 - A database of virtual anatomical models, which includes 3D surface representations of textured skin, arterial/venal trees, lungs, and bones of the human body, as well as models of surgical instruments (in our case the needle parts),

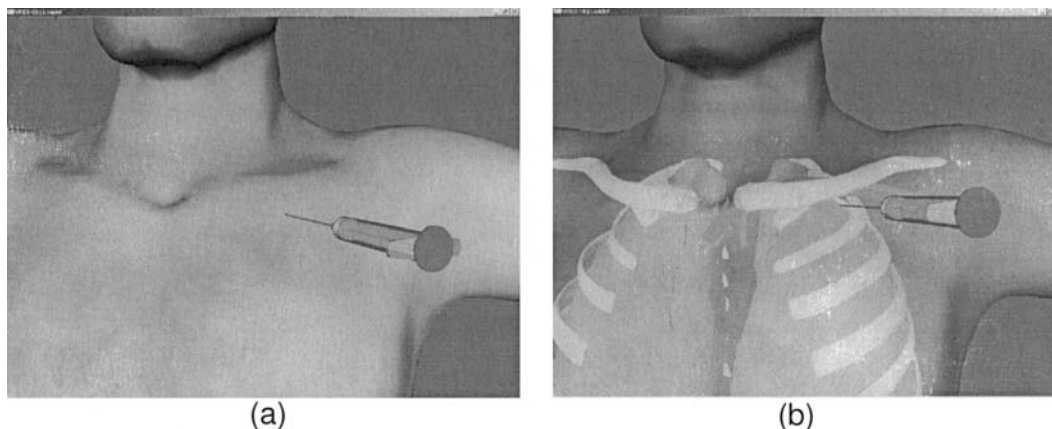


Figure 2. Virtual paracentesis simulation snapshots: (a) solid and (b) alpha-transparency rendering.

- Stereoscopic 3D graphics rendering routines, in our case implemented using OpenGL,
 - Collision detection routines, to detect in real time the contact point and intersection between the manipulated surgical instrument (needle) and the human anatomy models in VR space, and
 - Interactive simulation routines for soft tissue deformation (in this case, human skin deflection during needle insertion), and implementation of a dynamic remeshing algorithm using a level-of-detail technique.
2. The *haptic display engine*, which includes:
- Fast implementation of force computation models, based on local contact geometry approximations, and
 - Haptic interaction, force feedback, and control routines; in the first prototype platform, the simulator is coupled with a generic force-feedback PHANToM Desktop device from SensAble Technologies (<http://www.sensable.com/>), interfaced using the Ghost API.

2.2 Interactive 3D Visualization Engine

Figure 2 shows two screen snapshots of the virtual paracentesis simulation, including 3D renderings of the human skin in solid and alpha-transparency mode (with visible bony structures, lungs, and part of the vascular

tree involved in the considered paracentesis operation). In the current simulator configuration, the system continuously monitors the trajectory of the needle, and its intersection with the surface of the skin and inner organs, to enable the automatic computation of performance evaluation measures based on the simulation of all potential complications in accordance with the situations that may be encountered in the real clinical setting, as will be described more in detail later in this paper.

The first step in achieving a visually realistic and physically accurate paracentesis simulation is to develop models for the deformable tissues and anatomical structures of the human body involved in this invasive procedure. Elastic deformation modeling of soft tissues is a wide research area in the field of 3D computer graphics and animation, with many applications particularly in VR-based systems. The two most widely employed generic methodologies are based on: (a) mass-spring models, where the deformable object is modeled as a lattice of point masses interconnected by spring/damping elements; and (b) finite element (FE) models, which are based on the discretization of a continuum elasticity model (Gibson & Mirtich, 1997).

The use of FE models is typically preferred for surgical simulation applications (Bro-Nielsen, 1998; DiMaio et al., 2003), mainly because these models are tuned

more intuitively than mass-spring nets. However, to enable real-time interaction with 3D FE models, long preprocessing steps are required, imposing additional constraints related to shape changes and deformation limits. The use of “intelligent model simplifications” for particular applications may overcome these problems in a task-specific context. In this framework, we have developed a simplified 2D (planar) finite element model to simulate deformation of the skin under point load. More details on this approach can be found in Tzafestas, Koumpouros, and Birbas (2004a, 2004b). It must be pointed out here, though, that in the pilot study presented in this paper, visual rendering of the skin deformation during needle insertion was disabled, in order to eliminate any force-substitution effects that this may add to the interactive simulation (since our focus in this study, as will be more clearly described in the rest of the paper, was on the effect of a direct haptic component).

Another important issue here is real-time collision detection between moving objects in the virtual scene. In the first prototype platform used during the pilot study of this paper, the necessary routines have been implemented based on the “ColDet” 3D collision detection library for generic polyhedra, which is freely available on the Web.¹ The 3D surface models of the human body (outer skin), as well as those for the different anatomical structures (inner organs) involved (venal tree, muscle groups, lungs, and bony structures), are adapted from commercially available 3D models of human anatomy,² which are imported in an OpenGL/GLuT application using a specially developed loader for obj (standard ascii) formatted files (for portability reasons).

2.3 Haptic Display

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

realism of the simulation and the physically-based 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 modeling 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. Similar 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, for example, Colgate, Stanley, and Brown (1995) and Mahvash and Hayward (2005). It is needless to say, though, that such a soft coupling between any 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 pres-

¹<http://photoneffect.com/coldet/>.

²For the time being, we use models available at <http://www.cacheforce.com>.

ence 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 of primary importance, even if it is to the disadvantage of the accuracy of the geometrical representations used in force feedback. This was also subsequently validated during experimental trials through qualitative assessments performed by experienced surgeons. Such local force models, or linearization techniques, have also been proposed in other similar contexts, for example, in Kim, Otaduy, Lin, and Manocha (2003). Of course, different approximation models need to be applied for different simulation scenarios, also keeping in mind the particular requirements of the haptic skills involved in each specific application.

Regarding now the needle reaction forces applied by the skin along the needle insertion direction, we assume inhomogeneity characteristics in depth (stiffness coefficients increasing with the deformation magnitude u), resulting in a nonlinear stiffness coefficient that follows a typical exponential formula:

$$K = K_0[e^{m(u-u_0)} - 1] + K_{\min} \quad (u > u_0)$$

This term is used to model the instantaneous elastic component of force response for soft tissues exhibiting viscoelastic behavior (such as human skin) under compressive loads. The variable u_0 is the value of small deformation below which (i.e., when $u \leq u_0$) stiffness is experimentally observed to have a minimum constant value K_{\min} . The terms K_0 and m are constants that can be determined from available experimental data. Following the experimental analysis reported in the study of Pawluk and Howe (1999), who validated a similar exponential model for the specific case of human finger-pad deformation under flat-plate loading, we have set parameter m to a value of 2 mm^{-1} (which corresponds, in fact, to the mean value of m obtained from the experimental data reported in the aforementioned study), with K_0 scaled down appropriately to a value of $K_0 = 0.01 \text{ N/mm}$ (since we treat the case of a point load instead of a flat-plate load). This gives a force response of

approximately 5 g for 1 mm indentation. These values should increase empirically by a factor of approximately 2 and 4 for line and flat plate loading, respectively.

These results are also consistent with data presented in Dandekar, Raju, and Srinivasan (2003).

This model is applied until the needle punctures the skin surface, which is simulated to occur when the normal force (or equivalently the skin deformation) exceeds a (variable) threshold. When this occurs, the second phase of the paracentesis procedure is initiated, which involves (as will be described more in detail in the following section) needle reorientation and insertion into subcutaneous tissues toward the deep vessel. In this case, the applied force is computed as a sum of: (a) a variable viscosity term for the normal force, to simulate the penetrating motion of the needle through various underlying tissue layers, (b) a linear stiffness term for tangential reaction forces from the skin, and (c) a reaction force component from rigid contact of the needle with bony structures, which in this paper involves a model of the clavicle bone, as discussed above. Local geometry approximation techniques are also applied to speed up force computation, as already mentioned above, and involve: (a) a planar model for needle-skin interaction, and (b) a cylindrical local model to simulate the needle-clavicle interaction. More details on the force computation models can be found in Tzafestas et al. (2004b). In the rest of the paper we focus on the first pilot study conducted with this system, on the methodology employed and on the results obtained.

3 Pilot Evaluation Study: Methodology and Results

To evaluate the effectiveness of a VR medical simulator as a training tool, long-term clinical studies are required, especially if the goal to be assessed is the capacity of such a system to enhance users' surgical skills. Nevertheless, an initial proof-of-concept can be established if we validate the reliability of such a system as a tool, not directly for training and enhancing surgical skills, but for objectively assessing the clinical skill and dexterity of the user. Indeed, performance evaluation in

terms of objective skill assessment can provide an initial 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 (Aucar, Groch, Troxel, & Eubanks, 2005; Acosta & Temkin, 2005; Moody, Baber, Arvanitis, & Elliot, 2003), such as, for instance, studies demonstrating the effectiveness of the MIST-VR laparoscopic simulator (Gallagher, McClure, McGuigan, Crothers, & Browning, 1999; McNatt & Smith, 2001).

This section focuses on the first pilot study conducted with the virtual paracentesis system presented in this paper, and describes the metrics defined to objectively assess user performance, as well as the methodology employed and the results obtained.

3.1 Objective Metrics for Surgical Dexterity and Skills Assessment

In today's clinical practice, there are no specific objective evaluation measures to assess reliably the performance and skill of a clinician; evaluation is in fact based principally on the actual final clinical outcome. There is of course a correlation between the clinical outcome (successful or unsuccessful result and clinical complications) and the performance of the doctor, but not in an absolute manner. Furthermore, to draw safe conclusions regarding the skill of a clinician in a specific operation requires performing this operation in a sufficient number of patients, and then analyzing the statistics of the clinical outcomes. Therefore, to achieve our objectives related to the computation of objective assessment scores by the simulator, and conduct this pilot study as specified, it was necessary to devise a procedure

to quantify performance of the user during the whole simulated clinical act.

There are several surgical techniques used to perform subclavian vein paracentesis (each one with its advantages and disadvantages). For the purposes of this pilot study, one specific technique, as described in surgery textbooks, was selected and implemented consistently throughout the study, to exclude any effect that surgical technique variations can have on performance. According to this paracentesis procedure, 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 syringe is gently withdrawing until a free flow of blood appears in the syringe.

This technique was analyzed into two phases: (1) skin puncture, and needle penetration along the normal direction into an appropriate depth, and (2) needle reorientation, followed by the needle directed steadily toward the deep vessel. A set of critical parameters were defined in each step of the procedure, as a means to evaluate a set of objective performance measures, which include the following:

1. 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,
2. Deviations from an optimal needle path including: selecting the needle insertion point, as well as needle orientation in the venal access track,
3. "Economy of movements" measures, as an indicator on the smoothness of needle motion, and
4. 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

Table 1. A Partial Snapshot of a Sample Log File, with System Generated Scores

USER_ID	FF	COL_LUNG	COL_RIB	COL_ART	COL_VEIN	PHASE_1	PHASE_2	ECON_MEAN	ECON_VAR	TIME	SUCCESS
1	1	0	0	0	1	6	0	18	27	31524	1
USER_ID	FF	COL_LUNG	COL_RIB	COL_ART	COL_VEIN	PHASE_1	PHASE_2	ECON_MEAN	ECON_VAR	TIME	SUCCESS
8	0	0	25	0	9	4	982	12	51	157840	1

each one of these score categories, as well as a total assessment score, at the completion of each experimental trial.

It must be pointed out here that the “optimal” needle insertion path, mentioned above, cannot be defined in a unique way and depends on the specific surgical technique employed to accomplish the operation. In this pilot study, the subjects were instructed to perform the simulated subclavian vein paracentesis employing consistently the specific surgical technique described above, in order to evaluate performance in a systematic way (the simulator is of course general and can be used to teach and assess any related surgical technique). An important issue concerns, then, calibrating the system to obtain a parametric model of this optimal needle insertion path. For this reason, we conducted a set of training sessions where three experienced surgeons were asked to reproduce this technique (in simulation) in a sufficient number of times (at least 50 times each). These trials were in fact used to train the system, that is, to extract a statistical model of the critical parameters for each phase of the procedure, including: (a) needle insertion point and penetration depth, for the phase-1 of the procedure, and (b) needle orientation and direction along a line segment for the phase-2 of the procedure. The data collected during these training sessions were, thus, processed to define an optimal mean and acceptable deviations for this parametric representation, based on which subsequent computation of the respective objective assessment measures during the trial sessions of the pilot study was performed. It must be noted here that the surgeons who used the system to train the skill-assessment module did not participate in the subsequent pilot evaluation study.

A partial snapshot of a sample log file generated by the system is shown in Table 1, indicating the main

score values for two sample trials, with and without force-feedback ($FF = 1$ and $FF = 0$, respectively), where:

- **USER_ID**: the unique identification number of each subject (user),
- **COL_LUNG**; **COL_RIB**; **COL_ART**; **COL_VEIN**: the scores (penalty marks, or error indicators) related to unnecessary collisions (i.e., potential injuries) with the following vital anatomical structures, respectively: lungs, ribs and skeletal structure, arteries, and subclavian vein, and
- **TIME**: completion time (successful, if $SUCCESS = 1$, unsuccessful if $SUCCESS = 0$).

To the above scores, we add the performance measures related to:

- The “optimality of needle insertion track”: **PHASE_1** and **PHASE_2**, estimated as a static error with respect to the optimal path, in terms of initiating the two phases of this procedure, where phase 1 corresponds to the accurate selection of the needle insertion point, and phase 2 to the appropriate re-orientation of the needle after insertion into the correct depth, and
- The “economy of movement” indicators: **ECON_MEAN** and **ECON_VAR**, that is, mean value and variance, respectively, of the continuously recorded needle position and direction deviation with respect to the approximate optimal path (parameterized linear segment for phase 2 of the procedure).

All these performance measures are then weighted together to obtain an overall (total) score indicator for each performed procedure. The exact numerical values for the weights used in the computation of the total

Table 2. Numerical Values of Scoring Weights Used in the Metrics for Skill Performance Assessment

COL_LUNG	COL_RIB	COL_ART	COL_VEIN	PHASE_1	PHASE_2	ECON_MEAN	ECON_VAR	TIME
500	50	50	10	2	0.2–0.4	2	20	0.01

score during the present pilot study are shown in Table 2. Regarding Time, for instance, a weight factor of 0.01 is used, meaning that a score value of, for example, 30,000 (that corresponds to 30,000 ms, i.e., 30 s completion time) will contribute 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 (e.g., 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 (note that for the PHASE_2 score value, a margin of acceptable behavior was considered close to the mean optimum, leading to the definition of a nonlinearly increasing penalty score, which was approximated by a three-margin piecewise linear formula: `if (<1000) weight = 0; elseif (<2000) weight = 0.2; else weight = 0.4`). The score values obtained this way can then be seen, in fact, as normalized accordingly to reflect an average equivalent cost in terms of successful completion time (the choice made is of course empirical and reflects the expert knowledge of the surgeons collaborating in the design of this pilot study). 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.

In the rest of this section, we describe the experimental protocol and the procedures followed during the pilot study, as well as the results obtained.

3.2 Pilot Study Methodology: Experimental Protocol and Evaluation Procedures

The pilot evaluation study was conducted with two groups of users:

- Group I (novices): consisting of nursing graduates. The subjects selected to participate in this group should have no prior (direct or indirect) hands-on experience on the specific clinical procedure (that is, the subclavian vein paracentesis technique employed in this study), but with prior experience in the use of a syringe for simple needle insertion procedures on the human body. The nursing graduates selected to participate in group I of this study were qualified according to the above requirements.
- Group II (experienced): consisting of trained surgeons, selected based on their recent (within the last year) experience on this paracentesis technique (more than 100 operations performed), and on their qualification as trainers in such operations in the surgical clinics where they work.

A total number of 20 (N) subjects participated in this pilot study, 10 users in each group. Particular attention was paid 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 interface allowed the user to selectively switch between three prespecified fixed orthographic views of

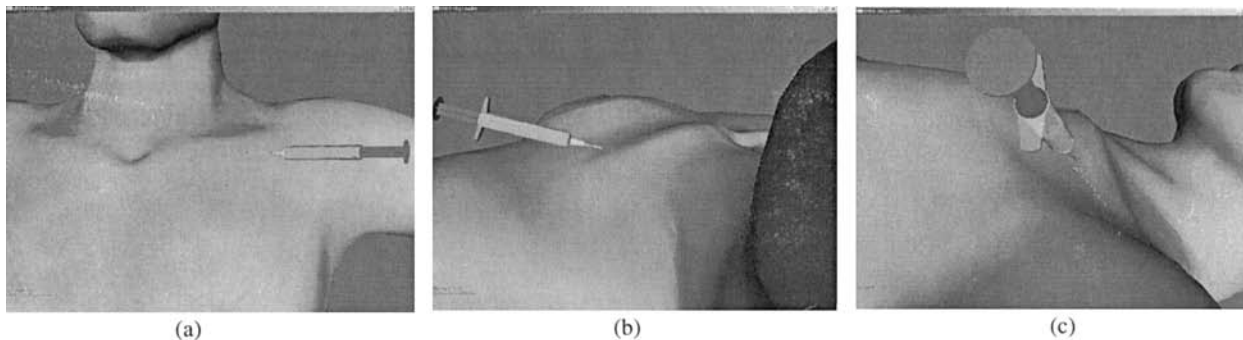


Figure 3. Orthographic 3D views of the virtual patient used during the simulated paracentesis operation.

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 (Figure 3). Furthermore, visual rendering of the skin deformation during needle insertion was also disabled, to eliminate any force-substitution effects that this may add to the interactive simulation, since our focus in the present pilot study is on the effect of a direct haptic component. In future studies, we plan to explore other multimodal and sensory substitution configurations.

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, that is, simulation without force-feedback (active force-feedback disabled).
- Condition (b): FF, that is, simulation with active force-feedback enabled.

A total 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 predefined level of difficulty and interaction features.

3.3 Evaluation Results and Discussion

The goal of the pilot study presented in this paper was to explore the correlation of the different system-generated performance scores (i.e., the dependent measures, automatically computed by the simulator at the end of each experimental trial) with:

1. The actual level of experience and surgical skill of the user, as this can be objectively deduced by the group to which the user belongs (group I: novices, group II: experts),
2. The activation or not of haptic feedback in the simulation, our goal being to study the extent to which the haptic display component plays an important role in increasing the performance of the simulation and training system. In other words, we aim to evaluate first the extent to which the objective scores computed by the simulator are consistent (in statistical terms) with the clinical skill of the user. Our goal is then to study the degree to which the presence (or absence) of a haptic (in this case, kinesthetic) display during the VR-simulated paracentesis procedures has any effect on the per-

formance of the system. Such an effect can be evaluated with respect to: (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 training basic surgical skills.

In line with these objectives, our interest during this pilot study focused on assessing two main research hypotheses:

Hypothesis 1 (Learning rate 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).*

This performance improvement (i.e., 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 second research hypothesis to be evaluated in the sequel.

To test the validity of our first research hypothesis, our goal was to assess an overall, average performance improvement in the course of each experimental session, that is, from the beginning (trial 1) to the end (trial 10) of each session. An indicative measure of this performance improvement, which should relate to the learning capacity of the system (if any), was obtained by comparing the mean values between the scores obtained during the first two trials (trials 1–2) and those of the last two (trials 9–10). This was used as a measure of an overall learning rate, for which results are depicted in Table 3.

This table presents the mean values and standard deviations for the scores obtained by group I users (time

Table 3. Learning Rate for Group I Users, Under Experimental Condition (a): NO Force-Feedback

GROUP I (a, NO)	Time (s)	Total score
Mean (trials 1–2)	71.5	1,918.99
SD (trials 1–2)	27.1	1,163.55
Mean (trials 9–10)	55.2	1,223.76
SD (trials 9–10)	42.3	767.60
Learning rate (%)	22.84%	36.22%

and total score), for experimental condition (a) (i.e. NO force-feedback). It is found that, for the time score, the learning ratio is approximately 22%, while for the total score this improvement is 36%. It should be noted, though, that (by applying a *t*-test comparison of the scores between trials 1–2 and 9–10) this improvement was not found to have a statistical significance in this case ($p = .162$ for the time). This result shows that, although there is an apparent performance improvement for group I users (in the NO experimental condition) in the course of each experimental session, this is not sufficiently strong and cannot be considered as a reliable, statistically founded result. This can of course be due to the wide variability of users' performance (as also seen by the high values of the standard deviations), which is natural in such a difficult (in the sense of delicate/dextrous) and complex (in the sense of combining many intermediate skills) simulation task, and which, given the limited number of users participating in this first pilot study, does not allow for a statistical proof to be obtained in this case. Nevertheless, a clear tendency, related to the learning rate of novice users in the system, is shown by this pilot study, as is also shown that without any haptic display (condition (a), NO) the learning rate exhibits a wide variability that limits a statistically reliable conclusion to be drawn with certainty.

It should be further noted that, on the contrary, as opposed to group I (novice) users, group II (experienced) users show no apparent learning gain at all, 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

Table 4. Learning Rate for Group I Users, Under Experimental Condition (b): FF, with Active Force-Feedback

Group I (b, with FF)	Time (s)	Total score
Mean (trials 1–2)	89.9	1,636.37
SD (trials 1–2)	41.4	1,447.34
Mean (trials 9–10)	46.3	859.32
SD (trials 9–10)	28.3	346.26
Learning rate (%)	48.42%	47.48%
<i>t</i> -test (<i>p</i>)	0.007	0.065

any apparent improvement rate in the learning curve, and this is irrespective of the activation or not of the haptic display. 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. Indeed, this can be attributed to an adequate realism of the simulator system, within which experienced users can intuitively perform the required procedures employing their normal skills and dexterities.

3.3.1 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 4, which concern group I users for experimental condition (b) (FF, i.e., with active force-feedback). Comparing the results in this table with the ones presented in Table 3, we can observe a performance improvement of approximately 48% for time (with respect to the value of 22%, in Table 3), and 47% for the total score (as compared to 36% for experimental condition (a) above). Furthermore, the learning ratio for time is now found to have a statistical significance ($p < .01$), while this was not the case for the results obtained in the NO experimental condition, as discussed above. This improvement provides an initial indication on the positive impact that the activation of

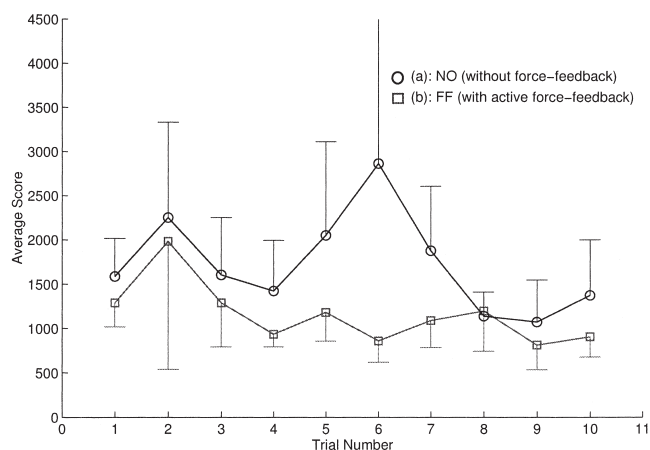


Figure 4. Experimental results for group I users: Average score over all users for each experimental trial, comparatively for the two experimental conditions. (a) without force-feedback (NO), and (b) with active force-feedback (FF).

haptic display (experimental condition (b)) has on the performance of the system, as evaluated here with respect to the learning rate for novice users.

The results concerning the learning curve for group I users, comparatively for the two experimental conditions, are depicted in Figures 4 and 5, for: (i) the average total score and (ii) the time-to-complete, respectively. These figures show 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. These data demonstrate: (i) a consistently better performance (in terms of the average total score) when force feedback is active (FF), and (ii) an overall tendency for a steeper learning curve in the case of experimental condition (b), that is, when the haptic display is active. This improvement is more apparent in the case of completion time, in Figure 5 (where statistical significance is also established, as also shown before, in Table 4), and 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 *SD* values for trials 9–10 in Table 4 being reduced as compared to the corresponding values

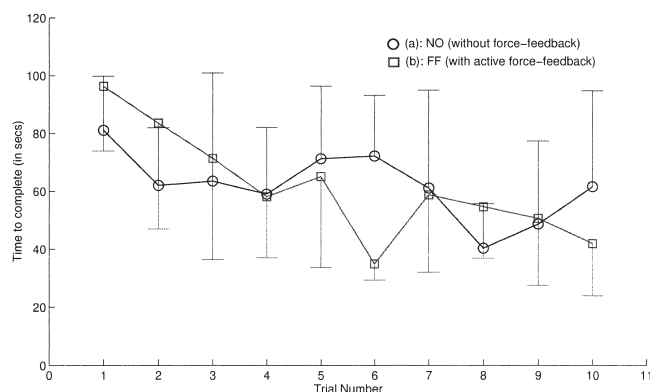


Figure 5. Experimental results for group users: Average time to complete (s) over all users for each experimental trial, for the two experimental conditions, (a) without force-feedback (NO), and (b) with active force-feedback (FF).

in Table 3). These results provide an initial indication of 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. This finding is particularly interesting if it is interpreted as referring to a more efficient skill acquisition process for novice users (trainees), when haptic display (active force-feedback) is enabled. Of course, larger-scale experimental studies need to be conducted before more persistent conclusions can be drawn reliably, with respect to the beneficial effect of haptic display on the learning curve. Furthermore, these simulator-generated score values need to be correlated with the actual skills that are needed to perform the clinical procedure in reality, which is in the scope of our second research hypothesis.

Hypothesis 2 (Skill assessment):

The system has the potential to assess reliably the dexterity and skill level of the user in an objective way, by means of the performance scores automatically computed after each experimental trial. The simulator-generated performance scores must correlate positively with the user group (that is, group II experienced users must perform significantly better—lower scores—than group I novice users).

The validation of this hypothesis can indeed support the direction of developing such realistic, VR-based,

Table 5. Overall Results for Group II Users, for the Two Experimental Conditions (a) Without (NO), and (b) With Active Force-Feedback (FF)

Group II	Time (s)	Total score
Mean value, (a): NO	53.4	1,580
SD, (a): NO	35.4	900
Mean value, (b): FF	32.7	753
SD, (b): FF	14.7	430

interactive simulation technologies as a tool for objective assessment and, potentially, accreditation of the user's competence and level with respect to specific clinical skills. To explore the validity of this hypothesis, the results obtained in this pilot study were analyzed comparatively for the two user groups. The overall results obtained for group II users (for the time-to-complete and total score values) are given in Table 5. Comparing these results with the results of group I above (even with best group I performance of trials 9–10 in Table 4), 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: 32.7 s (group II), as compared to 46.3 s (for group I); $SD = 14.7$, for group II, vs. 28.3 for group I.
- Average total score: 753 (group II) vs. 859 (for group I); $SD = 430$, for group II, and $SD = 346$ for group I.

Therefore, these results (comparison of means) demonstrate an apparent performance improvement for group II (experienced) users with respect to group I (novice) users, providing an initial indication about the skill assessment performance of the system. However, this conclusion is found to have statistical significance only when haptic display is active (condition (b): FF), as further explained in the sequel.

3.3.2 The Effect of Haptic Display on Skill Assessment. Figure 6 illustrates the overall comparative results between group I and group II, for the two

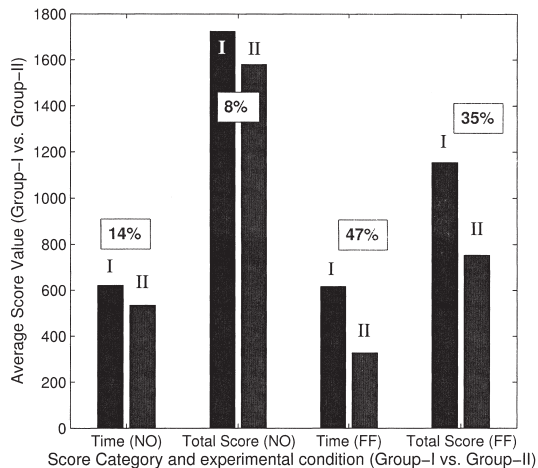


Figure 6. Overall comparative results between group I and group II, for the two experimental conditions: (a) without (NO) and (b) with force-feedback (FF).

experimental conditions (that is, with and without haptic display). As shown in this graph, the performance improvement for 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 group II users is apparently much more important in the case of experimental condition (FF), that is, when haptic display 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, as shown by the *t*-test results in Table 6 ($p < .01$ for the total score, and $p < .001$ 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 (group I vs. group II

users). Furthermore, what is even more interesting, is that this important correlation seems to appear only when haptic display is active. This finding supports our research hypothesis about the significant importance of haptic display in such interactive surgical simulator and training systems. 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 evident when one observes the time-to-complete score, which shows a particularly significant improvement ($p < .001$) when activating haptic display in the simulator, meaning that in this case experienced surgeons seem to become “familiar” with the experimental setting and perform the simulated surgical procedure with “confidence” and success. In this case, one can conclude that the 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.

The same conclusion can be drawn analyzing some more qualitative results, which have been collected by means of a questionnaire that the users were asked to complete after each experimental session. The experienced users (group II) have clearly expressed their feeling about the usefulness and necessity of active force-feedback. The associated haptic sensation conveyed by the system, although artificial, was stated to provide an irreplaceable feature for the realism of the interactive simulation, which seemed absolutely necessary for the experienced users to intuitively reproduce the dexterous manual operations that are required to safely and efficiently perform the simulated paracentesis procedure.

4 Conclusion and Future Work

This paper focused on the first pilot study conducted on a virtual reality (VR) based simulation and training system designed for a class of paracentesis operations. The specific case study presented treated a special surgical technique used for the catheterization of the subclavian vein. Two user groups participated in the study, namely: (group I) novice users and (group II)

Table 6. *t*-Test Results for Comparative Analysis Between the Two User Groups

Group I vs. Group II	<i>t</i> -test (<i>p</i>)
(a), NO (without force-feedback): time to complete	0.149304
(a), NO (without force-feedback): total score	0.292364
(b), FF (with active force-feedback): time to complete	0.0009
(b), FF (with active force-feedback): total score	0.0053

experienced surgeons. The simulator system automatically provides quantitative assessment scores on user performance, applying a set of objective metrics involving also measures regarding optimality of needle insertion path, economy of movement, as well as indicators of maneuvering errors or potential major complication risks (in terms of injuring vital anatomic structures). Performance, in terms of training and skill assessment, was evaluated in a twofold manner, regarding respectively: (a) the learning curve exhibited by users, and (b) the correlation of the system-generated scores with the actual surgical experience of the user, in other words, the extent to which the system can reliably predict and assess the actual skill level of the user, as related to the group to which each user belongs.

These performance measures were also evaluated with respect to the activation of haptic feedback and to whether this has any beneficial effect (or not). Our main objective in this study was to obtain quantitative evaluation results regarding the impact of haptic display on the performance of the paracentesis simulator (particularly with respect to the kinesthetic feeling evoked when the manipulated needle is constrained by the clavicle bone, which is commonly exploited by skilled surgeons as a “haptic guide” to control the needle insertion track).

The results obtained during this pilot study are presented and analyzed in this paper. The conclusions drawn support our basic hypotheses regarding learning and skill assessment performance of the system, and can be summarized in the following statements:

1. The system seems to simulate the considered paracentesis operations realistically, as can be deduced by analyzing the performance of experienced

(group II) users who perform the required manual procedures intuitively, reproducing the associated clinical dexterity and skills (no apparent learning gain for group II users in the course of each experimental session).

2. Initial findings with respect to the learning curve of group I users seem to support the idea that the proposed simulator can indeed be used by novice students or trainees to acquire basic skills and enhance their performance.
3. Statistical analysis of the results, comparatively with respect to the two user groups, supports our hypothesis that such a VR-based simulator can be used to reliably assess the clinical skill and experience of the user, based on objective metrics and performance scores automatically generated by the system.

Regarding particularly the beneficial effect of haptic display on the performance of the system, the following conclusions can be drawn:

4. The presence of the haptic display considerably improved the learning rate for novice users, leading to a steeper and more consistent learning curve.
5. Activation of the haptic display had a statistically significant effect on skill assessment performance of the system, as deduced by a comparative analysis of the scores obtained between novice and experienced users.

It was, thus, the goal of this research work to explore aspects of this beneficial effect and to obtain relative quantitative evidence, at least in a specific simula-

tion and training configuration. We must always bear in mind that this study concerned synthesizing an artificial kinesthetic feedback, where the issue is to explore how the users may react to the presence (or absence) of such a virtual component in a VR-based simulator. Quantitative results based on objective evaluation measures, particularly with respect to the effect of haptic/kinesthetic display, are in fact rather missing from the literature, and it was our goal to add relative evaluation data. A long-term goal is to characterize skills and to establish rules about how different dimensions in a simulation/training process can benefit from specific feedback modalities, in order to obtain guidelines for the design of multimodal training interfaces. Furthermore, it must be noted that larger-scale, long-term clinical studies are required before safe conclusions can be drawn regarding the use of such a simulator as an efficient training tool.

Besides that, future work is also oriented in the following directions:

- From a technical perspective, further development is needed particularly with respect to: (a) exploring more complex 3D deformation and force-feedback models, also studying haptic rendering stability and transparency issues, (b) improving the realism and the impact of the simulation by integrating bimanual coordination and operation in a coregistered virtual simulation space, and (c) implementing other training modes, particularly computer-guided training.
- From an application point of view, further development is needed to increase the scope of the simulator as a dynamic training platform, by: (i) integrating more complex and parameterizable 3D models, as well as patient-specific models reconstructed from CT and/or MRI data, and (ii) implementing other use cases and clinical scenarios, to create a more general-purpose platform.

Feedback collected from users (nursing graduates, MDs, and experienced surgeons) who have tested the system 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 per-

formance) specific, particularly challenging requirements that are imposed within the medical education and training sector.

Acknowledgments

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

References

- Acosta, E., & Temkin, B. (2005). Haptic laparoscopic skills trainer with practical user evaluation metrics. *Proceedings of Medicine Meets Virtual Reality 13, Studies in Health Technology and Informatics, 111*, 8–11.
- Aucar, J. A., Groch, N. R., Troxel, S. A., & Eubanks, S. W. (2005). A review of surgical simulation with attention to validation methodology. *Surgical Laparoscopy, Endoscopy & Percutaneous Techniques, 15*(2), 82–89.
- Baur, C., Guzzoni, D., & Georg, O. (1998). VIRGY: A virtual reality and force feedback based endoscopic surgery simulator. *Proceedings of Medicine Meets Virtual Reality (MMVR'98)*, 110–116.
- Bro-Nielsen, M. (1998). Finite element modeling in surgery simulation. *Proceedings of the IEEE, 86*(3), 490–503.
- Burdea, G. (1996). *Force and touch feedback for virtual reality*. New York: John Wiley & Sons.
- Burdea, G., & Coiffet, P. (2003). *Virtual reality technology* (2nd ed.) New York: John Wiley & Sons.
- Colgate, J. E., Stanley, M. C., & Brown, J. M. (1995). Issues in the haptic display of tool use. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'95)*, 140–145.
- Dandekar, K., Raju, B. I., & Srinivasan, M. A. (2003). 3-D finite-element models of human and monkey fingertips to investigate the mechanics of tactile sense. *Journal of Biomechanical Engineering, Transactions of the ASME, 125*, 682–691.
- DiMaio, S. P., & Salcudean, S. E. (2003). Needle insertion modelling and simulation. *IEEE Transactions on Robotics*

- and Automation: Special Issue on Medical Robotics, 19(5), 864–875.
- Frisoli, A., Borelli, L. F., Stasi, C., Bellini, M., Bianchi, C., Ruffaldi, E., et al. (2004). Simulation of real-time deformable soft tissues for computer assisted surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 1(1), 107–113.
- Gallagher, A. G., McClure, N., McGuigan, J., Crothers, I., & Browning, J. (1999). Virtual reality training in laparoscopic surgery: A preliminary assessment of minimally invasive surgical trainer virtual reality (MIST VR). *Endoscopy*, 31(4), 310–313.
- Gerovich, O., Marayong, P., & Okamura, A. M. (2004). The effect of visual and haptic feedback on computer-assisted needle insertion. *Computer-Aided Surgery*, 9(6), 243–249.
- Gibson, S. F., & Mirtich, B. (1997). A survey of deformable modeling in computer graphics. *Tech. Rep. TR97-19, Mitsubishi Electric Research Laboratories (MERL)*, Cambridge, MA. Available at <http://www.merl.com/reports/index.html>.
- Gorman, P., Krummel, T., Webster, R., Smith, M., & Hutchens, D. (2000). A prototype haptic lumbar puncture simulator. *Proceedings of the Medicine Meets Virtual Reality Conference*, 106–109.
- Heng, P. A., Cheng, C. Y., Wong, T. T., Xu, Y., Chui, Y. P., Chan, K. M., et al. (2004). Virtual reality based system for training on knee arthroscopic surgery. *Proceedings of the 12th Annual Medicine Meets Virtual Reality Conference*, 130–136.
- Kim, Y. J., Otaduy, M. A., Lin, M. C., & Manocha, D. (2003). Six-degree-of-freedom haptic rendering using incremental and localized computations. *Presence: Teleoperators and Virtual Environments*, 12(3), 277–295.
- Kühnapfel, U., Çakmak, H. K., & Maasz, H. (2000). Endoscopic surgery training using virtual reality and deformable tissue simulation. *Computers and Graphics*, 24(5), 671–682.
- Mahvash, M., & Hayward, V. (2005). High-fidelity passive force-reflecting virtual environments. *IEEE Transactions on Robotics*, 21(1), 38–46.
- McNatt, S. S., & Smith, C. D. (2001). A computer-based laparoscopic skills assessment device differentiates experienced from novice laparoscopic surgeons. *Surgical Endoscopy*, 15(10), 1085–1089.
- Moody, L., Baber, C., Arvanitis, T., & Elliot, M. (2003). Objective metrics for the evaluation of simple surgical skills in real and virtual domains. *Presence: Teleoperators and Virtual Environments*, 12(2): 207–221.
- Nienhuys, H.-W., & van der Stappen, A. F. (2003). Interactive needle insertions in 3D nonlinear material. *Tech. Rep. UU-CS-2003-019*, Institute of Information and Computer Sciences, Utrecht University. Available at <http://www.cs.uu.nl>.
- Okamura, A. M., Simone, C., & O’Leary, M. D. (2004). Force modeling for needle insertion into soft tissue. *IEEE Transactions on Biomedical Engineering*, 51(10), 1707–1716.
- Pawluk, D. T. V., & Howe, R. D. (1999). Dynamic lumped element response of the human fingerpad. *Journal of Biomechanical Engineering, Transactions of the ASME*, 121, 178–183.
- Sutton, C., McCloy, R., Middlebrook, A., Chater, P., Wilson, M., & Stone, R. (1997). MIST VR—A laparoscopic surgery procedures trainer and evaluator. *Studies in Health Technology and Informatics*, 39, 598–607.
- Tzafestas, C. S. (2003). Whole-hand kinesthetic feedback and haptic perception in dexterous virtual manipulation. *IEEE Transactions on Systems, Man and Cybernetics (Part A)*, 33(1), 100–113.
- Tzafestas, C. S., Koumpouros, Y., & Birbas, K. (2004a). Haptic interaction in VR-based paracentesis simulation for dexterity enhancement and assessment. *Proceedings of the 18th International Congress and Exhibition on Computer Assisted Radiology and Surgery (CARS 2004)*, 413–418.
- Tzafestas, C. S., Koumpouros, Y., & Birbas, K. (2004b). Paracentesis modeling and VR-based interactive simulation with haptic display for clinical skill training and assessment. *Proceedings of the International Conference on Integrated Modeling & Analysis in Applied Control & Automation (IMAACA ’2004)*, 13M: International Mediterranean Modeling Multi-conference.
- Webster, R. W., Zimmerman, D. I., Mohler, B. J., Melkonian, M. G., & Haluck, R. S. (2001). A prototype haptic suturing simulator. *Proceedings of the Medicine Meets Virtual Reality Conference*, 567–569.
- Zhang, J., Payandeh, S., & Dill, J. (2003). Haptic aided design: A case study. *Proceedings of the 11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (HAPTICS 2003)*, 254–261.