The Role of Simulation Fidelity in Laparoscopic Surgical Training

Hyun K. Kim[1], David W. Rattner[2], and Mandayam A. Srinivasan[1]

The Touch Lab,
[1]Department of Mechanical Engineering and The Research
Laboratory of Electronics
Massachusetts Institute of Technology, Cambridge, MA 02139
hyunkim@mit.edu
[2]Division of General and Gastrointestinal Surgery
Massachusetts General Hospital, Boston, MA 02114

Abstract. Although there have been significant advances in the development of virtual reality based surgical simulations, there still remain fundamental questions concerning the fidelity required for effective surgical training. A dual station experimental platform was built for the purpose of investigating these fidelity requirements. Analogous laparoscopic surgical tasks were implemented in a virtual and a real station, with the virtual station modeling the real environment to various degrees of fidelity. After measuring subjects' initial performance in the real station, different groups of subjects were trained on the virtual station under a variety of conditions and tested finally at the real station. Experiments involved bimanual pushing and cutting tasks on a nonlinear elastic object. The results showed that force feedback results in a significantly improved training transfer compared to training without force feedback. The training effectiveness of a linear approximation model was comparable to the effectiveness of a more accurate nonlinear model.

1 Introduction

Minimally invasive surgery has revolutionized surgical procedures over the past few years, due to quick recovery time and minimal scarring and pain. However, the method of training for MIS still requires continuous personal supervision of the trainee by an expert. Development of a VR-based surgical simulator proposes to provide a better method of training that improves on the current method of apprenticeship [1,2]. However, there still remains an underlying question about the level of fidelity of the simulation required for effective training to occur.

How realistic does the simulation need to be for effective training? On the one hand, it is possible to learn the surgical practices with an unrealistic model, which may lead to negative training transfer. However, because of the learning abilities and perceptual limitations of the sensory, motor, and cognitive system of the human user, perfect simulation is unnecessary. Furthermore, given the large variations in human anatomy and physiology, there is no single perfect model, and wide variations exist in geometry and material properties of the organs. The main question is how simple a simulation can we get away with, while at the same time preserving a level of fidelity

between the virtual and real organ behavior that leads to positive training transfer. Morever, this question becomes more important if one considers the difficulties in simulating tool-tissue interaction very accurately in real time.

There has been training transfer studies done for assembly tasks [3], pick and place tasks [4], and flight simulations [5] in virtual reality. However, a study of measuring training transfer under various levels of fidelity in surgical simulations with force feedback has not been conducted up to this point. The difficulty lies in measuring the improvement of real world surgical skills from virtual environment training. Measuring surgical skills within the virtual environment can be done relatively easily. However, some of these skills might only pertain to the virtual environment alone and the training may have no significant or even negative effects on real world skills. Therefore, what is truly required is the measurement of improvement in real world skills, but the surgical trainers that are available currently are not mature enough for such a study to take place in a clinical setting. A current method that is widely used to validate low fidelity simulators is to measure the inverse transfer of training. This involves comparing the performance of an expert surgeon with a novice surgeon on the simulator, and if the expert surgeon performs significantly better the simulator is deemed to be effective [6]. However, such inverse arguments are not sufficient enough to conclude that the simulator gives positive training transfer. The study described in this paper provides an alternate method for measuring training effectiveness directly.

2   Experimental Platform

The new paradigm involves a two-station experimental platform. The two stations are laparoscopic surgical stations with analogous surgical tasks. In one station the surgical tasks are performed on virtual objects whereas the other station contains instrumented real world objects. The real world surgical scene consists of simple inanimate objects, for which the shape and material properties are well known. This allows the real objects to be very accurately modeled in the virtual environment. Therefore, an opportunity to degrade the fidelity of the simulation from very high to low fidelity is provided.

In the setup the real environment station would be the testing station and the virtual environment setup acts as the training station. Subjects would be evaluated on the real environment test station before and after virtual environment training. Figure 1 shows the setup for the experimental platform. Both stations are covered with a rubber model of the abdominal wall (Limbs and Things, Inc.). Two laparoscopic tools are inserted through the abdominal wall via trocars, as done in real laparoscopic procedures. In the real station, the objects that are to be operated on are instrumented and are placed inside the abdominal area. A laparoscope is also inserted through the abdominal wall to provide visual feedback on to a CRT screen. In the virtual environment station the surgical tools are connected to two Phantom (SensAble Technologies) haptic interface devices. The objects are generated in the virtual world by the computer (450MHz Pentium III PC) and displayed to the user graphically through the computer monitor and haptically through the Phantom devices. Both the real station and the virtual station contain foot petals to activate markers on the end of the surgical tools that simulate harrmonic scalpels.
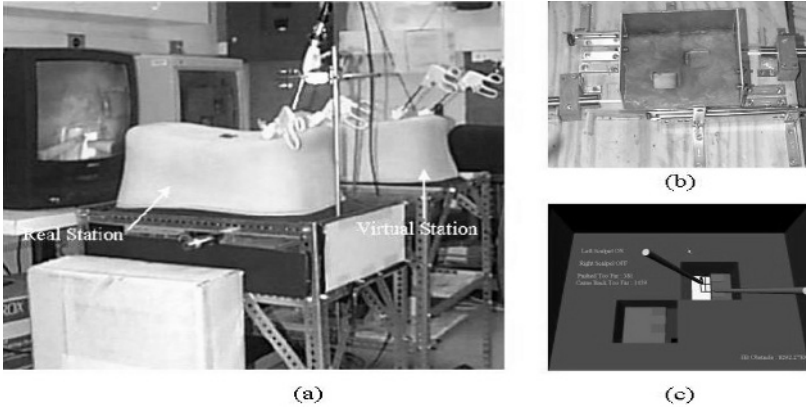
Fig. 1. (a)Experimental platform, (b) surgical scene in the real environment and (c) surgical scene in virtual environment.

2.1   Design of Task

The surgical tasks that were implemented in the two stations were selected with relevancy to laparoscopic surgery in mind. Although the operations were performed on inanimate objects and not human organs, the actions had to be very similar to what is done in real laparoscopic surgery. Additionally, the chosen task needed to provide a graded "mimicability" of the real tasks in the virtual environment. Choosing tasks and objects that could be modeled to varying degrees of accuracy, both haptically and graphically, was important in order to have various levels of fidelity. Also, The tasks needed to have an appropriate level of difficulty so that a number of training sessions would be necessary. Finally the tasks had to have a significant number of metrics so that the performance of the subject can be quantified during both training and evaluation.

A bimanual pushing and cutting task, similar to what is done in Heller's myotomy was chosen. The scene consists of three layers. The top layer is the obstacle and two spring-mounted blocks are underneath it. The task is to avoid the obstacle, push and hold the blocks at a certain location and then make an incision mark with the freehand tool on the uncovered bottom layer. The incision mark is made on a designated grid and has to be consistently thin, straight, and consistent in depth. The task is bimanual in that one hand has to maintain the spring-loaded block in a finely defined position, while the other hand has to make accurate incision marks. The action is analogous to the pulling with a gripper and cutting with a harmonic scalpel action that is central to Heller's myotomy. Figure 2 shows the instrumented real environment surgical scene and the virtual surgical scene.

2.2   Performance Measurements

A set of performance metrics was required to quantify the performance of the subjects. Traditionally, the measurement of surgical skill is more qualitative than

quantitative. However, for this study, a numeric measurement of performance was necessary to assess the relative levels of skill improvement. In fact, there have been several prior studies done on establishing metrics for laparoscopic surgery [7,8]. For this study, a customized set of metrics was established to fit the performance criteria of the particular tasks. The total performance was defined as a linear, weighted summation of various sub-scores, defined by the equation,

$$Performance = s_1\tilde{T} + s_2\tilde{P} + s_3\tilde{C} + s_4\tilde{O} \tag{1}$$

where $\tilde{T}$, $\tilde{P}$, $\tilde{C}$ and $\tilde{O}$ are the normalized scores of time, push accuracy, cut accuracy and tool control, respectively. Push accuracy was quantified by measuring the block position and velocity, the cut accuracy measured the straightness, depth, depth consistency and accuracy of the cut mark. Tool control measured the number of collisions with the obstacle and accuracy of tool position. An expert laparoscopic surgeon determined the weights, s_1-s_4, by assessing what he considers is important in successful surgery. Push accuracy and cut accuracy were given the highest weighting because these were tasks where tool coordination and depth perception was important. Also these two metrics were the quantities that described the success or failure of the main objective of the tasks. The time to completion is not as important in most laparoscopic procedures, since they are generally not emergency procedures. Therefore, the time taken to complete the tasks was given a relatively smaller weighting.

3   Experimental Design

For the first experiments on this platform, the role of haptic fidelity on training transfer was investigated. One of the key questions that remain about haptic fidelity is how accurately nonlinearity of material elasticity needs to be modeled. Characterizing the in-vivo force-displacement properties of organs is not a trivial problem and efforts are continuously being made to measure tissue properties accurately. However, it is not known if the user using the surgical simulator can actually distinguish between a linear and a non-linear elastic model. Even if they could, the difference in training effectiveness may not be significant.

For this investigation a non-linear spring was loaded on to the sliding blocks in the real station. The force-displacement data for the spring and block assembly was measured and plotted on Figure 3. The figure shows a typical behavior observed in most tissue material where the stiffness increases as the displacement increases. The stiffness of the nonlinear spring was comparable to that measured for in-vivo pig esophagus by Tay et al. [9]. For the fidelity study, two approximations were made for this curve. A linear approximation and a nonlinear approximation was made using a least-squares curve fitting technique. Therefore three different levels of haptic fidelity were possible in representing the spring in the training station: the nonlinear approximation was the most accurate high fidelity model, the linear approximation was the medium fidelity model and the lowest fidelity was a model without force feedback.

For the sensory resolution of the hand, the JND for force varies from 5-15% depending on the force magnitude, muscle system and experiment method [10]. The

resolution deteriorates at forces below 0.5N, with a minimum resolution of 0.06N at these small forces [11]. The maximum force difference in the operating region between the chosen linear and nonlinear approximation is 0.16N. While performing the task, this difference is amplified three to four times due to the long tool being pivoted near the handles. Also, the stiffness difference between the two models varies from 0-70% in the operating region. The JND for stiffness is known to be 23% [12]. Therefore, if a simple discrimination experiment were being done, the subject would most likely be able to distinguish between the nonlinear and linear elastic model.

For the sensory resolution of the hand, the JND for force varies from 5-15% depending on the force magnitude, muscle system and experiment method [10]. The resolution deteriorates at forces below 0.5N, with a minimum resolution of 0.06N at these small forces [11]. The maximum force difference in the operating region between the chosen linear and nonlinear approximation is 0.16N. While performing the task, this difference is amplified three to four times due to the long tool being pivoted near the handles. Also, the stiffness difference between the two models varies from 0-70% in the operating region. The JND for stiffness is known to be 23% [12]. Therefore, if a simple discrimination experiment were being done, the subject would most likely be able to distinguish between the nonlinear and linear elastic model.

Twenty-four subjects with no prior surgical training were tested in the experiments. The subjects were given the same detailed description of the tasks and scoring. Then they were initially evaluated at the test station performing the described tasks. The subjects were then divided into eight initial skill level groups depending on their initial performance score. One subject from each skill level group was treated by one of the three training treatments. The training for each subject lasted seven sessions. Each session involved completing the task on the virtual trainer from the beginning to the end. Once the training was complete the subjects were brought back to the real environment test station to be evaluated for the final time. The measure of training effectiveness was determined by the improvement in skill between the initial evaluation at the real station and the final evaluation, also on the real station.
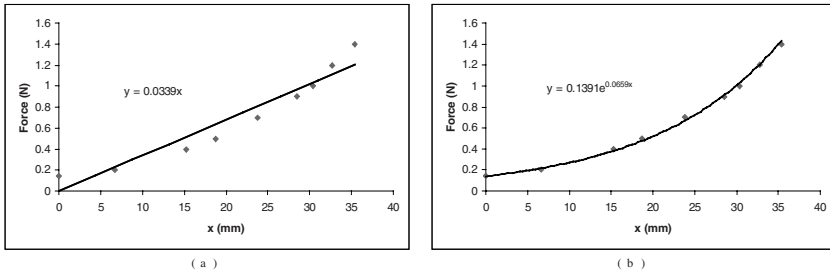

Fig. 2. (a) Linear approximation to force displacement curve, (b) Nonlinear (exponential) approximation to force displacement curve

4   Results

The mean performance improvement of each training treatment is plotted in figure 4 with the mean improvement in each sub-metric also plotted to the right of each total

improvement score. Error bars on the bar chart represent the interval for which a Tukey pair-wise comparison would show significant difference between the means.
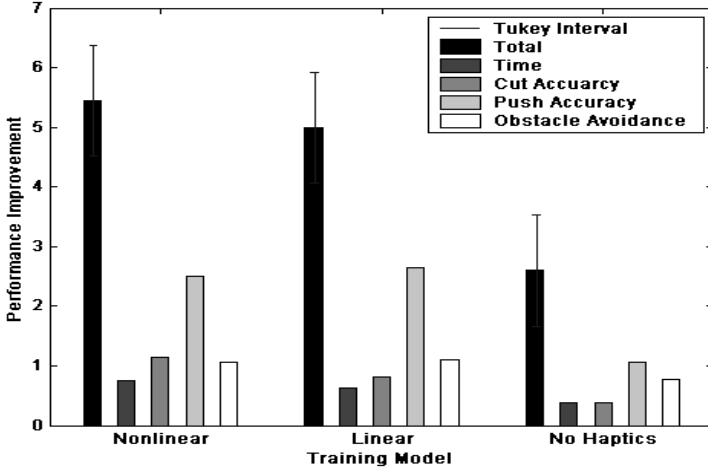

Fig. 3. Average training effectiveness plot

4.1   Observations

There is a considerable difference in the total score improvement between the haptic training modes and the training without haptics. There seems to be no significant difference between the score improvements for the nonlinear elastic and linear elastic models. The improvement in each sub-metric score was also comparable for the two haptic training modes. Among the sub-scores, the cutting and pushing accuracy seemed to deteriorate the most when force feedback was absent.

4.2   Statistical Analysis

A two factor analysis of variance was conducted on the results. It was assumed that the increase in performance was influenced by the training treatment received and the initial skill level of the subject, as shown by the equation,

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \qquad (2)$$

y_{ijk} is the performance increase for a subject from the initial skill level group i, who underwent training treatment j and has repetition index k. μ is the common effect, α_i is the training effect of trainer i, β_j is the effect of the subjects initial skill level group j, γ_{ij} is the interaction between factors α and β, and ε_{ijk} is the uncontrolled variation for this specific subject. Using this relationship two hypotheses

were set up. The first hypothesis was that the effects of all four training treatments were the same.

$$H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 \tag{3}$$

The second hypothesis, H_1, was that the effects of the five initial skill level groups, β_j were the same.

The analysis of variance test could be done under the assumption that the increase in skill level is normally distributed in the population of subjects. The ANOVA was conducted with the aid of MATLAB's statistical toolbox. H_0 could be easily rejected at 5% significance ($p=0.026$) as well as H_1 ($p=0.0001$).

A pair-wise comparison using a Tukey test for the three training treatments was also conducted. The training effect of the trainer was denoted as τ, and the Tukey test was done to see if each training treatment was significantly different from one another.

Table 1. Tukey pair-wise comparison

	$\tau_{nonlinear} = \tau_{linear}$	$\tau_{nonlinear} = \tau_{no\_haptics}$	$\tau_{linear} = \tau_{no\_haptics}$
p	0.81	0.003	0.011

The ANOVA test shows that the effects of the training treatments are significantly different. Therefore, the different levels of haptic fidelity in the spring model have affected the training. The pair-wise comparison shows that the training effect of the training without haptics was significantly lower than that of the training with force feedback. Also, there was no significant difference between the training transfer of the nonlinear elastic model and the linear elastic model. The statistical analysis agrees with the direct observations from Figure 4.

5   Concluding Remarks and Future Work

The study shows force feedback appears to be crucial for effective surgical training for procedures involving tasks contained in this study. The skill level improvement for the trainers with haptic feedback was significantly higher than that of the trainer without haptics. There are currently many surgical simulations that rely solely on graphical feedback without haptics. These systems have their advantages such as simplicity of hardware, slower computation time requirements and costs. Also, the results here show that a virtual trainer without force feedback would in fact give a positive training transfer on average. However, the results also show that haptics enhance the training effect considerably and that the performance is improved on a more consistent basis. Also, training with force feedback seemed to become more important for surgical tasks that require more accuracy and delicate tool control. Therefore, the results seem to support the idea that an effective surgical simulation is one that provides both graphical and haptic feedback.

The training effect of the linear elastic and the nonlinear elastic model were not significantly different for these tasks. This suggests that modeling the nonlinear

elasticity of tissue is not important for the type of tasks employed in this study. A simple linear approximation of the stiffness seems to result in training effects that are not significantly reduced, for stiffnesses that are comparable to that of organ tissue. Therefore, it seems accurate characterization of in-vivo nonlinear force-displacement properties is not always necessary in training for surgical pushing and pulling tasks. Due to the hardware limitations, it is possible that the additional forces on the surgical tools from the friction and inertia of the haptic interface and the resistive forces of the abdominal wall are large enough for the small differences in the stiffness between linear and nonlinear models to have any significant effect on training.

References

1. R. M. Satava and S. B. Jones, "Virtual Environments for Medical Training and Education", Presence, vol. 6, pp. 139–146, 1997.
2. C. Basdogan, C. Ho, and M. A. Srinivasan, "Virtual Environments for Medical Training: Graphical and Haptic Simulation of Laparoscopic Common Bile Duct Exploration", IEEE/ASME Transactions on Mechatronics, vol. 6, pp. 269–285, 2001.
3. R. Adams, D. Klowden, and B. Hannaford, "Virtual Training for a Manual Assembly Task," *Haptics-e*, vol. 2, 2001.
4. J. Kozak, P. Hancock, E. Arthur, and S. Chrysler, "Transfer of training from virtual reality," *Ergonomics*, vol. 36, pp. 777–784, 1993.
5. T. Carretta and R. Dunlap, "Transfer of effectiveness in flight simulation: 1986 to 1997," : Air Force Research Laboratory, NTIS, 1998.
6. N. Taffinder, C. Sutton, R. Fishwick, I. MacManus, and A. Darzi, "Validation of Virtual Reality To Teach and Assess Psychomotor Skills in Laparoscopic Surgery: Results from Randomised Controlled Studies Using the MIST VR Laparoscopic Simulator," presented at Medicine Meets Virtual Reality, 1998.
7. J. Rosen, M. MacFarlane, C. Richards, B. Hannaford, and M. Sinanan, "Surgeon-Tool Force/Torque Signatures - Evaluation of Surgical Skills in Minimally Invasive Surgery," presented at Proceedings of the MMVR Conference, 1999.
8. R. O'Toole, R. Playter, T. Krummel, W. Blank, N. Cornelius, W. Roberts, W. Bell, and M. Raibert, "Assessing Skill and Learning in Surgeons and Medical Students Using a Force Feedback Surgical Simulator," presented at MICCAI, Cambridge, MA, 1998.
9. B. K. Tay, S. De, N. Stylopoulos, D. W. Rattner, and M. A. Srinivasan, "In vivo Force Response of Intra-abdominal Soft Tissue for the Simulation of Laparoscopic Procedures" , Proceedings of the MMVR Conference, 2002.
10. M. A. Srinivasan, "Haptic Interfaces, In Virtual Reality: Scientific and Technical Challenges," *Report of the Committee on Virtual Reality Research and Development*, N. Durlach and A. Mavor, Eds.: National Research Council, National Academy Press, 1995.
11. S. Kilbreath and S. Gandevia, "Neural and biomechanical specialization of human thumb muscles revealed by matching weights and grasping objects," *Journal of Physiology*, vol. 472, pp. 537–556, 1993.
12. L. A. Jones and I. W. Hunter, "Influence of the Machanical Properties of a Manipulandum on Human Operator Dynamics;Part 1. Elastic Stiffness," *Biol.Cybern*, vol. 62, pp. 299–307, 1990.