COST-PERFORMANCE TRADE-OFFS IN HAPTIC HARDWARE DESIGN

by

IMAN BROUWER

Ir, Delft University of Technology, 2000

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

DEPARTMENT OF MECHANICAL ENGINEERING

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 2004

Library Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Iman BROUWER
Name of Author *(please print)*

25 / 08 / 2004
Date (dd/mm/yyyy)

Title of Thesis: Cost-Performance Trade-Offs in Haptic Interface Design Hardware

Degree: M.A.Sc   Year: 2004

Department of Mechanical Engineering
The University of British Columbia
Vancouver, BC   Canada

Abstract

The objective of this research was to determine whether low performance haptic hardware leads to the same surgical task performance compared to more expensive hardware in a virtual reality surgical simulator for laparoscopy. VR surgical simulators are currently being introduced in leading teaching hospitals around the world. While they provide great potential for improvement over current laparoscopic skills training methods, a major barrier to large-scale acceptance is their high cost. Therefore this study is performed to determine whether a reduction in quality and therefore cost can be obtained without affecting surgical task performance.

To perform user test at different levels of haptic quality, we developed software that can introduce friction, cogging, force saturation, inertia, and backlash into the haptic loop, simulating the characteristics of less expensive hardware on a high-end haptic interface. This software avoids the need for an expensive hardware redesign, while it allows varying the different parameters on a continuous scale, independent from each other, and within a realistic range.

In a pilot study expert surgeons performed a clip application (2 participants) and a dissection task (3 participants) on commercial haptic hardware and VR laparoscopic software. Each surgeon performed the task(s) 3 times under 5 different settings while forces and kinematic data were recorded. Two settings were picked from the ones mentioned above; the other three consisted of the unaltered high fidelity setting, zero force feedback, and a combination of force saturation, cogging, friction, and inertia. We compared tissue-interaction forces, velocities, tool-tip path lengths, and completion times between the high fidelity setting and each of the other 7 settings. At a significance level of 0.05 a Friedman test showed that only the no force feedback condition was significantly different from the high fidelity condition in applied 95 percentile forces. In the clip application task both 50 percentile, and 95 percentile forces were significantly different in the no force feedback condition compared to high fidelity. None of the other settings showed significant differences in any of the performance measures.

These preliminary results suggest that low performance components can be used in haptic hardware for laparoscopy without affecting task performance, potentially creating a significant cost reduction.

Table of contents

List of Figures

List of Tables

Acknowledgements

Chapter 1   Introduction

Currently surgical residents acquire their laparoscopic skills mainly in the OR. In search of a more suitable training environment, leading teaching hospitals are currently acquiring computer based surgical simulators. These simulators provide an interactive visual simulation of the surgery, while the more advanced systems also let the users feel the forces that they exert on the tissue with the instruments. Research has shown that the ability to feel these forces makes a difference in how surgeons execute the task [Moody, 2002; Wagner, 2002]. Unfortunately, systems with force-feedback are expensive, creating a barrier for hospitals to acquire these devices. In this project we would like to investigate the relation between force-feedback quality and user performance, to formulate new design requirements that may lead to a more cost-effective design of hardware for laparoscopic surgical simulation.

This chapter will introduce the important concepts involved in this study: in Section 1 we introduce minimally invasive surgery, computer-based simulation with force-feedback, and human haptic and multi-modal perception. In Section 2 we discuss the main roles of simulation in surgical applications: training, performance evaluation, and practice on patient specific anatomy. The main cost contributing factors in haptic hardware design will be introduced in Section 3; the components of the hardware, what designs are currently used, and the influence of market size and other factors. Finally, in Section 4 previous work on the relation between haptic quality and task performance will be reviewed and Section 5 will state the objectives of this thesis research.

1.1 Minimally Invasive Surgery and Computer Simulation

1.1.1 Minimally Invasive Surgery

In open surgery, a surgeon manipulates the tissue at the target site directly by hand. To provide access for the hands a fairly large incision has to be made. Minimally invasive surgery (MIS) techniques are employed in an effort to decrease the amount of tissue damage for access purposes. Special long and thin instruments are used that are inserted into the body through small incisions (Figure 1-1). The surgeon operates these instruments from outside the patient.
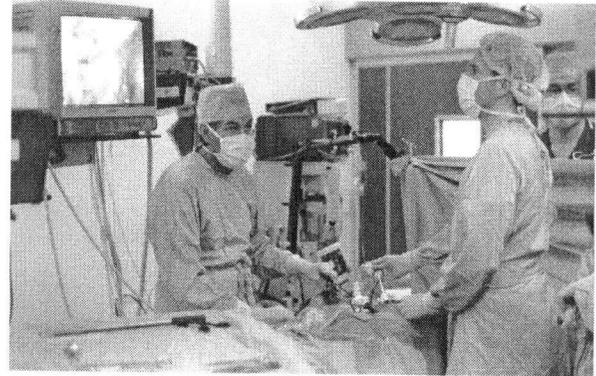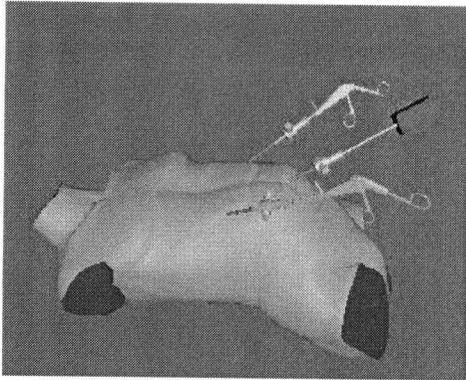
Figure 1-1: Minimally invasive surgery: Placement of the instruments and trocars relative to the abdominal wall (left), and an overview of a minimally invasive surgery setup in the OR (right).

Laparoscopic surgery is MIS applied to the abdominal area. During laparoscopic procedures, the abdominal wall is inflated to provide a tent-like workspace. After inflation, access ports, called canula, are inserted into the abdominal wall to provide airtight entry ports for the instruments. To provide visual access to the inner structures, a camera is inserted through one of these canula as well. This camera is usually not directly controlled by the surgeon but held by an assistant. The surgeon looks on a monitor to view the camera image (Figure 1-1). The book by [Berci, 1998] gives a good overview of clinical aspects of MIS.

Minimally invasive procedures have important benefits compared to their open counterparts. Due to the decrease in tissue damage necessary to access the site of operation, these procedures are usually associated with shorter recovery times and less scar tissue. As a result, minimally invasive surgery has seen an explosive growth in the last decade. By 1995 60-80% of all abdominal surgical procedures in the United States were conducted laparoscopically [Taylor, 1995]. This percentage is expected to increase with advances in technology, allowing more complex procedures to be performed minimally invasive as well.

1.1.2 Haptic Perception and Multimodal-Task Performance

The word 'haptic' is defined as relating to the sense of touch. It is derived from the Greek verb haptesthai: to touch. Haptic perception is derived through two main mechanisms:

- Tactile perception through the receptors just beneath the skin. Tactile perception is used to feel textures on surfaces, small vibrations, and for initial contact detection.

2

- Kinaesthetic perception is acquired through receptors in the muscles and tendons. From these receptors information about forces on the muscles and body position is obtained. The just noticeable difference (JND) in force perception is about 7% [Tan, 1994]. Our kinaesthetic perception is better at estimating relative forces than absolute forces. The JND in joint angle position were reported by Tan for the wrist (2.0°), elbow (2.0°), and shoulder (0.8°) [Tan, 1994]. If more than one joint is involved in perception of position, the error due to resolution will accumulate.

In laparoscopic surgery, the surgeon relies mainly on two senses: vision and touch. A very important question for this research is how surgeons integrate information from vision and touch. How much do they rely on vision and how much on touch? An important aspect of surgical skill is the ability to exert the right amount of force on tissue and to judge the stiffness. Work by Srinivasan and colleagues showed that visual information can strongly influence kinaesthetic perception of stiffness. In experiments to judge the stiffness of a spring that participants could feel but not see, participants relied more on indirect, visual information of the spring deformation then on the kinaesthetic haptic feedback [Srinivasan, 1996]. Ernst et al. performed experiments in which subjects were asked which one of two bars was higher. The bars could be perceived both visually and by touch. The setup allowed the experimenters to vary the amount of visual information and haptic information. Based on their results the authors proposed that humans integrate vision and haptics in a statistically optimal fashion, similar to a maximum likelihood estimator [Ernst, 2002]. Their hypothesis implies that the brain will assign more weight to information coming from the sense that provides the most reliable information at that time.

In minimally invasive surgery, the amount of force feedback is reduced compared to open surgery due to indirect touch and the low mechanical efficiency of the laparoscopic tools [Breedveld, 1999]. Stylopolous et al. suggest in [Stylopoulos, 2004] that surgeons have learned to adapt to the reduced force feedback conditions during laparoscopy in the OR by using something they call 'visual haptics': surgeons judge the amount of forces exerted on the tissue on visual cues as tissue color, contour, and adjacent tissue integrity and this is combined with a general gentleness. This would explain how expert surgeons are currently able to train residents how much force to exert on tissue while not being able to directly feel those forces.

1.1.1 Virtual Environments

This paragraph aims to give a short introduction to virtual environments. The reasons that virtual environments have been proposed for surgery will be discussed later in this chapter. There is no agreement on how to define a virtual environment (VE) exactly. A general definition is given by Webster's 2003 New Millennium Dictionary of English as "A computer-generated, three-dimensional representation of a setting in which the user of the technology perceives themselves to be and within which interaction takes place". A milestone in the creation of these environments is the Sensorama [Heilig, 1962] (Figure 1-3). It was developed during the 1950's by Morton Heilig, who would later invent the first head mounted display [Heilig, 1960]. The Sensorama consisted of a pre-recorded film in colour and stereo that was augmented with binaural sound, wind, scent as well as vibration experiences. Around the same time, a milestone in computer graphics was set by the work of Ivan Sutherland when he developed the Sketchpad (Figure 1-3): a computer that allowed an engineer to make a drawing that would be displayed on a CRT monitor and edit it while immediately seeing the results [Sutherland, 1963]. This was the first step in computer simulation, which can be described as the techniques of representing the real world in a computer program. While the Sensorama aimed at creating the most realistic multi-sensory experience, it was not interactive like the Sketchpad. But the Sketchpad did not have the immersive properties of a virtual environment yet.
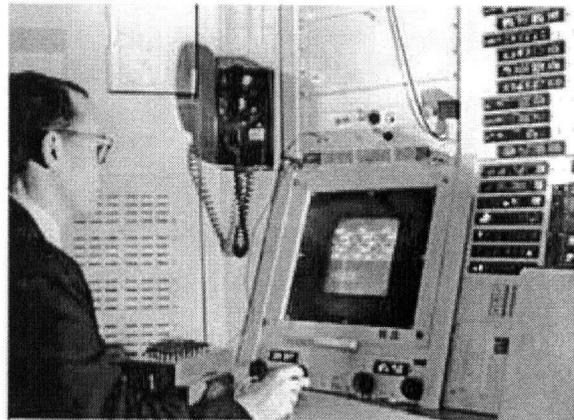
Figure 1-3: Two milestones in virtual reality: The multi sensory but non-interactive Sensorama (left) and the first graphical computer interface Sketchpad (right), operated by its inventor Ivan Sutherland.

With the enormous increase in computational power and the tremendous improvements in computer graphics in the last 40 years, it is now possible to simulate physics-based interactive 3D graphical environments (Figure 1-5). These environments have applications as far ranging as the treatment of anxiety disorders [Krijn, 2004], architectural education by recreation of historic buildings [Chan, 2003], engineering design [Sensable Technologies, 2004], and virtual training environments for virtually all kinds of surgery (e.g. laparoscopic [Tendick, 2000], [Basdogan, 2004], open abdominal [Bielser, 2003], and brain surgery [Larsen, 2001]).

Figure 1-5: A commercial simulator for minimally invasive surgery (left): two instruments and the camera are inserted into the simulator. The monitor shows the visual representation of the virtual organs interacting with the instruments while the instruments reflect the forces that can be felt by interacting with these organs. (source: www.xitact.com). The right picture shows an image of the virtual environment of another commercial simulator (source: www.simbionix.com).

1.1.2 Haptic Rendering

Haptic rendering is the process by which haptic cues are delivered to the person interacting with the application, to convey information about virtual haptic objects [Salisbury, 2004]. These can be cues reflecting the weight of objects, friction, texture, inertial forces etc. Haptic rendering is a relatively new field and there remain many unsolved problems. The main challenges are:

- Multi-contact interaction: In many applications today, there is only a single point or two points of contact interacting with the environment. To imitate real environments, multi-point, multi-hand, multi-person environments need to be developed. A great advantage of laparoscopic surgery is that there is no direct contact between the hands of the user and the VE: the indirect manipulation of the VE through the laparoscopic instruments greatly simplifies haptic rendering and is the main reason that there are so many more virtual environments available for laparoscopic surgery compared to open surgery.

5

- Level of detail: increase of level of detail increases the computational load of the VE to calculate forces, but also requires a high quality haptic interface to display subtle haptic cues.
- Stable rendering of simulations.

Stable rendering of simulations is a first requirement for any simulation but not an easy one to achieve. The cause of stability issues can be illustrated with Figure 1-7.
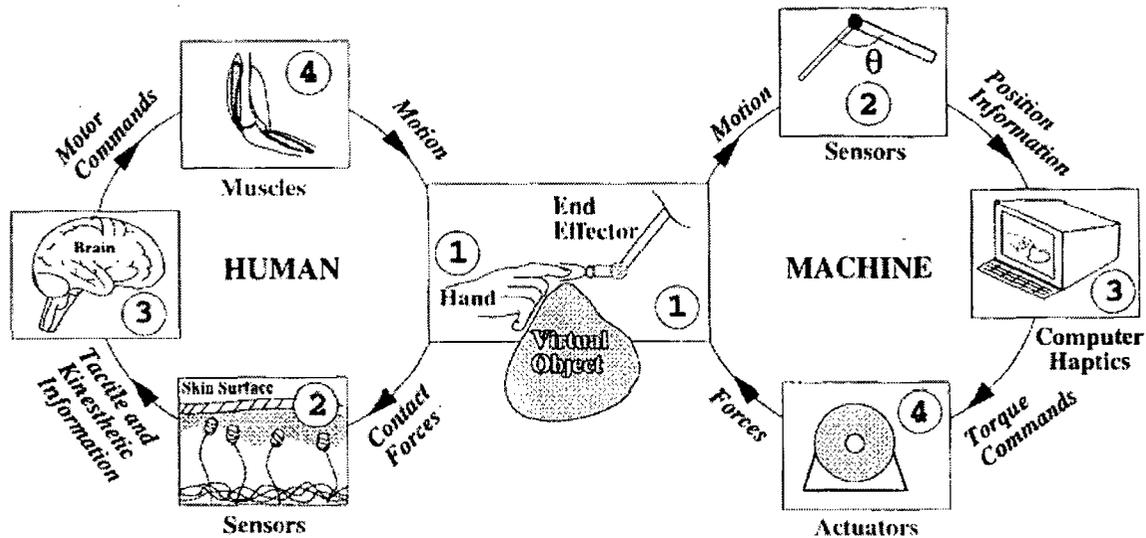
Figure 1-7: How do we interact with a haptic virtual environment? (source: [Srinivasan, 1997])

On the left side of Figure 1-7 is the human who is interacting with a virtual object. On the right side of the figure is the computer system with the haptic interface. Both systems are made up of a sensing part and an actuating part, with a delay in between. Position sensing in the haptic system is discrete in time. For a system to be stable, it needs to be passive: it should not be possible to extract energy from the system. Because position is only sampled at discrete times, the position is constant in between updates. The delay associated with this zero-order hold in position destabilizes the system by introducing energy. This can be illustrated by the common haptic example of exploring a virtual wall with a haptic interface. Let's assume that the wall is modelled with a spring. As a result of the time delay, the position of the probe in the VE will always trail behind its real position. This means that as one penetrates a virtual wall, the felt forces will trail behind the 'ideal' forces that would have to be overcome without the time delay, and the user has to perform less work than in the case of a real wall. On the way out of the wall,

the position and forces will trail again and this time the forces will be higher than in the case of a real wall. As a result, the work returned by the system will be higher than in a real system without the zero-order hold. Overall, the haptic system returned more energy than the user put into it. This is a very important issue in the modeling of inertia later in Chapter 2. A more detailed explanation of this phenomenon is given in [Gillespie, 1996]. To address this issue, update rates of 500 Hz are considered a minimal requirement for haptic simulation, though even higher update rates are required for environment that are very stiff or dynamic.

1.1.3 Haptic Interfaces

In 1971, the first force feedback system, GROPE I, was realized at the University of North Carolina (Figure 1-9). A robot was modified to allow a user to move it around freely in 2D. It would display forces present in a simulated environment to the user [Batter, 1971]. Unfortunately, this project was stopped for about 10 years because of lack of computational power available at the time. Force feedback became a major research area in the nineties and the first commercial haptic interface, the Phantom [Massie, 1994], was released on the market in 1995. More than 1600 of these devices had been sold by 2003 and it is currently the most popular haptic interface for professional use. Meanwhile force-feedback devices for consumers have outnumbered professional force feedback devices boosted by a considerable drop in price that made these devices available for less than US$100. More then 4 million force feedback gaming peripherals had been sold by October 2003 according to Dean Chang, Chief Technology Officer and Vice President of Technology Adoption and Partner Services at Immersion [Immersion Corporation, 2003].
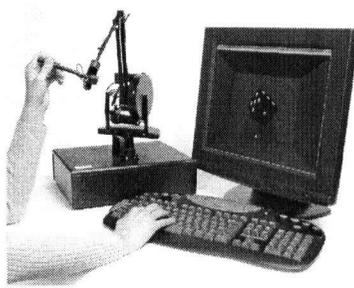
| GROPE I | Phantom | Joystick | LSW |
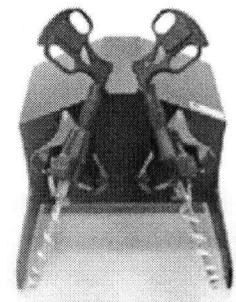
Figure 1-9: Examples of haptic interfaces (from left to right): The GROPE I is the first haptic interface, developed at the University of North Carolina in 1971. The Phantom is the first commercial haptic interface (released 1995). Commercial Joystick with force feedback and the Laparoscopic Surgical Workstation from Immersion Corp.

Multiple research groups at universities have developed haptic interfaces for laparoscopic surgery. An overview of designs by these groups and commercially available designs will be given later in this chapter.

Very basic tactile haptic interfaces have made their way into consumer electronics as vibrating cell phones, pagers etc. Essentially the same vibro-tactile technology is also available commercially on instrumented gloves. More complex technology aims at providing tactile actuation at a spatially higher resolution, usually in the form of pin arrays (e.g. [Moy, 2000; Siegel, 2002]). Though these displays are still a topic of active research, some displays are commercially available as for example Braille displays. Haptic interfaces that are designed to activate the tactile component of touch are to the best knowledge of the author not commercially available for surgical training at the time of writing.

1.1.4 Surgical Simulation: Advanced Virtual Environments with Haptics

Surgeons sometimes ask "Why do we only now start to have VE's for surgical training while pilots have had flight simulators for decades?". A simple answer is that surgical environments are extremely difficult to simulate. Face validity has been shown for 3 simulators in a recent review of 7 commercially available simulators for laparoscopic surgery [Schijven, 2003]. Still, there are many challenges ahead to improve realism.

Surgical VE's are computationally intensive for both graphics and haptics. They require a computer model that contains a geometrical, mechanical and graphical representation of the organs involved in the operation and their interaction with the instruments. Both the mechanical properties and graphical properties will be discussed in the next two paragraphs.

Mechanical properties of living tissue are much more complicated than common engineering materials. Living tissue is anisotropic, inhomogeneous, nonlinear, and viscous. Obtaining accurate mechanical models of living tissue is still an active research area [De, 1999; Mayrose, 2000; Brouwer, 2001]. A surgical simulation must repeatedly analyze the tissue structure with calculations similar to those commonly performed on man-made structures in mechanical or civil engineering. However, in surgical simulation the goal of accurately simulating mechanical properties competes with the necessity to provide interactive update rates. To provide a smooth haptic experience of the simulation, update rates of at least 500Hz are necessary. The 2 ms time allowed in surgical simulation to calculate forces based on the new state of the tissues is in sharp contrast to the more traditional application of structural analysis in engineering where computation times of hours to multiple days are acceptable. Therefore the current simulation models are optimized more for high update rates and low latency than for accuracy. Many variations on mass-spring-damper, finite element, and boundary element methods have been used to speed up computation times or increase realism of the mechanical model [Berkley, 1999; James, 1999b; James, 2002; Kim, 2002a; Zhang, 2002; Gosline, 2003b; Gosline, 2004][Berkley, 1999; James, 1999a; James, 2002; Kim, 2002a; Zhang, 2002; Gosline, 2003a; Gosline, 2003b; Gosline, 2004]. All of these methods greatly simplify the mechanical model of the tissue and often do not allow the tissue to be cut. Mahvash and Hayward developed a promising methodology that does not rely on real-time calculation of forces but instead retrieves pre-calculated forces during task execution. Forces can be obtained either by off-line simulation or by measurement in the real environment [Mahvash, 2004].

The graphical properties of a surgical environment are very complex as well. There are many challenging phenomena: blanching of tissue under tension, properties of a cut surface, realistic rendering of the flow of fluids and the reflection caused by the humid environment, complex

textures, smoke from cautery [Jensen, 2001], and visual properties of tissue after it has been cauterized.

While a lot of progress has been made in the past decade, there are still many barriers to overcome before surgeons can experience a simulation of surgery that is realistic in both touch and vision.

1.2 Roles of Simulation in Surgical Applications

Computer simulations of surgery could until recently only be found in research labs. Due to the advances in technology and the great market potential, there were 8 companies[1] offering computer-based surgical simulations for laparoscopy in 2004. This section will start with a discussion of the currently most prevalent application of computer-simulated laparoscopy: training. After that performance evaluation for certifying surgeons and testing and comparison of instruments and procedures will be discussed. This section will end with a discussion of a potential application that can be expected in the future: training and practice of a procedure on patient-specific data.

1.2.1 Training

Why use a simulator to learn minimally invasive surgery? In open surgery, technical skills are acquired according to the apprentice model: the student first watches an expert performing an operation and will gradually participate more and more in the procedure until the student can do the whole procedure by him or herself. While the apprentice model is still the most important component in minimally invasive surgery training, the general consensus is that training in the OR should be reduced in favour of training outside the OR. The main reasons are:

- Due to design of the instruments and the limitations of the camera-monitor system, MIS motor skills are more difficult to acquire than their open counterparts. Impeding factors include the limitation of moving the tool tip in only 4 degrees of freedom, mirroring of motion and forces at the tool tip with respect to the instrument handle, and the reduction of 3D information in vision [Breedveld, 1999, 2000].

---

[1] Immersion, Simbionix, Reachin, Surgical-Science, Xitact, Select-IT VEST, Mentice, Haptica

- Reducing risks to patients: Surgeons are more likely to make errors at the beginning of their career [Grantcharov, 2001].

- Reducing cost of training: OR time cost is expensive and average operating time is reported to increase by almost 50 percent when a resident is present [Traverso, 1997]. The cost of training a resident in the OR is estimated at a total of US$48,000 over 4 years. [Bridges, 1999].

- Guaranteeing exposure to certain kinds of less frequently encountered cases that usually do not show up during a resident's training. It can for example be of vital importance to recognize unusual anatomical configurations in time.

- Currently a resident will read 'the manual' for an advanced procedure before going into the OR to learn it. It may be beneficial to actually practice the new procedure in simulation before stepping starting to perform on patients [Soper, 2004].

The need to train outside the OR was underlined by the Society of American Gastrointestinal Endoscopic Surgeons (SAGES) Education Committee's creation of the Fundamentals of Laparoscopic Surgery (FLS) at the annual meeting in 1998 [Villegas, 2003]. It was proposed that controlled inanimate training environments outside the OR should replace initial training on patients. In [Stylopoulos, 2004] a survey was administered to a panel of thirty expert surgeons attending the 8th annual meeting of SAGES. Ninety-five percent of the participants said that skills training should be incorporated into the residency programs and medical curricula.

Currently, there are two forms of non-patient training available other than computer simulation: 'bench-top trainers' and animal labs. Bench-top trainers, or phantom boxes, consist of a box that has openings through which standard minimally invasive instruments can be inserted to manipulate inanimate objects that are placed inside of the box. A standard laparoscopic video setup can be used for visualization (Figure 1-11). While these simulators provide an environment to train basic motor skills and in some cases provide a representation of anatomical structures in the form of latex tissues, they lack the complex physiology of the human body. Physical 3D models of organs are often expensive and can often not be reused.
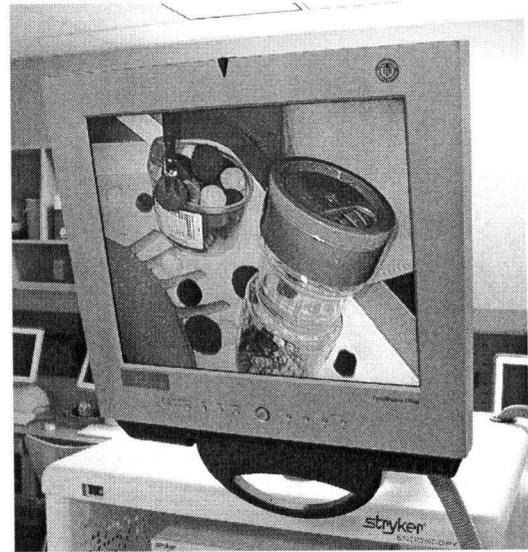
Figure 1-11: A student working on a 'bench-top trainer' (left): standard laparoscopic tools can be used to interact with inanimate objects. The right picture shows a close-up of the screen.

Training on porcine models provides a realistic physiological environment. While in 1993 its use was still recommended by some [Wolfe, 1993], ethical concerns have since been raised and led to a ban on using animals for surgical training in the United Kingdom [Lirici, 1997]. The costs associated with an animal training facility also make it a less attractive option today.

Trainers that are based on computer simulation technology have the potential to overcome many of the disadvantages that training with boxes and porcine models have. Computer-based simulators do not need a special facility; they can be used with minimal supervision, and at any time of the day. As the technology progresses and more data become available, the anatomy and physiology will be more realistic, it will also be possible to train on a specific anatomic anomaly or specific disease type. Because the movement of the instruments are tracked and the tissue interaction forces can be logged, automatic performance feedback can be provided and the student can be provided with automated personalized suggestions for improvement. Since all tissue is virtual, modules can be repeated as many times as necessary: students can focus on that aspect with which they have most trouble. For further reading on training for minimal invasive surgery, the article by Villegas *et al.* [Villegas, 2003] provides a good recent overview of its many aspects.

1.2.2 Performance Evaluation

Structured global rating forms are the current gold standard in evaluation of surgical performance. These assessments are performed on a videotape of the procedure in question. Expert surgeons evaluate the anonymous video, making judgements on issues such as dexterity, respect for tissue, spatial orientation, flow of the operation, assistant use, and knowledge of the procedure (e.g. [Reznick, 1997; Adrales, 2003]). While Adrales *et al.* reported that the surgeon-examiners shared a common perception of competence [Adrales, 2003], other researchers have found that expert surgeons often have different opinions on the relative importance of certain technical skills [Baldwin, 1999]. There are other disadvantage to global rating forms: they place a high demand on limited and expensive resources (expert surgeons), and they are not an objective assessment in the sense that individual human perception and opinion is still part of the analysis.

Another, relatively new, approach is quantitative performance analysis based on measurement of relevant aspects of surgery such as tool trajectories and forces. Algorithms are applied to the obtained data to extract performance measures. Though quantitative performance analysis may not be as comprehensive as expert surgeon's judgement, it has two main advantages:

- It is objective: The same algorithms can be applied to all data: no subjective judgements have to be made at any point.

- It does not require the time of highly skilled experienced surgeons. If the data is acquired during simulation, no human intervention is needed at all since a completely automated process can apply the algorithms to extract the performance indicators.

A wide variety of measures can potentially be extracted: At this moment evaluation algorithms are based on efficiency and evaluation of kinematics and global forces. The most commonly administered performance measure is completion time (e.g. [Chung, 1998; Derossis, 1998; Fried, 1999]). Some studies use penalty scores for errors made (e.g. [Derossis, 1998; Eubanks, 1999]), kinematic measures [Kinnaird, 2004], and force and torques [Rosen, 2001; Kinnaird, 2004] Future knowledge in areas like tissue injury mechanics can add to a more comprehensive evaluation of surgical skill on simulators [Schijven, 2003].

13

1.2.3 Practice on Patient Specific Anatomy

Not everybody has the same anatomy. While most people only show small variations in anatomy, there are certain outliers that can be the cause of increased risk during surgery if they are not recognized in time. In the future OR, it could be possible for a surgeon to actually practice a procedure on patient-specific data before performing the real procedure. Before the procedure the site of interest of the patient is scanned. From the images a 3D model is build to generate the VE to perform the procedure in. In a next step, the VR procedure could be recorded, and then optimized, with a robot finally performing the operation on the patient, with the surgeon only intervening when necessary [Satava, 2003].

1.3 Cost Contributing Factors in Haptic Hardware Design

As mentioned in the introduction, force feedback is a significant contributor to the overall costs of a surgical simulation system. For example, Immersion sells a two-instrument interface without force feedback for US$ 7,800, and a similar product with force feedback for US$ 28,000. What makes a haptic interface so expensive? For that we have to look at the major cost contributing factors:

- Components
- Labour / production facilities
- R&D
- Marketing
- Maintenance

1.3.1 Components

This section gives with an overview of the components of haptic hardware: motors, transmission, encoders, and electronics.

1.3.1.1 Motors

The motors are an essential part of the interface: they are the direct source of forces or torques that are exerted on the user. If all six degrees of freedom for each instrument are actuated in a laparoscopic interface with the standard 2 instruments, 12 motors are needed. Almost without exception, modern haptic interfaces make use of permanent magnet motors, either brushed or brushless. Cost of motors can vary greatly. While cheap motors as found in force feedback

joysticks can cost only a couple of dollars, high end motors used in many high-end haptic interfaces can exceed US$200 a piece. Motor companies contacted for this research were all leluctant to share pricing information and were only willing to specify prices for a few motors. The most important cost-contributing factors are torque output, cogging, friction, and inertia. These phenomena will be described in more detail in Chapter 2.

1.3.1.2   *Transmission*

The most commonly used transmission systems can be divided into three categories: direct drive, cable-based and geared. In direct drive, the rotational shaft of the motor is connected to the user-contact point of the interface either through a rigid connection (iDrive[3] , d'Groove [Beamish, 2004], Twiddler [Shaver]), or through a linkage system (e.g. a pantograph design). Figure 1-13 shows five examples of direct-drive haptic interfaces.
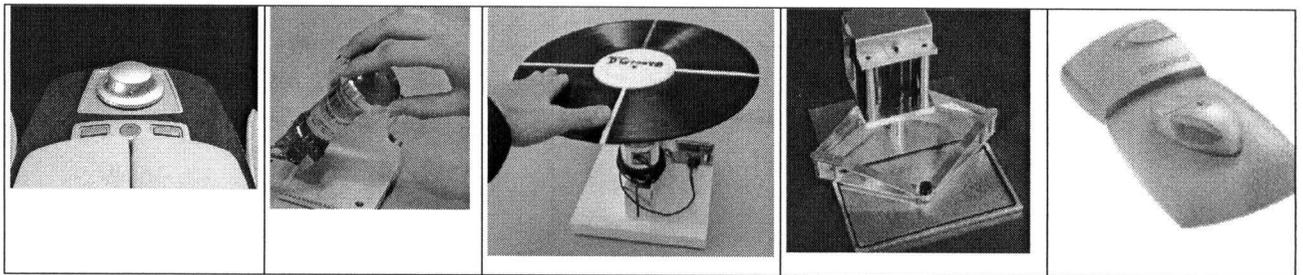
Figure 1-13: Five examples of direct drive haptic interfaces (from left to right): the BMW iDrive, the low-cost 1-DOF Haptic Twiddler, Tim Beamish's d'Groove digital turntable, a pantograph configuration by Vincent Hayward, and the Logitech Wingman a force feedback mouse with pantograph design (right)

In some designs gears are used to increase torque. This can add a significant amount of friction, backlash, and inertia and is used mainly in low cost devices like force feedback joysticks [Bin, 2002]. Some designs in force feedback joysticks use belts over directly interlocking gears to reduce backlash [Rosenberg, 2002b].

The pantograph design has been used in a consumer force feedback mouse (Logitech) that sold for below US$100 (Figure 1-13, right). Two important differences from the pantograph design by Hayward (Figure 1-13, second from right), is the absence of expensive machined components

---

[3] http://www.bmwusa.com/Joy/Drive/Technology/iDrive.htm

combined with cheaper motors in the Logitech design. The result is a haptic interface of noticeable lower quality, but also a significantly cheaper device.
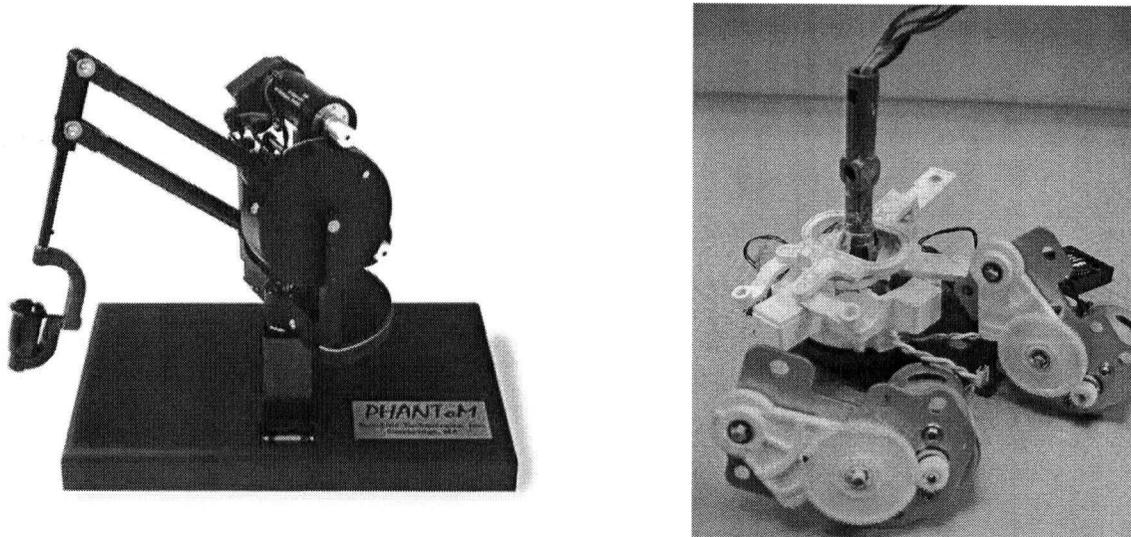
Figure 1-15: Cable drive and geared systems. The Pantograph (left) is a high-fidelity system with cable drive. On the right are plastic components found in the geared transmission of a force feedback joystick. The design has significant backlash and friction. The large amount of grease present could not compensate the latter.

Cable-drive systems can be found in many high-end multiple degree of freedom devices [e.g. the Phantom by Sensable Technologies and the Laparoscopic Surgical Workstation by Immersion Corp.]. Two advantages of cable-drives over direct-drive systems is that cable drive designs allow more freedom in the placement of the motors relative to the end-effector, and it allows more flexibility in choosing the transmission ratio. Compared to geared systems, cable-driven systems have virtually no backlash. Disadvantages of cable-drive systems are the high costs of manufacturing, assembly and servicing issues.

Finally, there are tension-based devices that use cables and pulleys instead of linkages (Figure 1-17) [Ishii, 1993; Kawamura, 1993; Kim, 2002b]. These systems combine low inertia with the absence of backlash but can be difficult to control because of the complex dynamics of the cables. A commercial version of this design is marketed by Mimic Technologies.
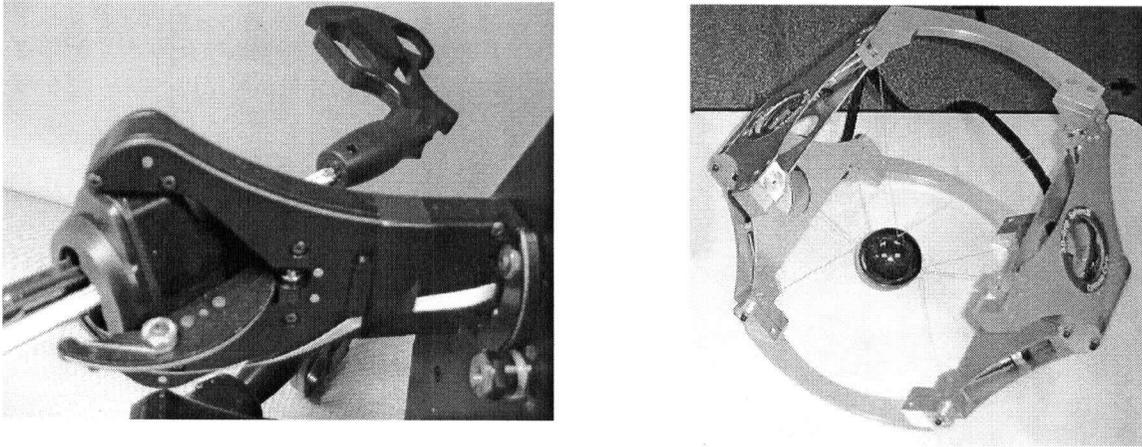
Figure 1-17: Two transmissions using cables: A cable-drive system in the LSW laparoscopic interface (left). Arrows indicate the location of some of the cables. The tension-based system on the right combines minimal backlash with low inertia.

1.3.1.3   Encoders

Most haptic interfaces use digital encoders to measure position. Their cost is highly correlated with the resolution. While haptic problems with low encoder resolution were discussed earlier in this chapter, low resolution can also lead to visible quantization of the tool movements. Costs of encoders vary from US$28 apiece, as found in one particular commercial design of a laparoscopic haptic interface, to roughly US$400 a piece for 6000 CPR encoders found in a high end pantograph design.

1.3.1.4   Electronics

The electronic components other than the encoders are the interface to the computer and the amplifier.

Interfacing electronics send the digital encoder signals to the computer and convert the digital control signals from the computer to analog signals for the amplifiers. Because of the high update rate, and many degrees of freedom, a very high bandwidth communication channel is needed. The Immersion LSW uses 3 custom-made PCI boards to control the hardware. In force feedback joystick design the communication bottleneck problem has been solved by downloading local models of the VE to a processor on the joystick to perform high update rates locally so that only low update rates between the interface and the computer are needed [Patent

17

US2002033841]. The cost advantage of this is that a standard communication port on the computer, like the serial port or USB-1, can be used. This has not been used in surgical simulation so far though Mimic Technologies is now shipping haptic interfaces with up to 6 DOF that connect to the computer through standard USB 2 ports. With new computer ports becoming faster, the communication bottleneck is diminishing. Surgical simulations could benefit from extra local processing power in the haptic interface to offload the main computer CPUs. While processing power of standard personal computers is constantly increasing, it may take a long time before the need for more processing power by surgical simulations is satisfied. Successful attempts to increase processing power in a cost-effective way has seen the use of powerful graphics card to perform haptic computation processing [James, 2002].

Amplifiers are more expensive with increase of power, dynamic range, and reduction in noise.

Since all the components described before are needed for each actuated degree of freedom, the total cost of components will go up roughly linear with the DOF. A common perception is that at the same time reliability will go down which can add to long-term costs.

1.3.2 Hardware Designs

Tendick and collegues at U.C. Berkeley developed a haptic interface that makes use of two Sensable Corp. Phantom haptic devices to provide force feedback at the tooltip in combination with custom made devices that provide torque-feedback on rotation of the shaft of the instruments [Tendick, 2000]. Another design based on modified Phantom's was developed by Ben-Ur at the MIT Artificial Intelligence Laboratory [Ben Ur, 1999]. A laparoscopic simulation station developed at the Forschungszentrum Karlsruhe that is now marketed by Select IT as the VEST System One makes use of modified Immersion Laparoscopic Engines[4]. The Laparoscopic Engines are the precursor of the Laparoscopic Workstation but had a much less sophisticated design. Its design has similarities with designs used for force-feedback joysticks [Moore, 2000] but with the addition of actuation of rotation of the instrument shaft. Baumann developed an interface based on a spherical remote-center-of-motion mechanical structure [Baumann, 2001] at

---

[4] The 'Karlsruhe Endoscopic Surgery Trainer' http://www-kismet.iai.fzk.de/TRAINER/mic_trainer1.html

the Swiss Federal Institute of Technology, Lausanne, Switzerland. Designs with hybrid spherical mechanisms were proposed by Payandeh *et al.* at S.F.U., Canada [Payandeh, 2003].

Currently the following companies produce haptic interfaces or haptic simulations systems for laparoscopic surgery: Immersion, Simbionix, Xitact, and Mentice.

The top model by Immersion Corp. is the Laparoscopic Surgical Workstation. Ten motors are located in the main housing (Figure 1-9). Cables run from each motor through the arm between the housing and the instrument to drive the main 4 DOF of the instrument. Only the motors to actuate the grippers are not placed in the main housing but in the handle instead [Patent EP1417547].

Xitact's patent on a haptic device is about providing force feedback to the shaft of the instrument for both the insertion and rotation. A method is described in which a drive block around the instrument's shaft engages with grooves made in the shaft [Patent WO03023736].

Simbionix partnered with Mimic Technologies in Seattle to apply Mimic Technologies' haptic interface technology to Simbionix's simulators. A patent search for "Simbionix" did not bring up any patents related to haptic interface designs of their own in the worldwide database. Mentice, the company that markets the MIST system, does not have patents related to haptic interface technology either.

1.3.3 Market Size and Relative Cost of Contributing Factors

Obtaining a generic formula for the breakdown of costs for haptic interfaces is practically impossible. A very important factor in the cost price of individual products is the production size. With high volumes, the cost of R&D and marketing can be spread over more products, buying power will bring component costs down, and cheaper manufacturing techniques can be applied as well as different designs more suitable for mass production. Given the fact that there are only a limited number of laparoscopic surgeons in the world, with only a limited number of institutions that are able to afford high-tech training environments for their surgeons, the

maximum number of haptic interfaces that can be sold for laparoscopy training may well be in the thousands.

For a 3 DOF haptic interface developed at a university, part costs contributed 30% to manufacturing costs, labour the other 70%. For a similar design, but extended to 5 DOF, the components accounted for 60% of the price. The cost of an amplifier was not included, nor the cost of development. Many commercially available haptic interfaces were originally developed at universities, for example the Phantom. It should be noted that haptic devices with high degrees of freedom are more prone to break down and therefore more expansive to maintain.

The relative cost advantage from big batch sizes may be reduced in the future by the advancement of completely automated flexible manufacturing plants that can make small batches at low cost. This technology, which was pioneered by Paul Wright at U.C. Berkeley [Wright, 1998] relies on customers using software supplied by the plant that automatically checks the customer's design for compatibility with the CAM plant. The approved design is submitted to the plant over the internet and produced with minimal intervention of humans. While the original technology was developed for fast prototyping machines, this service is now available commercially for milling, turning, CNC laser cutting, and CNC waterjet cutting of metals as well as injection molding, extrusion molding and thermofolding of plastics in low batch sizes (www.emachineshop.com), and fabrication and assembling of printed circuit boards (www.pad2pad.com).

The main cost-contributing factors in a haptic interface are R&D, components and manufacturing. If manufacturing costs for prototyping and manufacturing custom-parts in low batch sizes (50-5000) will substantially come down in the future, the importance of the costs of components will become relatively more important.

1.4 The influence of Haptic Quality on Task Performance

Studies have shown that for multiple tasks, the addition of force-feedback significantly influenced task performance[Adams, 2001b; Moody, 2002; Wagner, 2002] [Adams, 2001a; Moody, 2002; Wagner, 2002]. How accurate does the force-feedback have to be to be effective?

The limited number of studies that have examined task performance as a function of haptic quality suggest that task performance is often not affected by differences in haptic quality, even if an obvious degradation is perceived [MacLean, 1996; O'Malley, 2002; Richard, 2002]. This suggests that the force rendering in current surgical simulators may be of higher quality than necessary for their intended purposes.

1.5 The Big Picture

This research is part of a bigger project to assess the validity of training environments for laparoscopic surgery. Kinnaird assessed performance validity of the VR training environment used in this study and a bench top trainer by quantitatively assessing whether measurable quantities of performance are the same in the simulators as those in the OR [Kinnaird, 2004]. She found that the surgeons treated the VR environment differently than the OR. The impact of this on the results presented here is discussed in paragraph 3.5.4 and 4.2.1.2: "Upgrading the Gold Standard Hardware and Software". In a parallel study, Lim is investigating the influence of training in the VR environment on OR performance. This study addresses the issue of validity of the simulator by asking the following important question: "Does training on the simulator result in improved OR performance?". Publication of results of that study is expected in fall 2004. Together with the research presented here, these three projects aim at obtaining a simulation for training of laparoscopic technical skills that is both a valid alternative to training in the OR and cost-effective.

1.6 Thesis Objectives

Given the high costs of high-end haptic hardware, the possible advantages force-feedback simulators can bring for teaching hospitals, the barrier the high costs pose for teaching institutions to acquire these devices, and the plausibility that lowering the quality of the hardware may not bring a significant change in user performance, we would like to answer the following question:

" Can we degrade haptic quality in a state-of-the-art simulator without incurring a noticeable difference in user performance?"

21

Secondly, if we find that certain factors in haptic quality can be degraded without affecting task performance, can we rank them on impact? E.g. if both friction and backlash lead to a significant chance in task performance, which of the two affects task performance most? By answering these questions we hope to define new guidelines that will lead to more cost-effective designs of haptic hardware for surgical simulation. To do this, we developed a simulation with variable haptic quality (*Chapter 2*) and studied the impact of varying the hardware quality on task execution with user studies described in Chapter 3. Chapter 4 will contain the major findings, conclusions, and future work.

Chapter 2 Simulating Inexpensive Haptic Hardware

This chapter describes a platform devised to explore the dependence of task execution in a virtual environment on the quality, and therefore cost, of the system's haptic hardware: a complex haptic interface in which hardware quality can be varied in simulation. Software intercepts the position and force signals between the haptic hardware and the virtual environment software, and alters them to create the effect of increased friction, cogging, backlash, inertia and/or lower force output. All parameters of the introduced effects can be set independently or in combination and on a continuous scale. A primary contribution of this project is the creation of haptically realistic hardware models that are stable in combination on complex haptic hardware.

Section 1 will give an overview of our implementation of the different hardware quality effects and state our objectives. Section 2 will discuss the models that we use for each effect, introduce relevant previous work, and provide details on integration of the different models. Finally Section 3 provides experimentally obtained data from our models and results from tests on computational efficiency. A conclusion and future work section can be found in Chapter 3. The majority of the work in this chapter has been presented at ICRA 2004 [Brouwer, 2004].

2.1 Motivation and Approach

There is a general consensus that a significant part of surgical training for minimally invasive surgery will move towards computer based simulation. While the software for these simulators has significant room for improvement, the current commercially available hardware is able to render soft tissue simulations very well. A main drawback of the hardware is a cost too high for many hospitals to afford. Therefore, there is an incentive to make this hardware significantly less expensive. One approach is to invent lower-cost components that don't sacrifice quality; e.g. through optimizing electro-mechanical design (e.g. [Vlachos, 2004]), and signal processing to compensate for unwanted characteristics as friction [Kwon, 2000] or torque ripple [Lawrence, 2004]. Another approach, taken here, is to understand the dependence of task execution on the device's performance in order to establish design guidelines for the minimum haptic interface quality required for a given application. That is, we would like to see if, and under what conditions, low performance components will yield equivalent task performance as their high

performance counterparts, thus permitting systematic substitution of less expensive components where appropriate.

2.1.1 Goal of Simulations

To study the dependence of task execution on haptic hardware quality, we require an experimental test bed that:

- Allows us to vary the quality factors associated with major cost-contributing components of haptic hardware.
- Supports this variation in quality levels across a realistic range of possible designs.
- Allows quality factors to be tested singly and in combination.

Such a setup will allow us to rank hardware quality characteristics based on their impact on task performance: e.g. given the financial constraints, should the designer put more effort in minimizing friction or cogging?

2.1.2 Approach to Designing Simulations

Our test bed is based on a surgical virtual environment with force feedback in which we can vary perceived haptic hardware quality through simulation, by means of a custom software plug-in that intercepts the control loop between the virtual environment and a state-of-the-art haptic interface. Through modification of the position signal sent from the hardware to the virtual environment and the force signal sent the other way, this high-fidelity interface can be made to display characteristics associated with inexpensive hardware, such as friction, superimposed on the VE. As a result, the user will perceive forces from two simulations: the forces generated by the surgical simulation software, as well as the forces generated by the hardware-degradation simulation. Forces generated by the surgical simulation software are altered by the hardware-degradation simulation before they reach the user. The forces generated by the hardware-degradation simulation are dependent on instrument kinematics as well as the forces generated by the surgical simulation software.
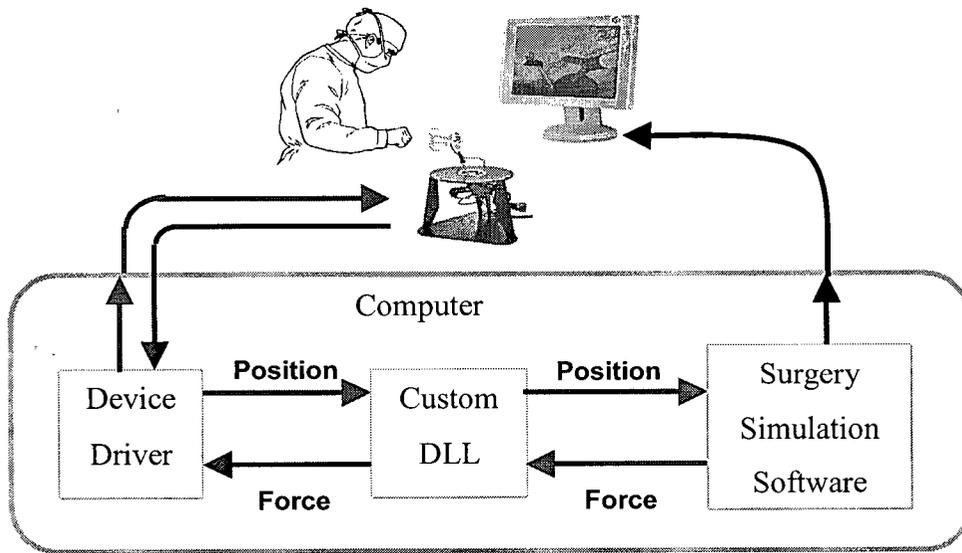
Figure 18: Interception of the signal flow between the surgical simulation software and haptic control algorithms allows the perceived haptic quality to change according to algorithms that model desired behaviour.

We chose the approach of simulated degradations over a hardware implementation because the latter would entail building haptic interfaces with a wide range of components and designs. While the use of real hardware instead of simulated hardware would give the highest possible fidelity, it would not allow us to change parameters easily or individually, while construction of the hardware would be both time consuming and expensive. To illustrate this, if the effect of higher cogging in the current hardware were to be studied, the motors would have to be replaced with motors with higher cogging. Since the variety of non-custom motors is limited, the new motors would almost certainly have different inertia as well as different dimensions from the one that are replaced. The former would lead to confounding of parameters in the user test while the latter could potentially lead to a very expensive redesign of the haptic interface.

Though the fidelity of a software implementation is limited by the accuracy of the models and the quality of the haptic interface used to display the models, the models of hardware degradation only have to be sufficiently realistic to result in comparable task performance. We maximized the quality of the hardware simulation by using one of the best laparoscopic haptic interfaces available at the time of experimentation. Overall, the advantages of software simulation outweighed the possible disadvantage.

2.1.3 Hardware and Virtual Environment

We use Immersion Corp.'s Laparoscopic Surgical Workstation[5] (LSW) (Figure 1-9; hardware specifications in Table 2.1). This device is currently the only commercially available high-end haptic interface for laparoscopy. This device has two 5-DOF laparoscopic instruments, each of which can move in and out of a 2-DOF pivoting point and rotate around a longitudinal axis. A virtual tool tip opens, closes and rotates relative to the main shaft. The first two of these three degrees of freedom are actuated. The main 4 DOF are actuated by low inertia, cogging-free permanent magnet motors while position is read by digital encoders (resolution ranging from 1024 CPR to 2048 CPR). Results of a preliminary study on some mechanical characteristics of the device is performed by Waldron and Tollon [Waldron, 2003].

Our software is designed as a stand-alone library compiled into a dynamic link library (DLL). A DLL is a file containing executable code and data that are dynamically linked with the application that uses them during program execution rather than being compiled with the main application. It can be shared by multiple applications. Our DLL intercepts the force and position signals between the hardware driver and VE software. It can therefore relatively easy be modified to work with different kinds of VE software or hardware.

TABLE 2.1. LAPAROSCOPIC WORKSTATION SPECIFICATIONS

	Range	Cont. Output	Peak. Output	Sensor Res.
Insertion	170 mm	11.0 N	19.0 N	.008 mm
Pitch	100°	0.47 Nm	0.85 Nm	0.01°
Yaw	100°	0.47 Nm	0.85 Nm	0.01°
Handle Twist	180°	0.04 Nm	0.07 Nm	0.03°
Virt. Tip Twist	Cont.	N/A	N/A	0.7°
Handle Grip	20°	0.15 Nm	0.32 Nm	0.04°

The virtual environment (VE) is a simulation of minimally invasive surgical procedures, intended to offer training in laparoscopic techniques. It incorporates abstract tasks such as pick-and-place tasks as well as more complex modules requiring the subject to dissect, staple and cut

---

[5] www.immersion.com/medical

26

structures. Force feedback is provided in Cartesian coordinates to the tool-tip on two instruments. The software is commercially available as the Reachin Laparoscopic Trainer by Reachin Corp[6]. More details about the available VE modules are given in Chapter 3.

2.1.4 Challenges

The main challenges in implementing our hardware degradation simulations are caused by the complex dynamics of the haptic hardware in combination with a low and varying update rate. Models that are stable in a computer simulation may not be when displayed haptically, because the virtual models interact with the real hardware dynamics, including the hardware's friction, inertia and coupled kinematics. For example, in our hardware, rotating the unactuated tool handle altered the dynamics enough to cause instability in the pitch and yaw direction in some cases. An accurate description of hardware dynamics, when available, can in principle be incorporated into a computer simulation [Hacksel, 1994; Grewal, 2001]; however, this is usually unavailable for commercial hardware, and it is difficult to obtain this information experimentally. Even if the model was available, this hardware is so complex it would still be hard to utilize since the algorithms require very accurate models.

Likewise, our need to simultaneously simulate a variety of hardware degradations vastly complicates the stability prognosis. These virtual models interact with one another, in addition to the real hardware. This has both influenced details of the model implementations and imposed limits on their parameterization.

Finally, Reachin's simulation software uses a sample rate that varies between 200-2000 Hz when working in conjunction with our haptic degradation simulation on a 2.0 GHz dual Xeon computer with 2GB RAM. We have no control over the sample rate, except for making the hardware degradation simulation as fast as possible (see paragraph 2.3.3 Computational Load). We can not impose a fixed sample rate but we can measure update times. Therefore our models must work with all sample rates in this range.

2.2 Methods: the Hardware Models

We have chosen to model several primary effects found in less expensive haptic interface hardware: cogging, inertia, backlash, friction and force saturation. Together with encoder

---

[6] www.reachin.se

27

resolution and refresh rate, these are the most prominent quality descriptors for haptic hardware. Figure 2-1 shows which hardware components are associated with each quality aspect.
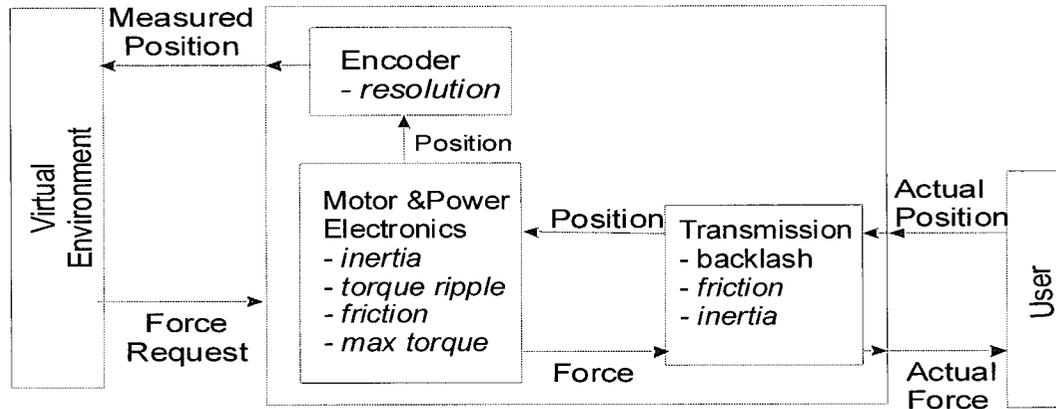
Figure 2-1: Flow diagram of forces and positions in a virtual reality system with haptics, and the factors that limit haptic fidelity.

We did not degrade refresh rate in our experiments because it depends on computing power rather than the haptic hardware, and the cost of computing power is rapidly declining. Nor did we degrade encoder resolution because the encoders used were not expensive and a reduction in encoder resolution would result in insignificant cost-benefits (in fact, our simulations of haptic-hardware degradation would have benefited from better encoders).

In the following paragraphs, each effect is initially presented as a 1-DOF translational model. We do not describe the rotational variants, also implemented, which are obtainable through a straightforward transformation. Each section begins with a short discussion of previous work relating to simulation of that physical phenomenon. The model parameters we used are listed in Table 2.2.

2.2.1 Inertia

2.2.1.1 Background

Inertia is the force or torque that resists changes in velocity of an object. An object with high inertia will require more force to speed up or slow down than an object with low inertia. The most straightforward way to simulate inertia is to multiply actual acceleration by the virtual inertia. However, an acceleration estimate obtained by double-differentiating the position signal is too noisy to produce a stable simulation. A common solution is to simulate the virtual inertia's

dynamics through integration of a 2nd order system, and virtually couple it to the probe position through a damper-spring combination (e.g. [Colgate, 1995]). To calculate damping, only a velocity estimate is required. The stiffness of the spring and damper coefficient determine the tightness of the coupling, which ideally is critically damped.

2.2.1.2   *Implementation*

Our system's temporal and position resolutions are such that the velocity signal tends to oscillate between a small number of values. Therefore we applied a first- order, low- pass Butterworth filter to the raw velocity before feeding it to the inertia model. We implemented this filter in the standard difference equation:

$$a(1)*v_f(n) = b(1)*v_{raw}(n) + b(2)*v_{raw}(n-1) + ... + b(nb+1)*v(n-nb)$$
$$-a(2)*v_f(n-1) - .... - a(na+1)*v_f(n-na)$$

(2.1)

In this equation v_f is the filtered velocity, v_{raw} the raw velocity, a and b vectors with the filter coefficients and na and nb the length of vectors a and b. A Butterworth filter was chosen because infinite-duration impulse response (IIR) filters require in general a much lower filter order then finite-duration impulse response (FIR) filters for a given performance level and accordingly introduce less delays, which is very important in this application. Also, the Butterworth filter has a maximally flat magnitude response in the passband, giving the smoothest response achievable. All standard implementations of these filters require constant sample rates over all the last na and nb updates, and therefore we could only use a 1st order filter as illustrated in formula (2.2). Filter coefficients are pre-calculated in Matlab™. The sample rate windows for which the same coefficients are used are 10Hz wide at low update rates (150Hz) and up to 200Hz wide at higher update rates. The lookup algorithm is designed such that the low-update-rate coefficients are retrieved first, so the least amount of time is spent when time is shortest in supply. The cut-off frequency was chosen at 70Hz to balance filtering high-frequency spikes due to time and positional discretization while passing user-induced rapid movements.

$$a(1)*v_f(n) = b(1)*v_{raw}(n) + b(2)*v_{raw}(n-1) - a(2)*v_f(n-1)$$

(2.2)

Because of the low starting quality of the velocity signal, an inertia signal based on its differentiation is still too noisy to be used. Therefore the inertia signal is low-pass-filtered again by averaging it over the last three updates or a 20ms time window, whichever is shorter. The resulting model is illustrated in Figure 2-2. Parameters are listed in Table 2.2.
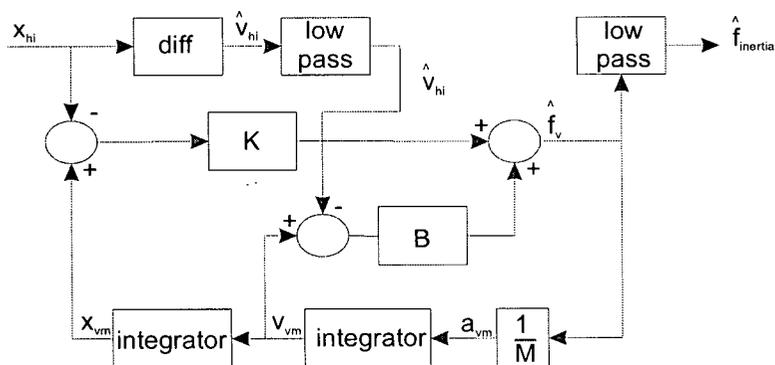
Figure 2-2: Virtual coupling with two low- pass filters to simulate inertia. Subscript 'vm' indicates that the variable is related to the virtual mass, subsript 'hi' indicates it is related to the haptic interface.

We hoped to be able to increase the stiffness of the virtual coupling through a stability analysis of the model [Colgate, 1994; Colgate, 1995], thereby making the K and B variables dependent on sample rate. High update rates will typically occur during free space motion, when there is no direct contact between the instruments and the tissue. Since we also expect the largest accelerations in user movements to occur at this time, it would make sense to stiffen the virtual coupling in this regime. However, the mathematical model we derived by applying Colgate's theory [Colgate, 1994] did not provide more insight on the relationship between the stability limits of B and K as a function of the sample rate. This is a direct result of the complexity of our transfer function compared to the transfer function used by Colgate et $al.$ ($H(z) = K + B\dfrac{z-1}{Tz}$).

Therefore we have continued to use the same virtual coupling coefficients regardless of sample rate. In a paper published after completion of our setup, Mahvash et $al.$ proposed a method that extends Colgate's results to the general class of nonlinear virtual environments and non-linear devices [Mahvash, 2004].

2.2.2 Backlash

2.2.2.1 Background

Backlash refers to the play between mechanical components that results in discontinuous motion transfer. Often backlash is present in gears due to play between the teeth of the gears. In haptic interface design, backlash can be present in the transmission between the motor and the part of the interface that is in direct contact with the user. Because the encoder is often mounted directly on the motor (Figure 2-1), backlash can also affect position measurement. The amount of backlash is dependent mainly on the design of the transmission. As discussed in Chapter 1, direct drive and cable driven systems have very low backlash or none at all, while geared systems can have considerable amounts of backlash.

In a system with backlash, motion transfer between two masses occurs within a finite gap, causing a discontinuity and impact upon changes in direction and velocity. A significant amount of simulation research has been done in the control theory community to accurately control systems with backlash. In a commonly used model, impact between the two masses in a transmission with backlash is approximated as occurring through a linear damped spring [Gerdes, 1995; Jukic, 2001] as depicted in Figure 2-3.
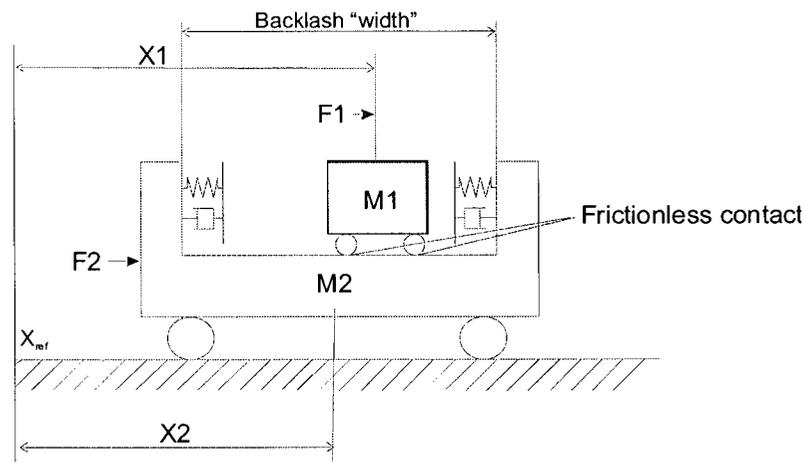
Figure 2-3: A backlash model commonly used in the field of control theory. X1 indicates the position of mass 1 (M1). X2 indicates the position of mass 2 (M2). A spring and damper combination on each side of the gap simulate impact of the two masses.

The spring forces F are a function of the indentation δ, and the damper forces G a function of the speed of indentation $\dot{\delta}$. The following equations describe the system when the masses are in contact:

$$F_1 - (F + G) = M_1 \ddot{x}_1 \tag{2.3}$$

$$F_2 + (F + G) = M_2 \ddot{x}_2 \tag{2.4}$$

When the masses are not in contact, these equations simplify to:

$$F_1 = M_1 \ddot{x}_1 \tag{2.5}$$

$$F_2 = M_2 \ddot{x}_2 \tag{2.6}$$

Implementation

We have adopted this model by attaching the virtual coupling to the gap-wall, engaging it when the user interface contacts either edge of the gap. In Figure 2-4, the virtual mass (M_{vm}) represents the simulated extra mass of the motor and transmission. The position of the virtual mass is x_{vm}. The user's input position is indicated by x_{hi}. We assume that there is negligible backlash in our real hardware's cable drive transmission and therefore consider the encoder signal an accurate estimate of x_{hi}.

(a): In the gap   (b): Outside the gap
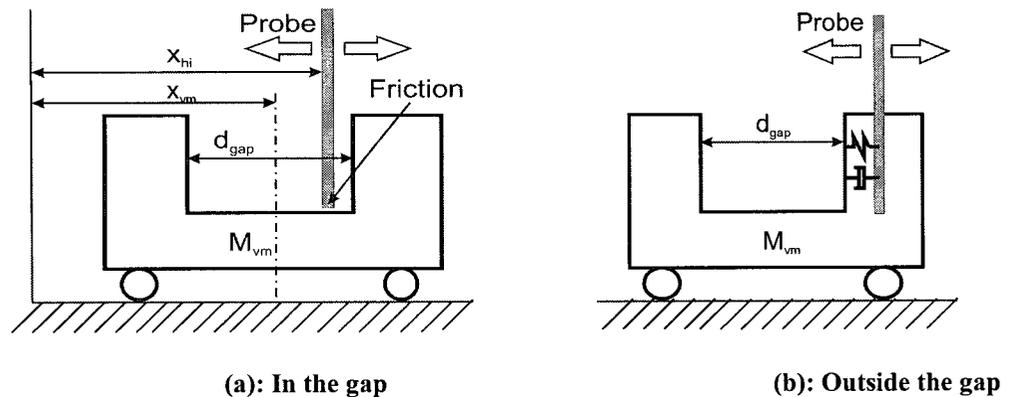
Figure 2-4: Backlash model

To enhance stability, we apply a small amount of friction on the mass M_{mv} while the masses are not interacting to prevent them from oscillating: in any real system some friction would always be present. This friction model is described in Section 2.2.3. When the probe is in contact with the mass, the virtual coupling engages the gap wall (b), and a variable p_{vc} holds the attachment point of the virtual coupling. p_{vc} is undefined when the probe is not in contact with the mass.

$$if\ x_{hi} > x_{vm} + 0.5 \cdot d_{gap}: \quad p_{vc} = x_{vm} + 0.5 \cdot d_{gap}$$
$$if\ x_{hi} < x_{vm} - 0.5 \cdot d_{gap}: \quad p_{vc} = x_{vm} - 0.5 \cdot d_{gap}$$

(2.7)

The force felt by the user is then described as:

$$if\ (x_{vm} - 0.5 \cdot d_{gap} < x_{hi} < x_{vm} + 0.5 \cdot d_{gap}):$$
$$F = 0$$
$$otherwise:\ F = F_{ext} + K(x_{vm} - x_{hi}) + B_2(\dot{x}_{vm} - \dot{x}_{hi})$$

(2.8)

2.2.3 Friction

2.2.3.1 *Background*

Friction refers to the force that resists the relative motion of two bodies in contact, or the tendency to such a motion. Since friction plays an important role in control of mechanical systems, a substantial amount of work exists on the modelling of friction. Relatively little work has been done on the haptic display of friction and how this modelled friction compares to real friction in the way it affects task performance.

On a microscopic scale all surfaces contain asperities, even if they look smooth (Figure 2-5). When two of these surfaces are in contact and one surface tries to move relative to the other, these asperities will lock or rub against each other, causing the friction force.
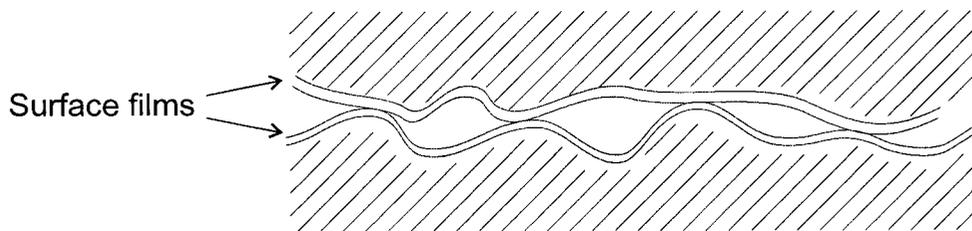
Figure 2-5: Microscopic contact between two surfaces. The actual surface of contact is small compared to the total surface.

The relative motion of two surfaces in dry friction can be categorized as follows:

33

- Presliding: During static friction, the asperities can deform in both normal and tangential directions. Experimental observations of ball bearings in small motions have shown an approximate linear relation between applied force and displacement during presliding. Experiments by multiple authors, summarized in [Armstrong-Helouvry, 1994] have shown that the breakaway displacement in engineering materials is in the order of 2-5 microns in steel junctions.

- Breakaway and sliding: Friction is proportional to the shear strength of the asperity junctions. The area of these junctions is in direct proportion to the total load. In surface contact without any contamination, the shear strength of the contact area can be as great as the shear strength of the materials themselves. In practice, surfaces are never really clean, and usually lubricants are added to lower the shear strength. Lubricants contain additives that bond to the metal surface and cause the shear strength in the contact area to lower.

Many friction models are described in the literature; Armstrong-Helouvry et al. provides a good overview [Armstronghelouvry, 1994]. The most important models will be reviewed in the next section.

Haessig *et al.* proposed a friction model that simulates the rubbing of asperities of two surfaces [Haessig, 1991]. The model introduces deformable bristles, sticking out from the surface, as an equivalent of the asperities. The bristles can attach to bristles on the other surface, deform under motion and snap when the force becomes too high. The friction force is dependent on the resultant deformational forces of all bristles (each bristle is basically a spring). The bristle model was designed for accuracy, without regard to use of computer time, to accurately simulate multi-point microscopic stick-slip contacts in real surfaces, and it is too computationally expensive for real time processing. Therefore Chen *et al.* developed a version for haptic rendering based on a single bristle that produces the dependency between normal and friction force [Chen, 1997]. The authors report mixed results, and we could not implement it because our interaction normal force is unavailable.

In Dahl(1976), the author proposes a friction model based on a single differential equation. In this frequently cited model, the time rate of change of solid friction is expressed as:

$$\frac{dF(x)}{dt} = \frac{dF(x)}{dx}\frac{dx}{dt} \qquad (2.9)$$

in which F(x) is a friction force that is only a function of displacement x. In 2000, Hayward & Armstrong showed that this model drifts if there are small oscillations in position while a small bias force is present [Hayward, 2000]. This is important in haptic rendering, since position is read from digital encoders that tend to oscillate due to their discrete nature and the natural tremor of the hand.

Hayward & Armstrong [Hayward, 2000] produced a 4-state version dependent only on position. However, the state transition process assumes a constant sampling rate, making it unusable for our system.

Karnopp introduced a friction model that incorporates stick-slip without pre-sliding: i.e. when the friction force is below F_{static}, the relative velocity between surfaces is set to zero, something that can not be done in haptic simulation if the velocity of one of the objects is directly controlled by the user [Karnopp, 1985]. In this model the friction force F_f is a function of velocity v, except for the speeds where the absolute velocity is smaller then a value v_{min}. In this velocity range, friction force is the resultant of all forces working on the object. If this force exceeds a set limit F_{stick}, the object will start to move and the friction force will be dependent on velocity again. Richard proposed a modified version of Karnopp's model that included viscous friction and asymmetric friction values for positive and negative velocities [Richard, 1999].

Nahvi & Hollerbach introduced a haptic friction model that has two states: stick and slip. In the stick-state, a virtual spring extends from the current position of the mass to the rest position of the mass to simulate presliding. This spring ruptures when the spring force exceeds F_{static}. The transition from slip-stick is made continuous by choosing the attachment position of the spring such that the static friction force is equal to the slip friction force [Nahvi, 1998].

2.2.3.2 Implementation

Multiple authors [Chen, 1997; Gosline, 2001] reported oscillation problems with friction forces based on velocity measurements in systems where velocity is not directly measured but obtained from discrete position measurements. With the haptic hardware used in this study, the update rates achievable, and the low velocity movements in this task, only 0 to 3 encoder tics pass in one update. Even after filtering this would result in inaccurate and possibly oscillating friction values.

Multiple authors identified friction to be non-symmetric (with regards to velocity) in friction identification experiments [[Armstrong-Helouvry, 1991; Johnson, 1992]. To minimize the number of variables in these experiments, we assume the friction to be symmetric.

For our main haptic friction effect, we use the model developed by Nahvi *et al.* but we assume the unknown normal force to be of unit size. In the slip phase the friction force F_f is then defined as:

$$F_f = -\mu_d \frac{v}{|v|} \qquad if \ |v| > v_{\min} \tag{2.10}$$

in which μ_d is the dynamic friction coefficient, v is the tangential velocity of the two moving surfaces in the haptic interface for which extra friction is modeled, and v_{min} is the minimum threshold velocity to remain in the slip state. When the relative velocity of the two surfaces drops below this minimum, the spring attaches so as to form a continuous transition from slip to stick state:

$$F_f = -\mu_d \frac{v}{|v|} = -K_f(x - x_c) \tag{2.11}$$

in which x_c is the attachment point of the spring to the second surface: the stick centre. The stick centre can therefore be found by:

$$x_c = x - \frac{\mu_d}{K_f} \frac{v}{|v|} \tag{2.12}$$

It reaches a maximum static friction (stuck state) at a pre-sliding displacement of 100 μm in translation and (0.0075 rad in rotation). Parameters are shown in Table 2.2.

In our backlash model, friction was necessary to dissipate kinematic energy from the virtual mass to prevent it from sliding back and forth between the walls. Since the velocity of the virtual mass in the gap is not controlled by the instrument position, we were able to directly apply Karnopp's friction model here. The advantage of Karnopp's model over the model proposed by Navhi *et al.* for objects not directly controlled by the user is its increased stability in the stick state: velocity can be set to exactly zero. In the stick state, this model can be described as:

$$ if \quad |F_f| < F_{\max} \qquad F_f = \sum F_{ext}, \quad v = 0 \qquad (2.13) $$

When the sum of all external forces exceed the maximum F_{max}, the contact will go into the slip state in which the state variables are defined similar to the model proposed by Navhi *et al.*:

$$ if \, |v| > v_{\min} \qquad F_f = -\mu_d \frac{v}{|v|} \qquad (2.14) $$

2.2.4 Torque Ripple and Cogging

DC brushed permanent magnet motors are the most commonly used actuators for kinaesthetic haptic force feedback interfaces. Ideally, their output torque would be independent of the position of the rotor. However, in most motors torque fluctuations occur as the motor rotates due to cogging and torque ripple.

Cogging

Caused by the preferential alignment of rotor and stator, cogging can be felt as a series of opposing and aiding torques as the motor is turned when unpowered (Figure 2-6).
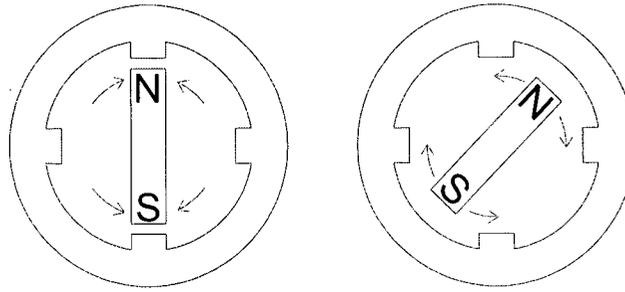
Figure 2-6: Cogging: A stable (left) and unstable (right) detent position of a permanent magnet motor.

Figure 2-6 (left) shows the rotor magnet aligned with the poles. This is called a stable detent position: any disturbance of the rotor position will result in a torque that causes it to realign with the poles again. The right side of Figure 2-6 shows the same magnet in another detent position: here also no resulting torque on the rotor is present but in this case any disturbance of the rotor position will result in a torque away from the shown position. A position like this is called an unstable detent.

In formula form, the cogging torque $T_{cogging}$ can be described as:

$$T_{cogging} = \sum_{1}^{P} \left(-\frac{1}{2} \phi_g^2 \frac{d\Re}{d\theta} \right)$$

(2.15)

with P the number of poles, ϕ_g the air gap magnetic flux, $\Re$ the reluctance, and θ the rotor position [Studer, 1997]. A motor design with more poles will usually result in a higher torque for the same current level. As the number of poles increases, the space for each pole will become smaller, while the total amount of between-pole space increases. This eventually leads to decreased performance, limiting the number of poles [Hanselman, 1994]. As a result, the cogging torque of a permanent magnet motor is a complex function of commutation, motor geometry, and material properties. Example cogging torques as a function of rotor angle are shown in Figure 2-7. Many motor design techniques have been applied to reduce cogging [Hanselman, 1994; Studer, 1997] successfully, such as skewing of the magnets or using ironless cores. A drawback of low and zero cogging designs is their high price.
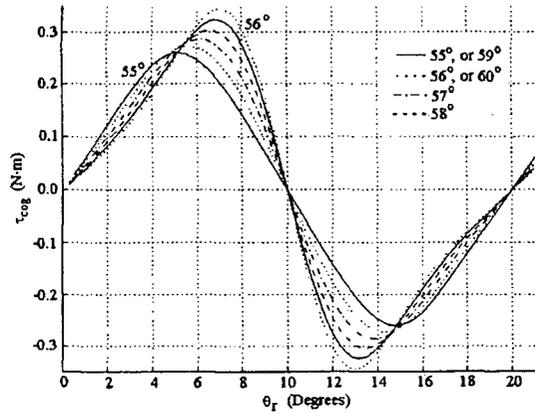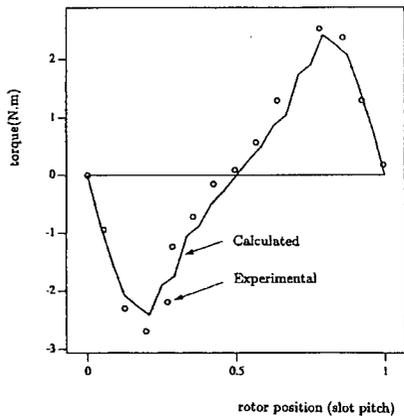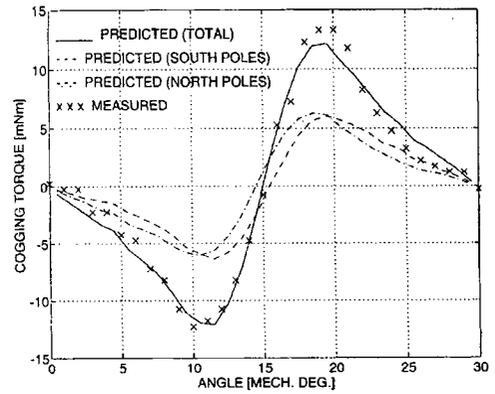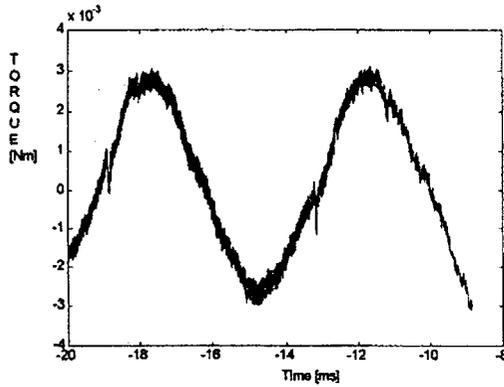
Figure 2-7: Examples of simulated and experimentally obtained data from various IEEE publications (source (left to right, top to bottom): [Benarous, 1999], [Deodhar, 1996], [Ishikawa, 1993], [Studer, 1997]).

2.2.4.1 Ripple torque

The main cause of ripple torque is fluctuations in back electromotive force (EMF) from their ideal value. Most permanent magnet DC motors are designed to have a trapezoidal back EMF [Kapuscinski, 1997]. In practice, deviations from perfect trapezoidal shape occur. Figure 2-8 illustrates the resulting rounded edges and nonuniform slopes.
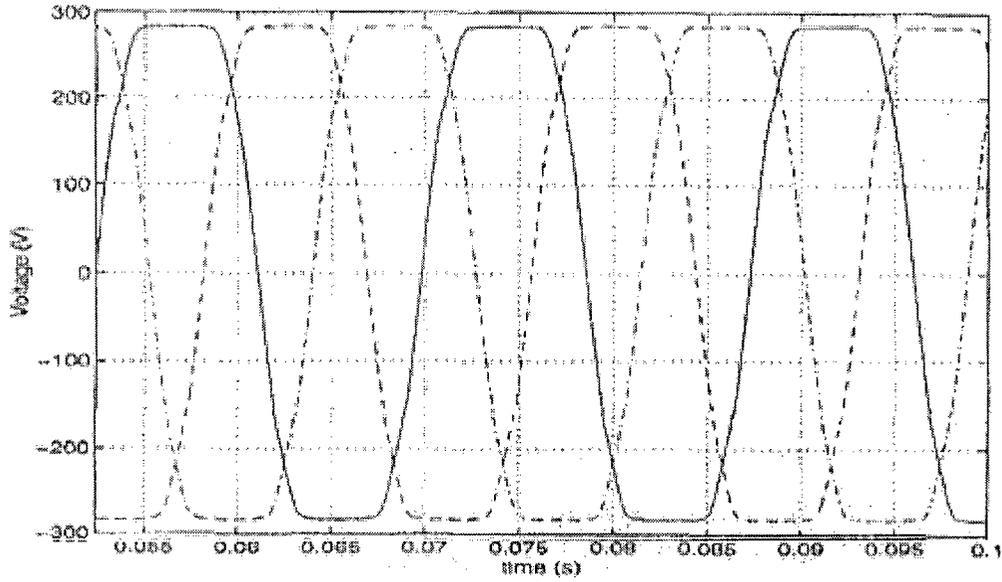
Figure 2-8: Experimentally obtained back EMF in a motor with trapezoidal back EMF (source: [Sozer, 1998]).

The effect of this on the torque output is seen in Figure 2-9 which show experimental results from work by Lajoie-Mazenc *et al.* [Lajoie-Mazenc, 1989] on ripple torque in a motor with trapezoidal back EMF. The left diagram shows the output torque as a function of rotor position for a motor with 3 phases. The right figure shows the results for an identical motor except for that it has 9 phases. According to the authors, cogging is negligible in both motors.
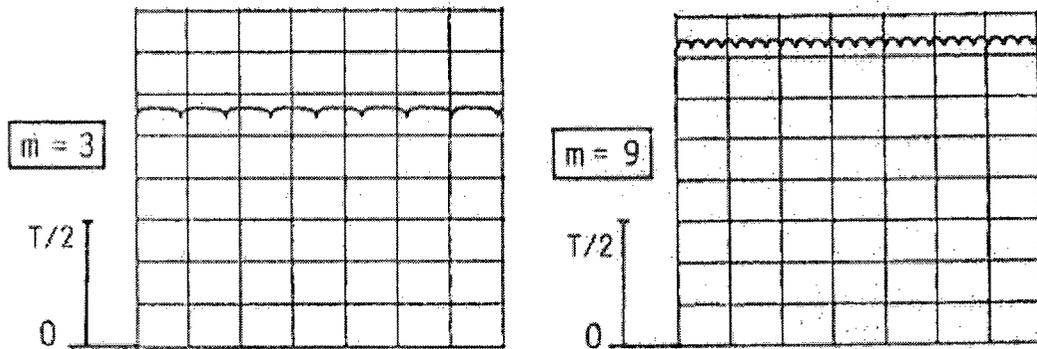
Figure 2-9: Torque ripple as a function of motor angle (source: [Lajoie-Mazenc, 1989])

Torque output is higher for the design with 9 phases. The angular frequency of torque ripple is higher in the design with 9 phases, but the torque ripple amplitude is the same in the two designs.

40

Three-phase motors are most commonly used since two- and three-phase motors minimize the number of power electronic devices required to control the winding currents.

Implementation

We produced a torque-angle shape match to experimentally obtained cogging data [Ishikawa, 1993; Deodhar, 1996; Studer, 1997; Benarous, 1999] which resulted in a sinusoidal relationship between torque and motor angle. Since both cogging and torque ripple produce a position-dependent torque fluctuation, we chose to capture both effects in a single model to reduce experimental variables. While in torque ripple the exact function shapes can vary as a function of torque, we felt this was less critical since torques are generally low in our simulation. Furthermore, much of the fine instrument positioning occurs while there is minimal tissue interaction. To minimize the number of experimental parameters, we only modeled cogging.

2.2.5 Torque Saturation

Electromotors are usually described by both continuous and peak maximum torque outputs; the peak torque can only be exerted for a limited time because of heat generated. Thus, while a motor has two design torque limits, the lower limit will be expressed in hardware as overheating and eventual damage to the motor rather than a haptically perceptible performance reduction. Therefore we applied only a single cut-off limit for motor torque: i.e., when in effect, the motor force is clipped to the imposed saturation level.
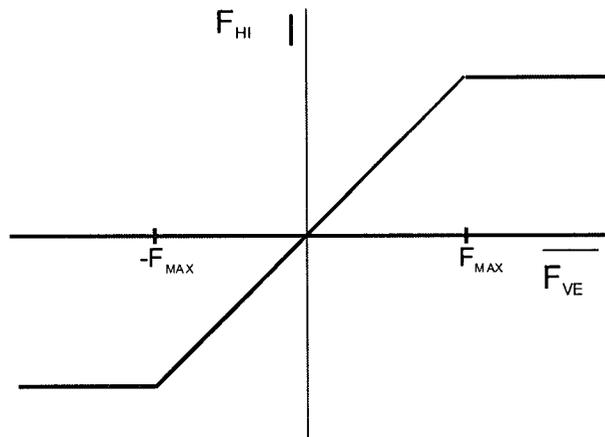
Figure 2-10: The relation between forces displayed at the haptic interface and forces commands send by the virtual environment when force saturation is turned on. F_{VE} is the force output by the virtual environment. F_{HI} is the force displayed on the haptic interface. F_{max} can be varied to saturate the force at the desired level.

41

2.2.6 Model Integration

We integrated our models in two stages. First, we combined the various degradations into a single DOF model so as to maximize simulation fidelity and stability. Next, we extended this 1-DOF model to the 3-DOF movement of the instrument tool-tip.

2.2.6.1   1-DOF integration

To the extent possible, we based our integration on the actual physical location of the respective degradations in a typical haptic hardware system (see Figure 2-1). In some respects, detailed and justified below, this was not practicable.

We first simplified our model of the real system (Figure 2-1) by lumping the mass of the motor and transmission. Backlash is then defined as the difference between the user's probe and the lumped mass. Friction is defined as the movement-opposing force between this mass and the 'ground'. Forces from the virtual environment are transferred through this backlash mechanism. As a result, the user feels forces from the virtual environment and from the degradation models only while the probe is in contact with the mass Figure 2-11. The mass of the motor and the parts of the transmission were all lumped together to avoid the difficult, if not impossible, process of realistically simulating the dynamics of impact of multiple small masses at very small distances from each other.
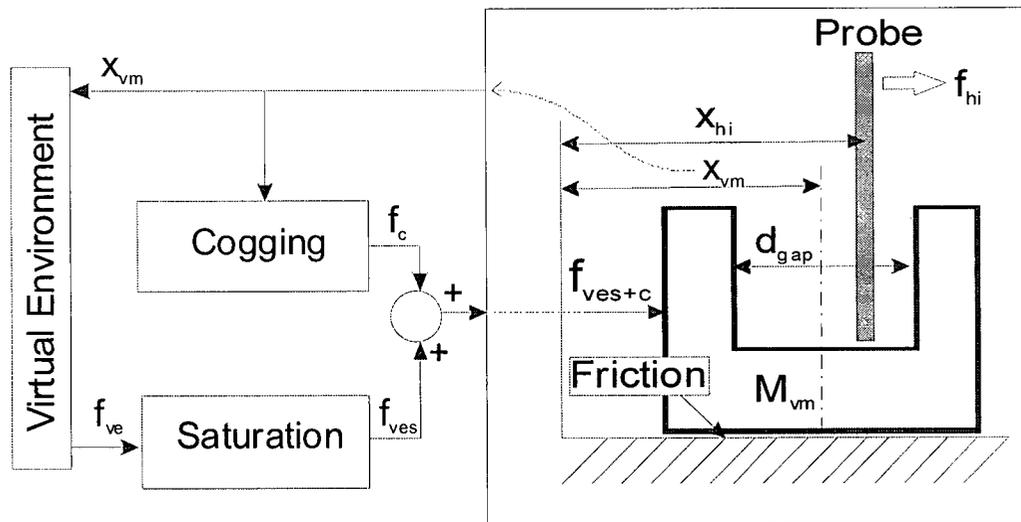
Figure 2-11: The basic physical representation of our model integration with backlash present. The following subscripts are used: hi=haptic interface, vm=virtual mass, c=cogging, ve=virtual environment, ves=saturated VE force.
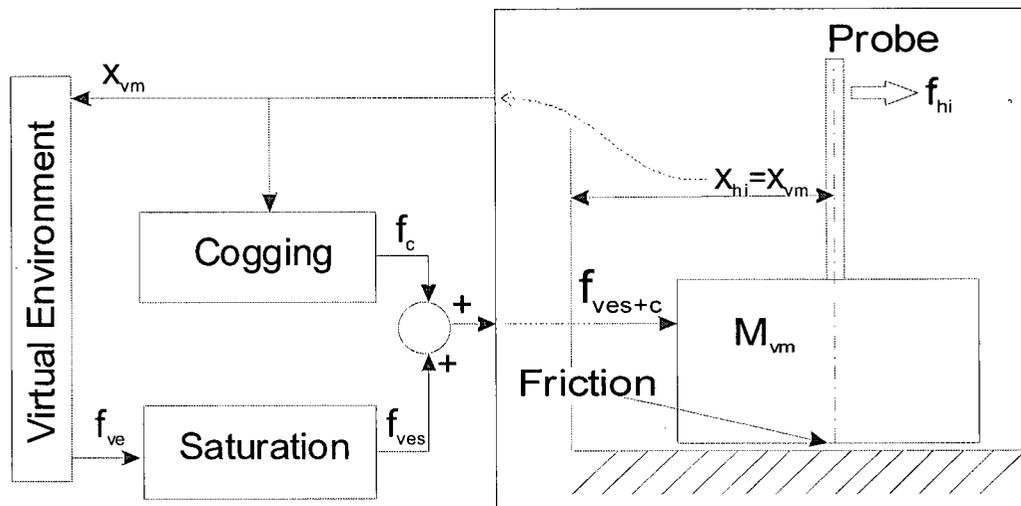
Figure 2-12: The physical representation of the model integration if no backlash is present

To maximize perceptual fidelity of the different models, the VE, cogging, and friction forces are applied directly to the instrument instead of to the virtual mass when possible: the virtual coupling is an device necessary to simulate inertia, but it also low-pass filters the other degradation models as well as the forces coming from the virtual environment. Therefore, the only case in which these forces are applied on the virtual mass is when the backlash is turned on. In this situation it is necessary to apply VE forces and the appropriate hardware degradation forces directly to the mass since these forces determine the position of the virtual mass in the backlash-gap: if a sinusoidal VE force should work on the tooltip, then backlash in the transmission can cause the motors to move back and forth while the user interface is being held in a rigid position. In this case the VE, cogging and friction models get positional input from the virtual mass, after this position signal has been corrected to stay within the range of the gap - i.e. when backlash is turned on these models receive a positional input signal that can not be more than half a gap width away from the instrument position.

2.2.6.2   Extension of 1 DOF to 3 DOFs

The previously described 1 DOF models are applied to 3 different axes on each of the two instruments: insertion (translation), yaw (rotation) and pitch (rotation). Environmental forces in world frame coordinates need to be transformed to these three axes and back as well as a tooltip position to insertion, yaw and pitch transformation.
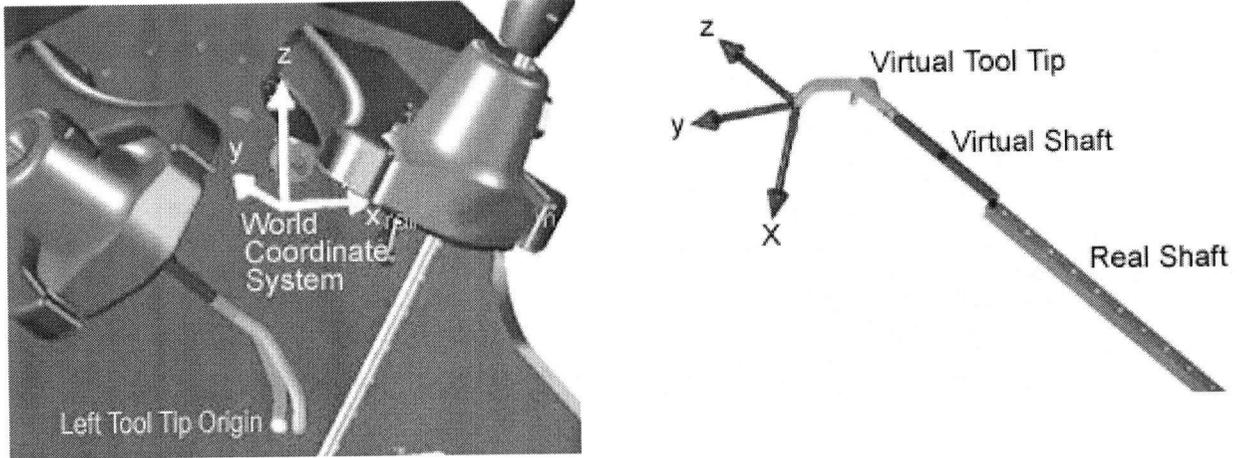
Figure 2-13: The coordinate systems in the LSW: the origin of the world frame coordinate system is defined in between the two trocars (left). The virtual tool shaft can be adjusted in length and rotated (right).

The environmental force F_e is transformed to tool tip coordinates after which these lateral forces are used to calculate insertion force F_i, yaw torque T_{yaw}, and pitch torque T_{pitch}.

$$F_i = F_{x,e} \sin(yaw) * \cos(pitch) + F_{y,e} \cos(yaw) * \sin(pitch) - F_{z,e} \cos(yaw) * \sin(Pitch) \quad (2.16)$$

The effective moment arm has to be calculated to compute the torque around the pitch axis of rotation:

$$\begin{aligned} L_{pitch} &= Insertion * \cos(yaw) \\ T_{pitch} &= L_{pitch} * (F_{y,e} \cos(pitch) + F_{z,e} \sin(pitch)) \end{aligned} \quad (2.17)$$

iIn which L_{pitch} is the effective torque arm of the environmental forces working around the pitch-axis. T_{pitch} is the torque around the pitch-axis. An identical calculation is used for the yaw direction:

$$\begin{aligned} L_{yaw} &= Insertion * \cos(pitch) \\ T_{yaw} &= L_{yaw} * (F_{x,e} \cos(yaw) + F_{z,e} \sin(yaw)) \end{aligned} \quad (2.18)$$

The transformation from insertion forces, yaw and pitch torques to forces in world coordinates is as follows:

$$F_{t,x} = F_i \sin(yaw)\cos(pitch) + \frac{T_{yaw}}{L_{yaw}}\cos(yaw)$$

$$F_{t,y} = F_i \cos(yaw)\sin(pitch) + \frac{T_{pitch}}{L_{pitch}}\cos(pitch) \qquad (2.19)$$

$$F_{t,z} = -F_i \cos(yaw)\cos(pitch) + \frac{T_{yaw}}{L_{yaw}}\sin(yaw) + \frac{T_{yaw}}{L_{yaw}}\sin(pitch)$$

Backlash causes a discontinuity between user input and visual feedback: due to play in the mechanism, the instrument handles can be moved small amounts before the virtual encoder will register it. To simulate this possibly important visual feedback characteristic, the position of the virtual mass is sent back to the graphical environment when backlash is turned on. To visually simulate infinitely stiff gap-walls, the difference between the real and virtual tooltip position is limited to half the gap width. In the VE, the instrument position is defined by two vectors: the position of the tooltip and the orientation of the instrument to calculate where the insertion points of the instruments are. Backlash-adjusted pitch, yaw, and insertion are used to update a positional offset vector and an orientation offset matrix. These are used in the next update to reflect the new perceived position of the instrument and instrument insertion point. To calculate the positional offset in Cartesian coordinates, we use Eq. (2.20):

$$Offset_c = \begin{bmatrix} \cos(-yaw) & \sin(pitch)\sin(-yaw) & \cos(pitch)\sin(-yaw) \\ 0 & \cos(pitch) & -\sin(pitch) \\ -\sin(-yaw) & \sin(pitch)\cos(-yaw) & \cos(pitch)\cos(-yaw) \end{bmatrix} \bullet \begin{bmatrix} Offset_{yaw} * Insertion \\ Offset_{pitch} * Insertion \\ Offset_{insertion} \end{bmatrix} (2.20)$$

And finally the offset in instrument orientation can be calculated by first calculating the offset in orientation of the tool tip frame $R_{offset,t}$ and multiplying this with the transformation matrix from tool tip coordinates to world frame coordinates R_w.

$$R_{offset,t} = \begin{bmatrix} \cos(Offset_{yaw}) & \sin(Offset_{pitch})\cos(-Offset_{yaw}) & \cos(Offset_{pitch})\sin(-Offset_{yaw}) \\ 0 & \cos(Offset_{pitch}) & -\sin(Offset_{pitch}) \\ -\sin(-Offset_{yaw}) & \sin(Offset_{pitch})\cos(-Offset_{yaw}) & \cos(Offset_{pitch})\cos(-Offset_{yaw}) \end{bmatrix} \quad (2.21)$$

$$R_{offset,c} = R_{offset,t} * R_w$$

2.2.7 Software Implementation

The hardware degradation software has a modular design. For each instrument, an instance of our HapticEffect class is created. The HapticEffect class creates instances of the different degradation classes for each degree of freedom. Because of this design, the code can relatively easily be re-used for haptic interfaces with a different number of degrees of freedom. The main limitation to reuse of the code for other devices is the use of Immersion Corp.-specific variables, mainly in the HapticEffect class.
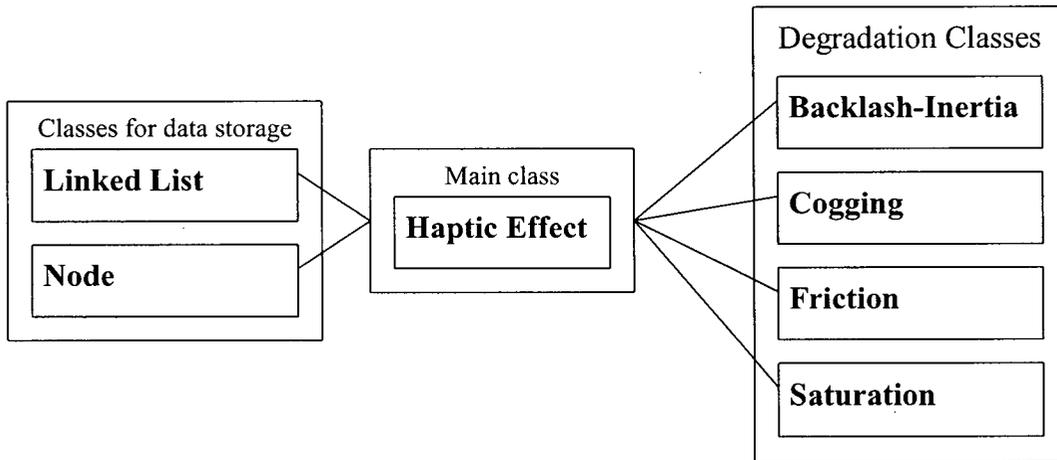
In the VE software, an instance of the Haptic Effect class is defined for each instrument. Text files contain settings for the different degradations. The VE contains a function that updates the haptic interface forces and retrieves new positions. In this VE function our haptic-effect update function is called that updates forces and positions in our hardware degradation model and returns the 'degraded' position and force values to the VE. A version of the code that was compiled into the executable of another virtual environment generates a window with sliders to allow the user to adjust the parameters at any moment during runtime. Because of memory sharing issues between the HapticEffect DLL and the main application it is not possible to use this window with the Reachin surgical VE.

For various reasons, the code was compiled with a Microsoft Visual C++ compiler into a DLL. This DLL is called by the Reachin device driver DLL. A problem was caused by the fact that the Reachin VE was created using a Borland compiler. Even though DLL files are an important component of the MS Windows platform, their format is not standardized; the major compilers do not produce the same DLL format. As a result, software compiled with a Borland compiler can not read DLLs compiled with a Microsoft compiler. Even different versions of the Microsoft compiler generated compatibility issues. There are solutions to work around this problem but many cannot deal with class definitions in C++. An exception is described in an article on bcbdev.com[7] which entails creating a separate DLL that creates virtual COM wrapper to translate between the two DLLs that are compiled with different compilers. We used this method here.

2.3 Analysis and Discussion of Simulation Methods

The models above were implemented on a dual Xeon 2.0GHz with 2 GB of RAM running Windows 2000, Service Pack 4. Table 2.2 lists key model parameters used in the integrated version of the models; the values were chosen through a combination of realistic levels we expect to see in inexpensive hardware components, and constraints imposed by simulation stability. Below, we discuss some of the more interesting features we observed in the simulations individually and in combination.

TABLE 2.2 MAIN PARAMETERS OF THE MODELS ON SIMULTANUOUSLY

		Translation	Rotation
Inertia	Mass	= 0.2 kg	= $2gm^2$
	K	= 600 N/m	= 4 Nm/rad
	B	0.3 critical damping	0.2 critical damping
Backlash	Gap Width	= 1 mm	= 2°
Cogging	Amplitude	0.6 N	0.04 Nm
Friction	Pre-sliding	= 100 μm	=5 millirad
	Stick Velocity	= 5 mm/s	= 0.1 rad/s

---

[7] http://www.bcbdev.com/articles/vcdll2.htm

Implementation of the cogging and saturation setting were very straightforward. The following sections will discuss details of the Friction, Backlash, and Inertia implementation.

2.3.1 Friction

Figure 2-14 shows a measured probe trajectory segment with only the friction degradation turned on. When the instrument handle is moved back and forth, we can see evidence in this data that the state of the friction model for this DOF is oscillating between stick and slip (the vertical dotted lines indicating state transitions have been added manually). The instrument is initially in the slip state while being moved in a positive direction. When the velocity drops below 0.05 rad/s, the model goes to the stick state: a spring is applied such that a continuous friction torque transition will occur between the two states, and the position trajectory immediately flattens. In between t=1.1s and t=1.9s, the model is the stick-state and the friction torque is entirely dependent on the position (pre-sliding stage in real friction). When the friction-torque exceeds a limit T_{stick}, a discontinuous transition to the slip state occurs. Because of the low velocities involved at this point, combined with the discretized velocity measurement, small oscillations between the two states are possible (though not perceptible during haptic simulation).
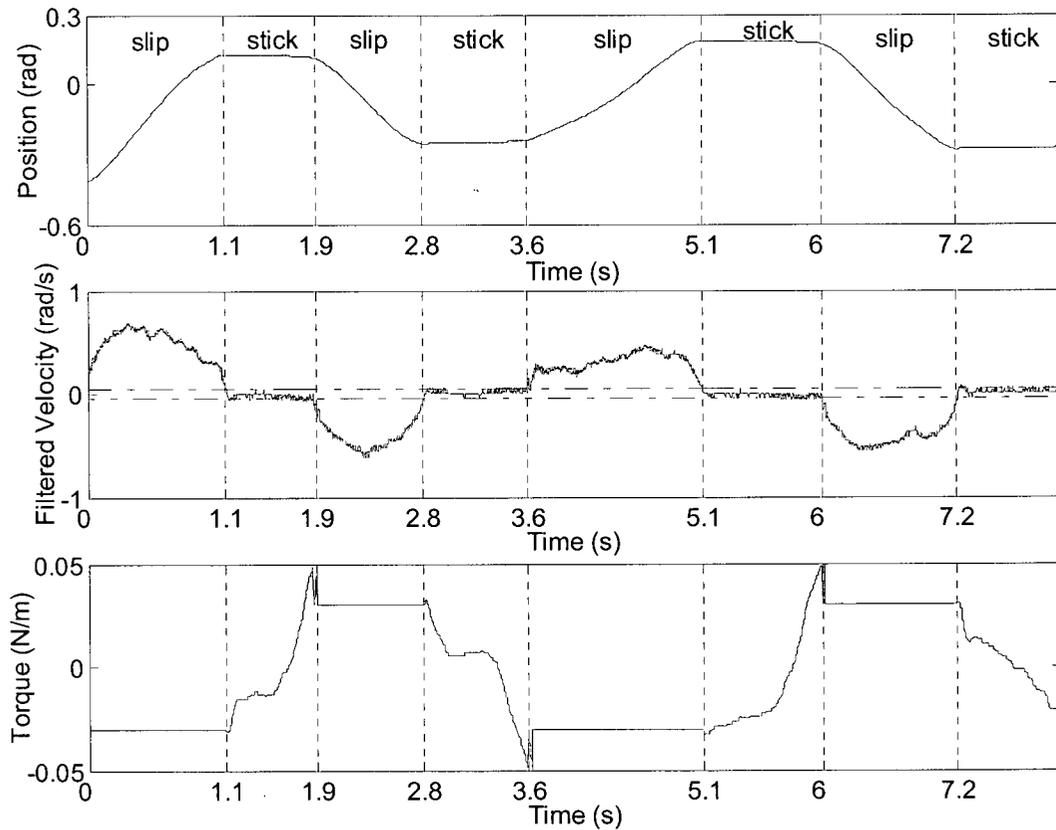
Figure 2-14: Friction model, recorded data. Instrument angle (top figure), instrument angular velocity (middle figure), and friction torque (bottom figure).

2.3.2 Backlash and Inertia

Figure 2-14 shows a measured trajectory segment of the user-controlled probe and the virtual mass with backlash turned on. At t=29.5s, the probe is pushing against one wall of the gap, dragging the virtual mass closely behind it. When the probe stops, the virtual mass continues until the other wall of the gap hits the probe. When the probe starts moving in the other direction, this repeats itself. The backlash gap-width in this example is 1 mm, and the simulated mass increment is 0.2 kg.

While our backlash model is structurally similar to that used in non-haptic simulation [Gerdes, 1995; Jukic, 2001], our simulated wall stiffness is significantly lower than could be expected in real metal on metal contact: K=600 N/m for the linear case. As a rough comparison, a 1 cm^2 contact area of a 1 cm^3 steel block has a K (EA/L) value of 200x10^7 N/m. This is reflected in the backlash model's feel: while there is a clearly perceptible play in the gap, the impact is not as

49

crisp as one would feel in a real system. One remedy (untried through lack of time) might be to apply a force impulse on impact with the mass, as described for crisp simulation of virtual walls [16, 22]. However, the small gap creates a serious risk of wall-to-wall oscillation. Instead, to limit the impact on perceived gap-width, the backlash models' output to the graphical environment is adjusted to have infinitely stiff walls.
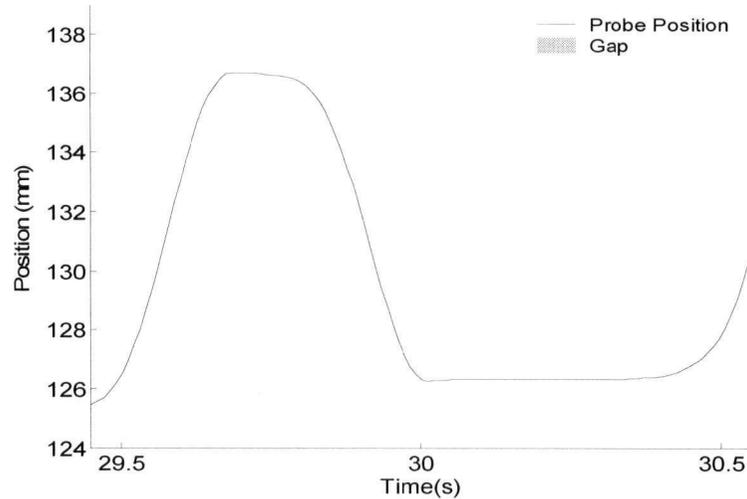
Figure 2-15: Backlash: probe position relative to the backlash-gap in the virtual mass. The path shown is based on data obtained by manually pushing the instruments of the LSW. Both instrument position and virtual instrument position were recorded. The virtual instrument position was expanded to both sides to show the backlash gap.

2.3.3 Computational Load:

The CPU effort required to simulate the various model aspects for the three degrees of freedom we degraded are listed in Table 2.3. Values were obtained by recording the time required to run each degradation independently and without the VE for 10,000 cycles, then substracting the time it takes for the loop to run without the models, and finally computing mean update time. Sharing of computational results by multiple models has the effect that the update time of all effects together is actually 2 μs smaller then the actual sum of the update times of the separate models.

Table 2.3 Computing times of the various degradations. The backlash model cannot be run without the inertia model; and therefore the listed total leaves single inertia out of its summation.

Effect	Update Time (μs)
0. All effects off	19
1. Inertia	+26
2. Backlash + Inertia	+136
3. Friction	+0.5
4. Cogging	+2.3
5. Saturation	+0.1
Total (sum of 1,2-5)	**156**

For the two instruments simulated, the total update time required when all models are running is 312 μs on one CPU (this includes the computational overhead present when all effects are turned off). The VE update frequency dropped in our experiments regularly to 200-500Hz with both CPU's working at 100%. At critical times such as these, the percentage of CPU time spent on running the simulated degradations is roughly 3%-8%.

2.3.4 The Models Combined

When all the models are turned on at the same time, cogging, friction and backlash can all be clearly distinguished. Inertia as an individual degradation is not as strongly present as the other degradations when all are displayed: while it is apparent in the backlash model and the force on impact with the gap-walls can be felt, during constant acceleration in one direction it is hard to distinguish it while the other degradations are present.

Unfortunately, in the dissection task of the VE the haptic interface becomes occasionally unstable while in contact with the tissue. This happens only when the inertia/backlash model and friction model are turned on at the same time and both instruments are interacting with the tissue models. At this point the update rate drops well below 500Hz (150-300Hz). While the exact tissue models used in the simulation are unknown, it is likely that the delays in the haptic loop introduce extra energy as described in Chapter 1 for virtual walls: holding on to tissue is similar to having a tool tip located exactly in between two walls: force fields will increase in either direction instead of just one.

In summary, we have described software that can simulate characteristics of inexpensive hardware on a high-end haptic feedback system. We observed good overall realism of the models at low computational load relative to the VE. Our conclusions based on these observations will be presented in Conclusions and recommendations.

Chapter 3   User studies

3.1 Abstract

This chapter comprises the experimental design and results of a pilot study to examine the extent to which task execution in a VR surgical environment is affected by quality of the force feedback hardware. Expert surgeons performed a VR dissection and clip application task on a simulator displaying with its maximum fidelity, as well as with simulated hardware degradations such as friction, cogging, force saturation, backlash, and/or inertia present. Preliminary results show that the forces exerted by the surgeons were not significantly affected by any of the hardware degradations except for the setting in which no force feedback was present. Median velocities of tool-tip movement, completion time, and path length were unaffected by the different hardware settings, across all surgeons.

3.2 Introduction

The motivation behind this study is the high cost of force feedback hardware: many believe it will have great potential in surgical education. If less expensive hardware can be used without affecting surgical performance, more institutions will be able to afford these simulators. A more extensive discussion of the motivation is available in Chapter 1.

3.2.1 Related work

Previous studies have compared surgical tasks performance under force feedback to no-force feedback conditions. Wagner *et al.* found in a simulated dissection task that the mean force applied to the artificial tissue increased by 50%, and peak forces increased with at least a factor of 2 in the absence of force feedback. The speed of dissection and the precision were not affected [Wagner, 2002]. Moody found that under force feedback conditions, completion time decreased and the precision of suturing in a virtual environment increased compared with no force feedback [Moody, 2002]. She did not report on tissue interaction forces. While both studies found that haptic feedback make a difference in task execution, the difference was found in different performance measures. In both studies participants were students of unspecified background. Heijnsdijk and colleagues studied the influence of force feedback and visual feedback on closing force of a laparoscopic grasper [Heijnsdijk, 2004]. In this study, 9 expert laparoscopic surgeons

and 12 participants with no laparoscopic experience were asked to pull on porcine tissue with 5 N without letting the tissue slip. The experiments were repeated with and without visual feedback and by varying the mechanical efficiency of the grasper from 30% to 90%. The average exerted grasping force was well above what had been established as the minimal grasping force to prevent slip. Visual feedback as well as changes in force feedback did not make a significant difference in exerted forces.

As discussed in Chapter 1, MacLean [Maclean, 1996] and Kilchenman [Kilchenman, 2001] found in independent studies that task performance is largely unaffected by haptic quality in haptic identification tasks. The task in MacLean's research involved identifying resemblance between emulated and real mechanical sliders. In Kilchenman's research participants judged surface hardness and the size of ridges on a surface. New in this study is a larger group of haptic quality factors that are varied and different tasks.

3.2.2 Goals

To objective of this user study is to answer the following questions related to task performance by expert surgeons in a VR surgical environment:

1. Do specific and/or combined hardware degradations result in significant changes in user task execution, when compared to high fidelity haptic rendering?

2. If certain hardware degradations do make a difference, which degradations result in the largest deviations from task performance under high fidelity conditions? E.g. if both cogging and friction significantly affect performance, which of the two affects performance most?

3.3 Methods

3.3.1 The Virtual Environment

Our virtual environment consists of both hardware and software. The software used in these experiments is a commercial product specifically designed for laparoscopic training. It consists of eight different basic training lessons, ranging from camera navigation to bimanual dissection, plus more advanced lessons that train specific parts of a cholecystectomy procedure. Unfortunately, technical details on tissue models and parameters are not publicly available. As

evidence of the high regard with which this product is held, its core haptic technology will soon be incorporated in what is arguably the best-known and most thoroughly studied simulator for laparoscopic surgery, the MIST-VR[8]. The software provides 3 DOF of force feedback on the tip, plus force feedback in the handle when a clip is applied.

Our haptic interface is the Laparoscopic Surgical Workstation (LSW) from Immersion Corp. Immersion is the largest force-feedback hardware company in the world, and the LSW is their top-of-the-line interface for laparoscopic surgery. It provides 2 instruments with 6 DOF each, five out of the six DOF are actuated.

Because both the software and hardware used are arguably the best available at the time of this study, we regard their unaltered behavior as the gold standard for high quality surgical simulation.

3.3.2 Choice of Tasks

The tasks chosen for the experiment are the dissection task and clip application task depicted in Figure 3-1.

3.3.2.1 Dissection Task

Dissection accounts for 25-35% of the time spent in most surgical tasks [Scott-Conner, 1999]. It requires bimanual dexterity: to properly dissect tissue, the non-dominant hand first puts the tissue under tension so that the dominant hand can be used to do the actual dissection. Due to the tension, the tissue will tear more easily.

In the simulator's implementation, a dark structure is shown that is covered with a light colored tissue. On one side, a red dot is shown. The goal of the task is to dissect away the covering tissue to expose a green line that starts at the red dot (Figure 3-1, left). The actual trajectory of the green line can vary somewhat, but it always starts at the red dot on the left side of the tissue and a second line will branch off halfway along the line.

---

[8] Currently marketed by Mentice Corp. as the MIST system.

3.3.2.2 Clip Application Task

Clip application involves similar bimanual coordination: the non-dominant hand will grasp the structure on which the clip will be placed, after which the dominant hand is used to place the clip. Precise motor skills are necessary to place the clip in the correct way. Clip application is an important part of the most commonly performed laparoscopic procedure, cholecystectomy (gall-bladder removal) as well as of more advanced procedures.

In the software implementation, a tube-shaped structure is shown connecting two larger structures. Somewhere on the tube, a black marker indicates the place to apply the clip (Figure 3-1, right). The clip is then placed bimanually.

3.3.2.3 Comments on Tasks

In light of the limited available time from the participants, the number of tasks was limited to two. Other tasks available in the software were rejected because they either took too long to complete, did not require enough technical skills, or simply did not provide any force feedback.

The two tasks chosen do have shortcomings. The software sometimes fails to detect a clip when it looks like it is properly applied. In that case a new clip has to be applied which naturally leads to an increase in completion time. Shortcomings in the software's collision detection and graphics algorithms sometimes cause the simulator to fail to display the proper order of overlaying anatomical structures and instruments, leading to an increase in difficulty of the task since the operator has to correct for objects that are not properly displayed. Another shortcoming of the dissection task was discovered during the actual experiments with expert surgeons. In the VE it is possible to dissect in a manner that is not possible in the OR, but allows for fast completion of the task without any penalties. The Discussion section of this chapter will go into more detail on this.
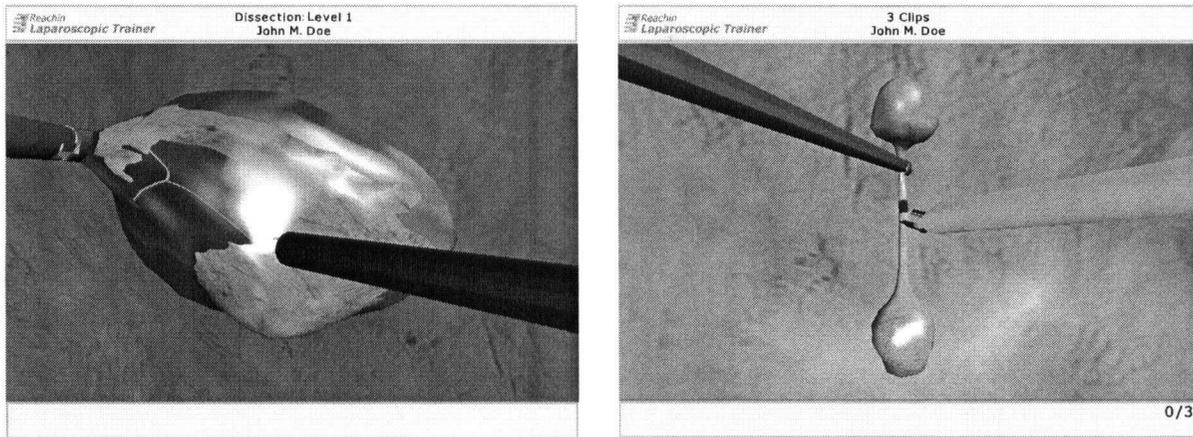
Figure 3-1 **Pictures of the dissection task (left) and clip application task (right). In the dissection task light coloured tissue covers darker sensitive tissue on which a green line is painted. The objective is to fully expose the (green) line by dissecting away the covering tissue. In the clip application task a grasper in the left hand is used to stabilize a duct while the tool in the right hand is used to apply a clip on the spot marked by a black line.**

3.3.3 Choice of Hardware Degradation Settings

The following considerations played an important role in the design of the experiment:

- Potential participants (i.e. expert laparoscopic surgeons) were limited in number and their time and cooperation was hard to get.

- We anticipated potential significant differences between the participants in terms of task execution (not in terms of quality, but in terms of actual kinematics and force profiles; similar to top athletes that play at the same level but have distinct strokes).

Therefore to reduce experimental time, only a limited number of quality settings was chosen, and each effect was only tested at a single level with the exception of No Force Feedback (implemented by setting the force saturation level to 0 N). These settings, which were discussed in detail in Chapter 2, are:

- High Fidelity
- No Force Feedback
- Inertia
- Inertia + Backlash

- Friction
- Cogging
- Force Saturation
- All Degradations

Due to the time constraints we only include one setting with multiple degradations on at the same time. This is the All Degradations setting, which consists of the superposition of the Inertia, Backlash, Cogging, and Force Saturation settings.

3.3.4 Strength of Haptic Degradations

To choose the strength of degradations, the following criteria were considered:

- The levels should be representative of those that might actually be found in a cost-cutting hardware design.
- The degradations need to be strong enough to be perceptible.
- The effects should not be so strong that one would expect ergonomic objections to using the machine.
- While any noticeable haptic degradation may make the tasks less pleasurable to perform, and this cannot be avoided, the effects should not be so strong that it would compromise subject cooperation in any way.
- The individual strengths of the degradations should be unaltered when they are combined in the All Degradations setting.

The final decision on the choice of parameters was made after a preliminary round of experimentations and assessments. The chosen amplitudes and other parameters are listed in Table 3.1. In a pilot study with graduate students, higher amplitudes had been used; amplitudes were scaled down for the final study to simulate more realistic values that might be implemented in actual commercial hardware.

Table 3.1: Parameters of models used in this study.

	Parameter	Translation	Rotation
Cogging	Amplitude	0.45 N	30 mNm
Friction	Amplitude	0.45 N	30 mNm
	Presliding	0.1 mm	0.01 rad
Backlash	Gap Width	2 mm	0.04 rad
Inertia	Mass	0.1 Kg	1 gm^2
Saturation	Amplitude	1 N	N/A

3.3.5 Performance Measures

The following performance measures were used in this study:

- Completion time
- Median absolute tool-tip force
- Peak (95 percentile) absolute tool-tip force
- Median absolute tool-tip velocity
- Path length

Both absolute velocity and force distributions are non symmetrical. The performance measures use the median instead of the mean since the median is less sensitive to outliers.

3.3.5.1 Completion Time

The completion time was chosen because it is a direct intuitive measure of performance. Multiple studies have shown that completion time decreases with experience [Rosser, 1997; Chung, 1998; Den Boer, 2001; Coleman, 2002] and is therefore a good candidate to measure differences in performance. In the Reachin VE, both tools need to be retracted completely to start a task. Completion time was measured from the time that the tools reach 20 mm insertion into the simulated tissue, until the last time the tissue is touched.

3.3.5.2 Median and Peak Absolute Tool-Tip Force

Tissue interaction forces are an important measure of surgical performance. To minimize the risk of tissue damage, it is necessary to exert forces that are not higher than what is required to execute the task. An important part of technical skills training in surgery is to learn to exert the right amount of force. While median force gives a good indication of overall forces applied, the peak force was also chosen as a separate measure since it directly relates to potential tissue damage. To improve robustness of the measure to outliers, we used 95 percentile forces instead of the 100 percentile force. Tissue interaction force is obtained for each tool in as a vector with x, y, and z coordinates. The absolute tool tip force is calculated as follows:

$$F = \sqrt{F_x^2 + F_y^2 + F_z^2} \tag{3.1}$$

This vector also contains all the zero force values for the time that the tools are not into contact with any tissue. Since this would result in force percentiles reflecting the amount of time that the tool-tip is not in contact with the tissue, force values close to zero (F_{abs}<1 mN) were filtered out.

3.3.5.3 Median Absolute Tool-Tip Velocity and Path Length

Velocity and path length are commonly used to describe tool-tip kinematics. The absolute value of velocity was taken according to Formula (3.2) to obtain a single measure that is also robust against the small variations in location of the target in each task.

$$v_{abs} = \sqrt{v_x^2 + v_y^2 + v_z^2} \qquad (3.2)$$

Measures calculated by the Reachin software were not used, since we were unable to learn what algorithms are used to calculate them and they do not provide results for segments of data.

3.3.6 Experimental Design

3.3.6.1 Participants

In order to assure expert performance, only expert laparoscopic surgeons that were performing laparoscopic procedures on a regular basis were allowed to participate in the experiments. Residents and surgical fellows were excluded from the experiments.

3.3.6.2 Experiment Size

Based on a pilot study with 3 graduate students from the Mechanical Engineering Department, we predicted that surgeon participants would be able to complete 3 repetitions of all settings in the dissection and 9 repetitions of the clip application task in roughly 90 minutes. The assumption was that the experienced surgeons would be able to execute the tasks at least as fast as the inexperienced graduate students. Unfortunately, this was not the case: the 1st participant took 3 hours to complete these tasks. Therefore, the experimental design was scaled back to have all participants do only those settings that were judged most important (High Fidelity, No Force Feedback, and All Degradations), plus two of the other five conditions (see Table 3.2)

Table 3.2 The settings to be executed by eight participants. This table applies to both tasks.

Participant	Hifi	NoFFB	All Dg	Bckl	Cogg	Frict	Inertia	Sat
Surgeon 1	*	*	*	*				*
Surgeon 2	*	*	*		*	*		
Surgeon 3	*	*	*			*	*	
Surgeon 4	*	*	*		*			*
Surgeon 5	*	*	*		*		*	
Surgeon 6	*	*	*	*		*		
Surgeon 7	*	*	*	*			*	
Surgeon 8	*	*	*			*		*

While we started to execute this design, repeated failure of the haptic hardware and the extreme difficulty in recruiting participants led to the decision to limit this thesis research to the first 3 participants only. Participant 1 and 2 performed both tasks, while participant 3 only performed the dissection task because the haptic interface stopped working at the end of the dissection session. Since no data was collected from the other 5 subjects, this limits the power of the statistical tests that are used, and makes the study more of a pilot study.

3.3.6.3   Instructions

One of the main constraints in this study was the limited time available from participants. Therefore we wanted to minimize the need for practice. Since laparoscopic techniques are not standardized but vary between surgeons, each participant was asked to use the technique that he/she use in the OR. Since the statistical tests used in this study have a within-subject design, differences between surgeons present in all settings do not affect the statistical outcome.

Also, if different techniques actually affect how haptic quality affects performance, it is beneficial to have this variability in the experimental population since this would reflect variation in the larger population.

The participants were allowed training time to repeat the tasks until they felt their performance was not improving anymore. Duration of training time used was recorded. For consistency during the experiment, they were instructed not to change their technique once the training period had ended.

3.3.7 Statistical Methods

We first want to test the hypothesis that none of the degradations make a difference. If this hypothesis was going to be rejected for any performance measure, we then wanted to do pair wise comparisons between the High Fidelity setting and the other settings to determine which haptic degradation is significantly different from the High Fidelity setting.

Since we expected variability between subjects, a within-subject design was chosen to avoid losing power due to inter-subject variability. Power was especially important since we had to reduce our experiment both in number of participants, and in the number of settings each participant executed. The most common parametric test for within-subject designs is the repeated measures ANOVA test. However, it had very low power for the data in our study for two reasons. First, the power of a repeated measures ANOVA does not increase with multiple repetitions. Second, our data was not symmetrical and Mauchly's test revealed that it strongly violated the sphericity assumption (Sphericity refers to the equality of variances of the paired differences between conditions). When data does not meet the ANOVA assumptions, compensation techniques can be applied but these generally lead to lower power. The multivariate ANOVA test does not require sphericity of the data and can be a more powerful substitute of the repeated measures ANOVA. Unfortunately, it loses this advantage in the case of a small number of participants compared to factors [Maxwell, 2003], as is the case in this study.

3.3.7.1   Friedman Test

The Friedman test [Friedman, 1937] used in this study is an alternative to the repeated measures ANOVA when the assumptions of normality or equality of variance are not met [Zar, 1996]. A significant result of the Friedman test indicates that in a set of k dependent samples (k represents the eight hardware settings in our case), at least two of the samples represent populations with different median values [Sheskin, 2000]. The relative power of the Friedman test compared to

the ANOVA (in case ANOVA assumptions are met) is $\dfrac{3k}{\pi(k+1)}$, which is 0.80 in our case.

Therefore the Friedman test should only be used when the assumptions of the ANOVA are not met. An example of how the Friedman test works is given below with fictional data.

	High Fidelity		No Force Feedback		All Degradations	
	Force	Rank	Force	Rank	Force	Rank
Dissect , S1, RH	0.5	3	2	1	1	2
Dissect, S1, LH	0.3	3	1.5	2	1.6	1
Dissect, S2, RH	0.8	2	1.2	1	0.7	3
...	...	...	...	...	...	...
Dissect, S3, LH	0.7	3	1.0	1	0.9	2
Sum		17		10		13

The examples shows forces obtained for surgeons completing the dissection and clip application task. Forces are compared within each row and ranked. The largest force value is ranked 1. The total sum of column ranks is calculated. The underlying idea of the test is that if the columns (hardware settings) have no effect, the column-wise sums of ranks should be roughly equal. To put a number on the term 'roughly', the following equation is then used to compute the chi-squared approximation of the Friedman test statistic [Sheskin, 2000]:

$$X_r^2 = \frac{12}{nk(k+1)}\left(\sum_{j=1}^{k}\left(\sum R_j\right)^2\right) - 3n(k+1) \tag{3.3}$$

If the calculated chi-squared value is greater than the chi-squared value at the specified level of significance, performance measures in at least two of the hardware settings have different median values. When multiple repetitions per cell are made, ranking is performed on all repetitions together [Zar, 1996].

If the result is significant, pair wise comparisons can be performed to calculate which medians are significantly different.

There was a complication in our experiment due to the fact that we had to scale the experiment down from each surgeon doing all 8 settings to each surgeon doing 5 settings. This made it

impossible to apply the Friedman test across all surgeons and all settings at once. Therefore an alternative approach was taken in which first the Friedman test was applied to data of each surgeon individually. If for at least one surgeon the null hypothesis of equal medians between settings was rejected, 7 pair-wise comparisons were performed between the High Fidelity setting and each of the other settings. Alpha values were adjusted to reflect the increased chance of a type-I error when multiple comparisons are made; if the chance of incorrectly observing a difference is 5% per observation, the chance of incorrectly observing a difference over 7 observations would be 30% ($1-0.95^7$). Therefore only P-values lower then ($1-\sqrt[7]{0.95}$)=0.0073 are judged significant.

3.3.7.2   Other Tests Considered

Before we turned to the Friedman test, a significant effort was made to develop a statistical methodology based on the Kolmogorov-Smirnov test [Smirnov, 1939] and bootstrapping [Effron, 1979]. This methodology was attractive for the following reasons:

- No assumption is made about the distribution of data.
- It uses a dimensionless difference measure that allows comparing across difference performance measures (differences in force can be compared to differences in velocities).
- This difference measure can be used for relative ranking of hardware degradations

Unfortunately, the power of the test proved lower than the Friedman test: we were not able to detect any significant differences. This is most likely due to a violation of the assumption of the bootstrapping method that the samples form a good representation of the real population; in our case we had 2 or 3 participants as a representation of the population of expert laparoscopic surgeons. For researchers who follow in this path, Appendix I: Data analysis with the Kolmogorov-Smirnov Test and Bootstrapping. provides a description of the methodology and a more detailed discussion of why it did not work for our data.

The standard test to compare means of different groups to a control (in our case the High Fidelity setting), is Dunnett's test for comparison with a control [Dunnett, 1964] . One of the assumptions of this test is that the population variances are equal which is not the case in our data. Therefore we used the Friedman test for the pair wise comparisons as well.

3.4 Results

The results section is organized as follows. Following a report of observations made during the experiments, we present: cumulative distribution plots of tool-tip forces and velocities, and data plots of median and 95 percentile forces, completion time, median velocities, and path lengths that show values of the repetitions. Finally, learning effects are presented.

3.4.1 Data

3.4.1.1 Experiment

We originally intended to recruit 8 participants. We initially intended to recruit experienced surgical residents. While surgical residents are busy the whole year through, our experiments coincided with exams. As a result none of the invited residents responded to our invitation. A subsequent effort to recruit expert laparoscopic surgeons resulted in 3 participants. Each of them had more than 5 years of experience with laparoscopic procedures. All three are male. As explained in the methods section, each surgeon was assigned to 5 out of 8 settings. The settings were presented in random order. Table 3.3 shows the assignment of settings to the participants.

Table 3.3 Assignment of settings to the participants.

Participant	Hifi	NoFFB	All Dg	Bckl	Cogg	Frict	Inertia	Sat
Surgeon 1	*	*	*		*		*	
Surgeon 2	*	*	*		*	*		
Surgeon 3	*	*	*	*				*

Surgeon 1 first performed the clip application task first, then the dissection task. Surgeon 2 also performed both tasks, but the dissection task first. Surgeon 3 only performed the dissection task.

3.4.1.2 Task Duration

A factor-of-two difference in completion times between mechanical engineering students in the pilot study and Surgeon 1 in the final study caused the total experimental duration of the first participant to approach 3 hours. Due to a failure in data collection, fifteen minutes of the experiment had to be repeated on another day. Assuming that the remaining surgeons would

65

need a comparable amount of time, we then reduced the experiment to the above schedule. Surgeon 2 completed the experiment in 1.5 hours. Surgeon 3 took 60 minutes to complete the dissection task, after which the grippers of the haptic interface became unresponsive, which made it impossible to perform the clip application task.

3.4.1.3 Cumulative Distribution Plots

Insight into the complete distributions of forces and velocities can be obtained from the cumulative distribution functions (CDFs) of the data (Figure 3-4 and Figure 3-5). These plots show the probability that a value takes a value equal to or less than the value presented on the x-axis. The value on the x-axis that corresponds to a y-axis value of 0.5 is the median value.

Figure 3-2 and Figure 3-3 show the variability in forces between repetitions of the High Fidelity, No Force Feedback and All Degradations settings compared to the inter-setting differences. The data shown are the CDFs from the right hand instrument of Surgeon 1 executing the dissection task and clip application task. Both plots show a general trend we noticed: the variation between repetitions of each setting is large compared to the differences between settings. In general the only exception is the No Force Feedback setting which can be quite different from the High Fidelity and the other settings. Note that Surgeon 1 executed the task under 5 conditions. Data from the other two settings are not shown to avoid cluttering of the plot. The other 2 settings show trends similar to the High Fidelity and All Degradation setting.

CDFs of Surgeon 1 repetitions of dissection in 3 settings

Legend:
— High Fidelity
— · No Force Feedback
····· All Degradations
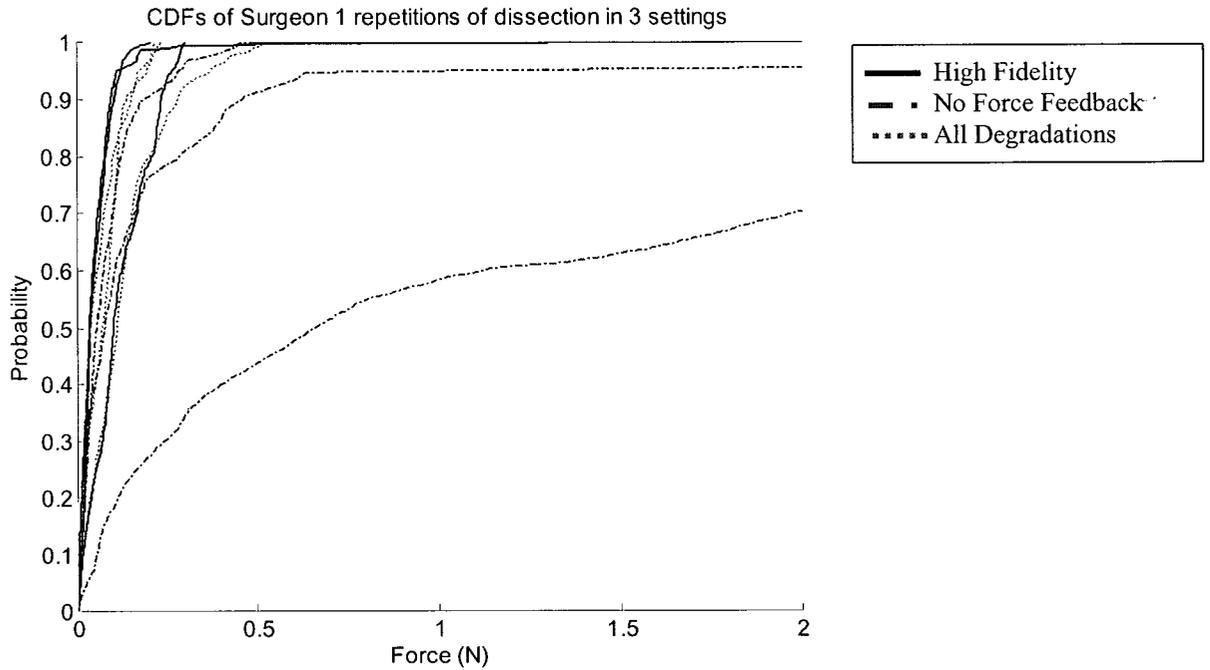
Figure 3-2   Cumulative distribution functions of forces exerted by the right hand tool by Surgeon 1. Three repetitions were executed under 5 conditions. To avoid cluttering the plot only data from the three most important conditions are shown here.

CDFs of Surgeon 1 repetitions of clip application in 3 settings

Legend:
— High Fidelity
— · No Force Feedback
····· All Degradations
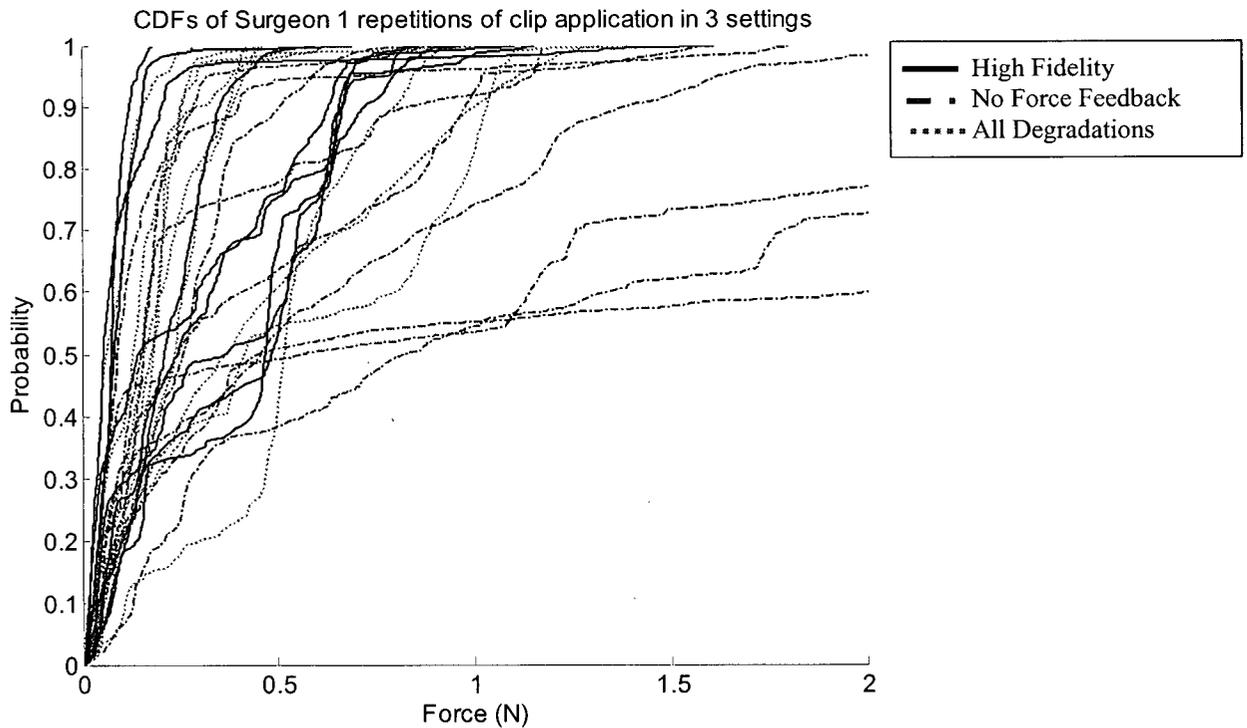
Figure 3-3   Cumulative distribution functions of forces exerted by the right hand tool by Surgeon 1 in the clip application task. Nine repetitions were executed under 5 conditions. To avoid cluttering the plot only data from the three most important conditions are shown here.

67

To limit the number of plots, data from multiple repetitions by a single surgeon are grouped together into a single curve, and curves for the different settings are included in a single plot. Since repetitions vary in duration and thus in number of samples taken, all repetitions were resampled to have the same number of samples as the repetition with the lowest number of samples, which in all cases exceeded 1500. Finally, x-axes on all plots have the same range for easier comparison.
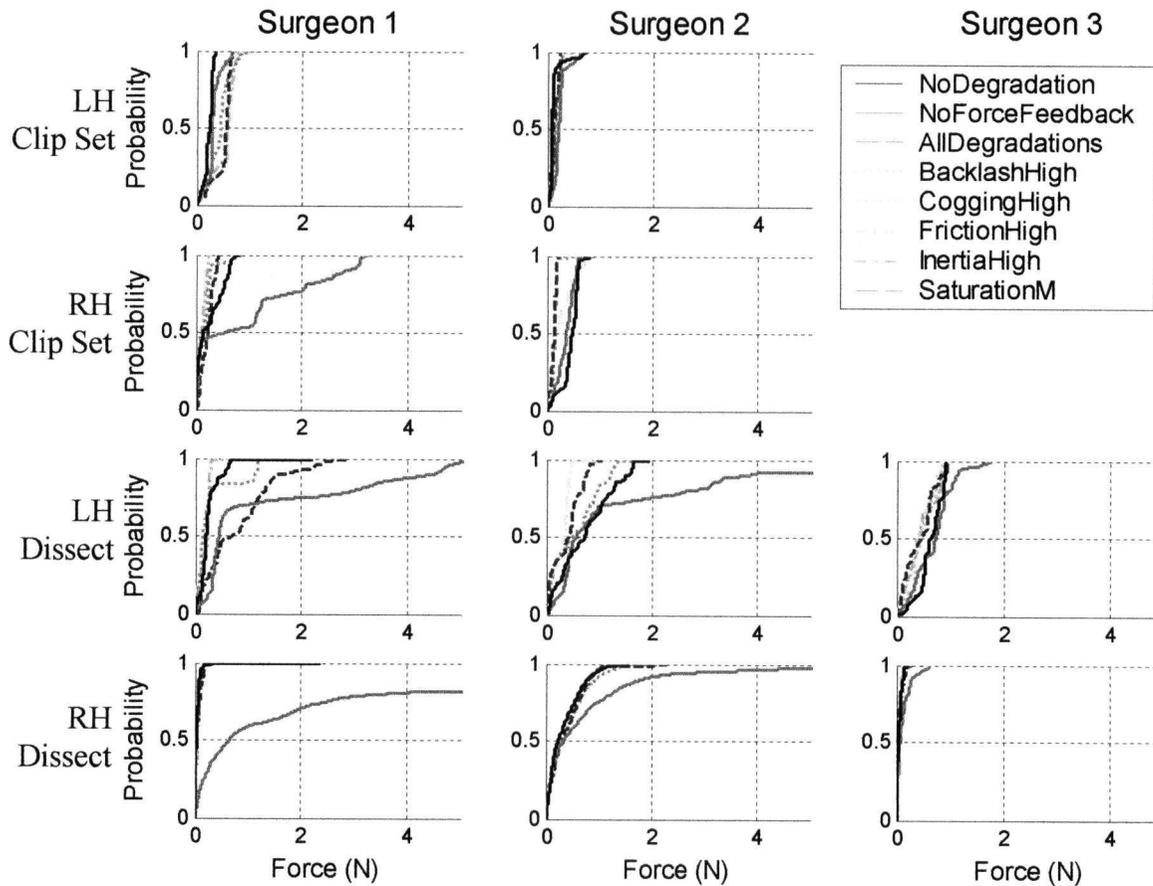
Figure 3-4: Cumulative distribution functions of tool-tip forces.

In the dissection task, the right hand of all surgeons is remarkably consistent between the different settings with the exception of the No Force Feedback setting (although the variability between repetitions does not show up in these graphs, they do indicate average behavior *between* settings). The three surgeons all have different force profiles. The left hand shows more variability in force profiles, and individual force profiles are not as distinct in the left hand. Surgeon 3 seems least affected by the different settings, both for the right hand and left hand.

In the clip application task, forces generally do not exceed 1 N. The exception to this is the forces exerted by Surgeon 1 in the No Force Feedback setting.
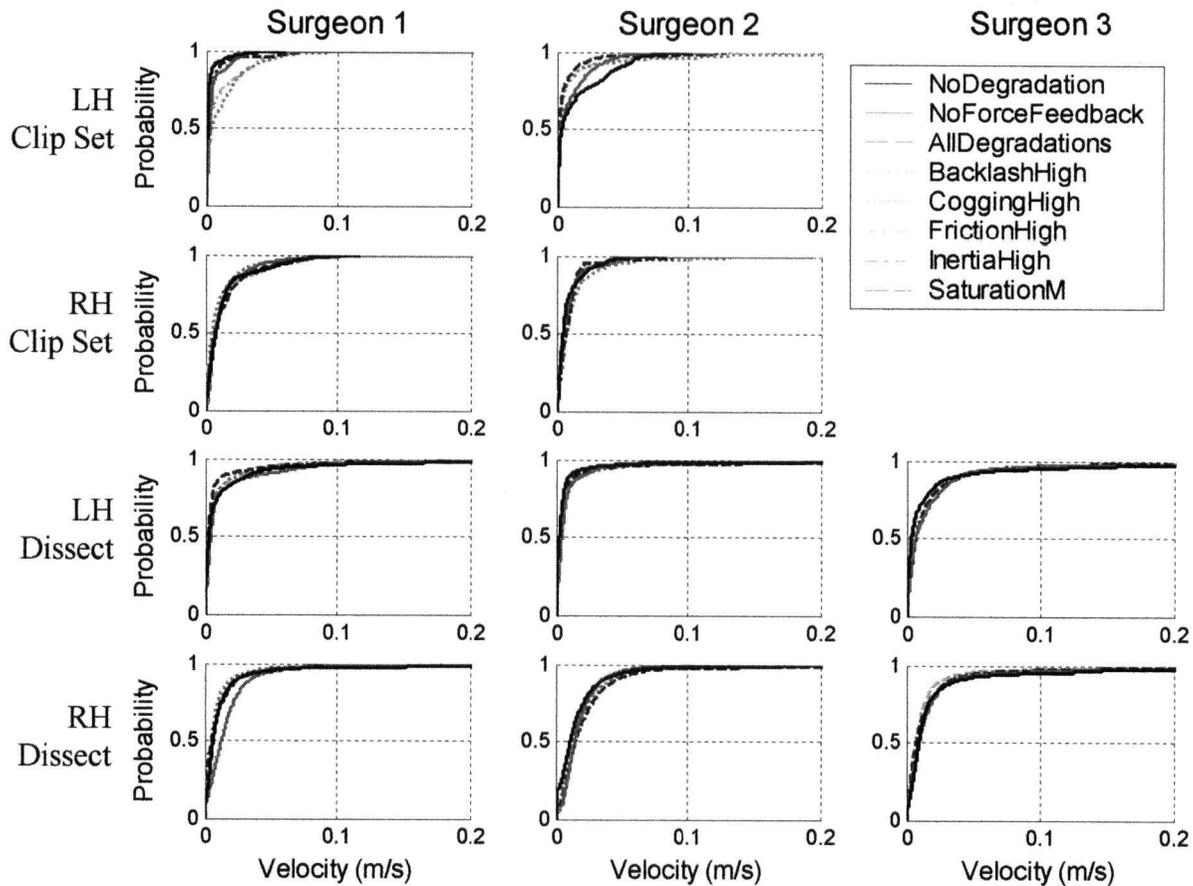
Figure 3-5 Cumulative distribution functions of absolute velocities.

Velocity CDFs look very similar along the different settings, except for the left hand clip setting data for Surgeon 1 and 2, and the No Force Feedback setting in right hand dissection Surgeon 1, and the All Degradations setting in the left hand dissection task of Surgeon 1.

There are clearly visible differences between surgeons, tasks, and hands: for example Surgeon 2 has a different velocity CDF than Surgeon 3 in left hand dissection, but both surgeons are consistent in maintaining their velocity profile over the different settings.

3.4.1.4 *Raw Data Plots*

The following plots show the median tool tip forces, 95 percentile forces, completion times, median velocities, and path lengths. Each data point represents an individual repetition of a task. The data are ordered per surgeon. For the clip application data, for each surgeon a bar shows the

range and the 25 percentile, 50, and 75 percentile values over the different repetitions of the measure of interest. Because in the dissection task only 3 repetitions were obtained, only the range is shown. Comments about trends follow each plot.
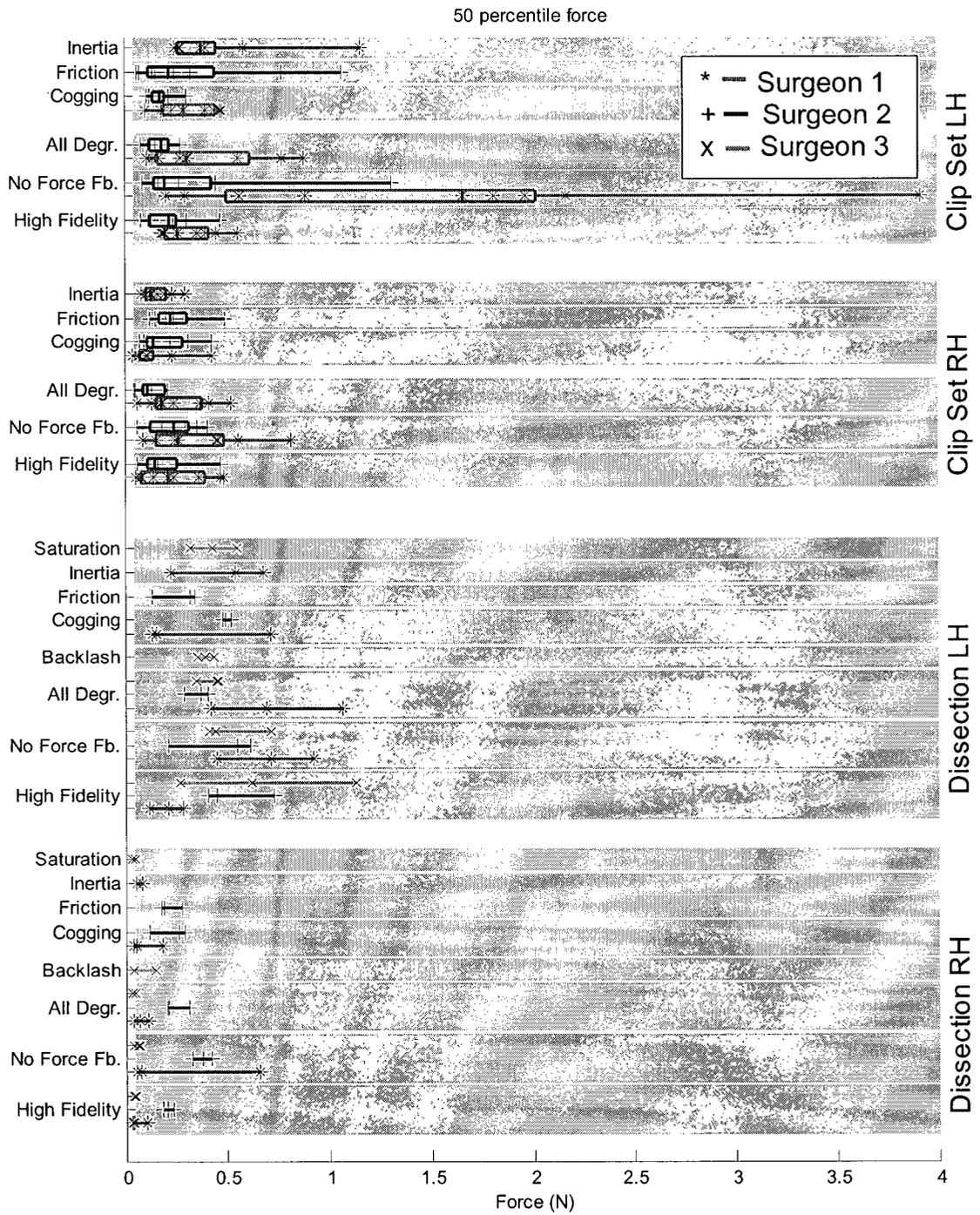
Figure 3-6   Median tool tip forces. Each data point represents a repetition.

50 percentile forces (Figure 3-6):

- Dissection: Right hand forces and range of forces within the same setting are small compared to left hand forces.

- Right hand Surgeon 3: Range of forces within the same setting are very small. Left hand variations are more in line with the other two participants.

- Clip application: significant range of forces over repetitions compared to differences between surgeons and between settings.
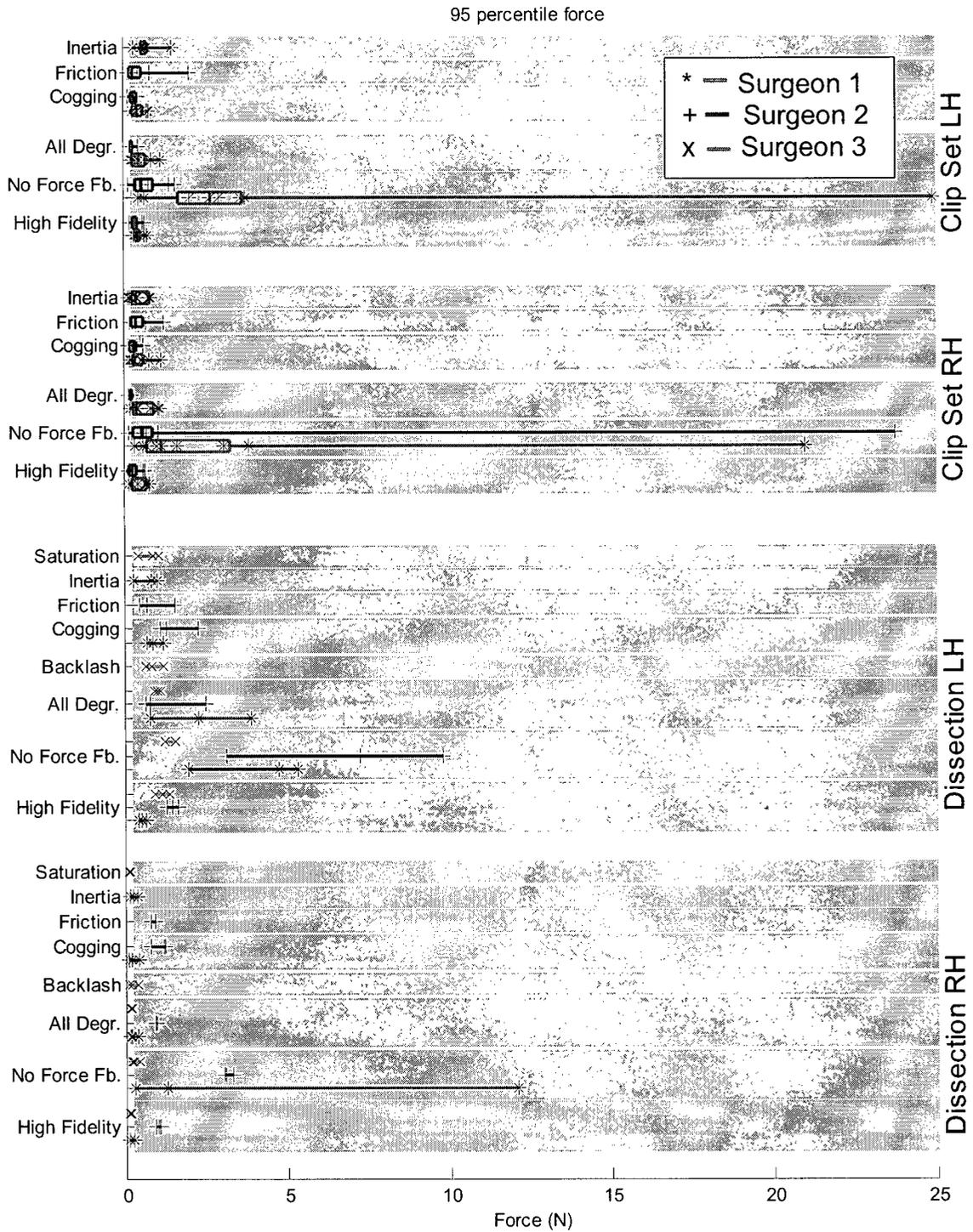
Figure 3-7 95 percentile tool tip forces. Each point represents a repetition.

95 percentile forces (Figure 3-7):

- Right hand dissection: differences between surgeons and between settings are larger than differences between repetitions at the same setting (exception Surgeon 1, No Force Feedback setting).

- Surgeon 1 and Surgeon 2: the largest variability between repetitions is present in the No Force Feedback setting.

- Dissection: higher forces and larger variability between repetitions in left hand compared to right hand.

- Dissection: Surgeon 2 often exerts the highest forces.
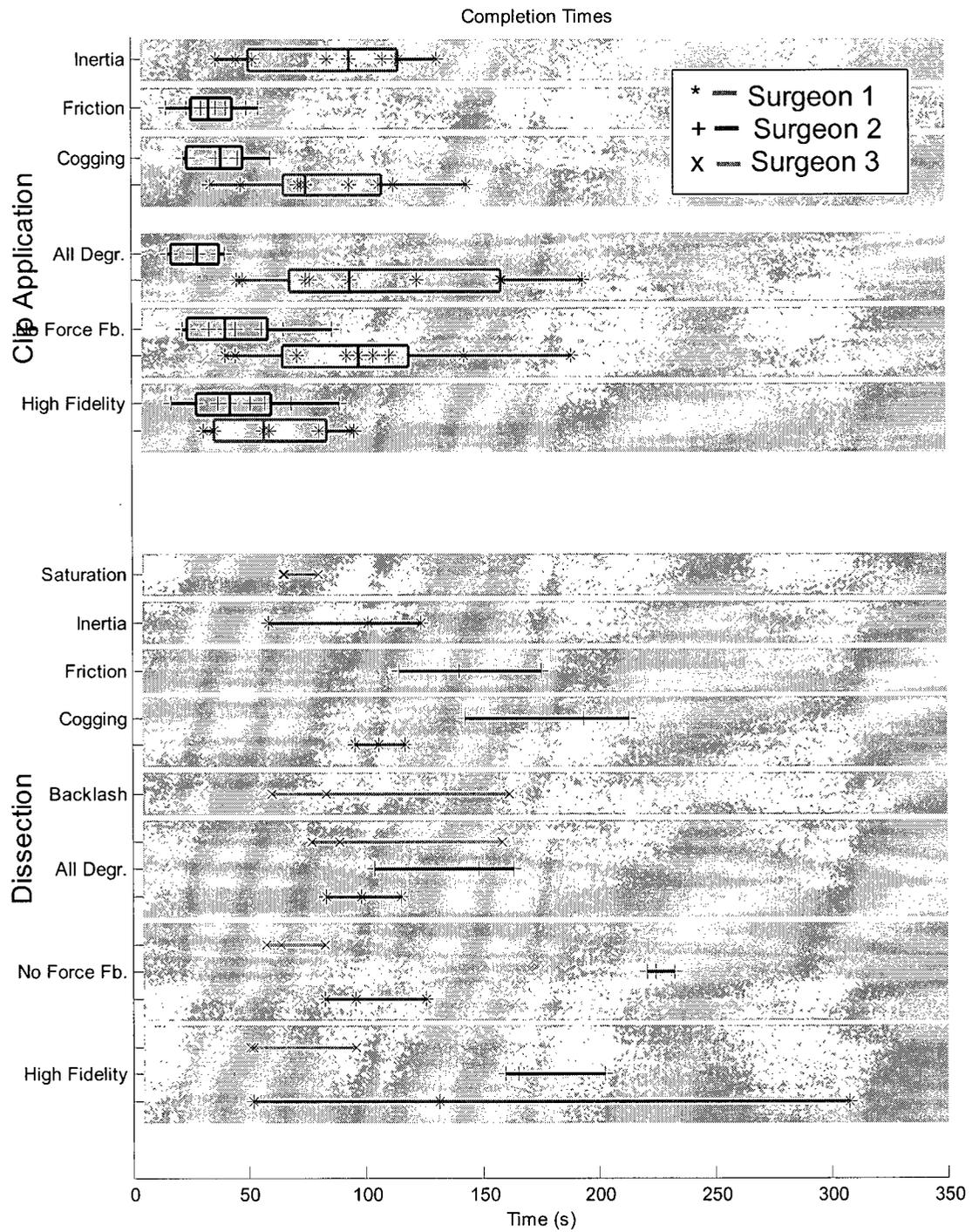
Figure 3-8   Completion times. Each point represents a repetition.

Completion times (Figure 3-8):

- There is a wide variability between completion times of the repetitions.

- Surgeons do not seem to maintain relative speed compared to other surgeons when going from one task to the other: On visual inspection, Surgeon 3 seems fastest in the dissection task and Surgeon 2 slowest, while in the clip application task Surgeon 2 looks fastest.
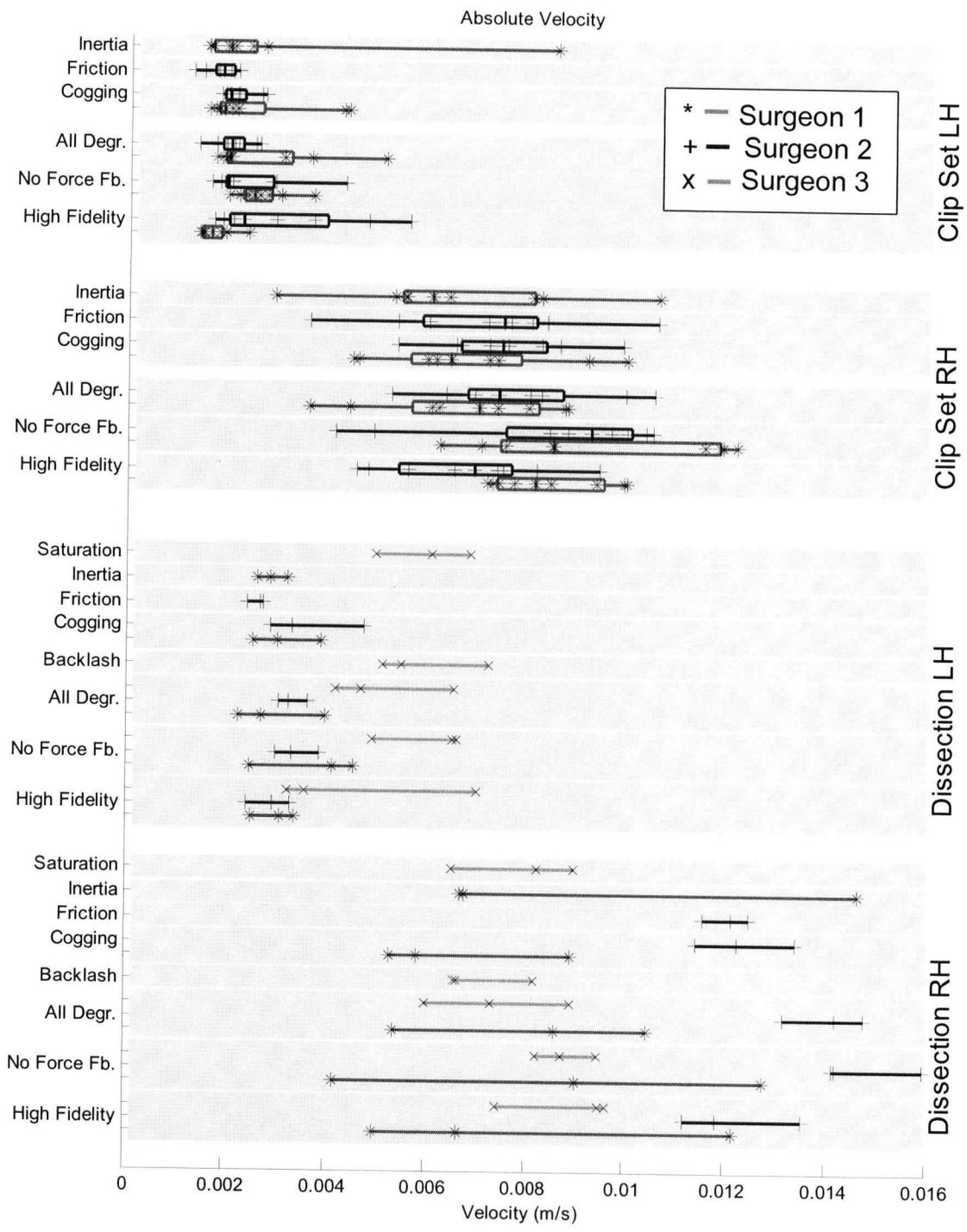
Figure 3-9   Median absolute velocities. Each data point represents a single repetition.

Median absolute velocity (Figure 3-9)

- Median velocities are lower for left hand, though magnitude of this difference is dependent on the surgeon

- In the clip application task, the two surgeons' median velocity range overlap considerably, whereas in the dissection task each surgeon has a more distinct profile: the difference between the left hand and right hand of surgeon 3 is small compared to the other two surgeons.

- In the dissection task, as in the clip task, Surgeon 2's right hand moves consistently faster than Surgeon 3, but the inverse relation is true in the left hand.
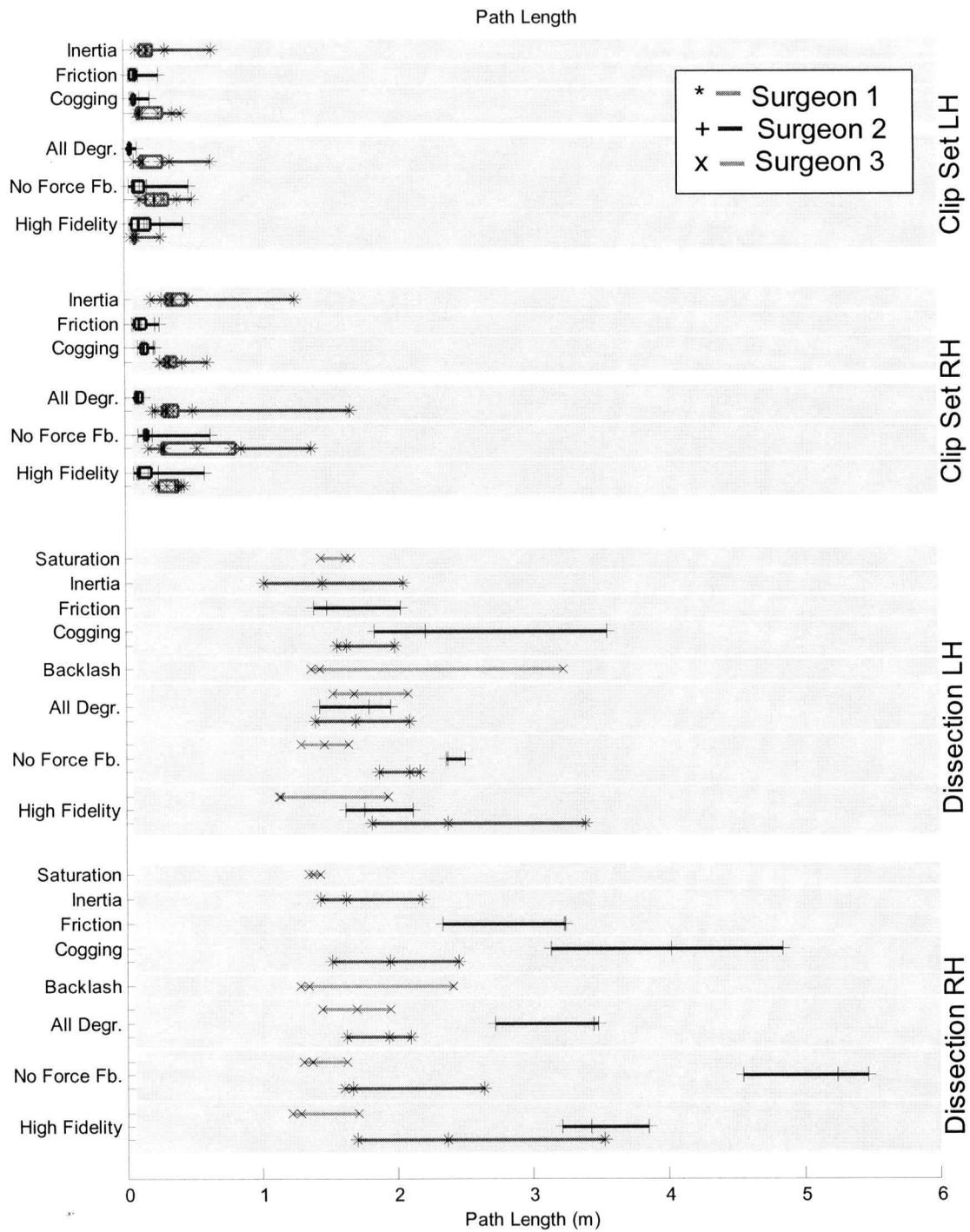
Figure 3-10 Path lengths. Each data point represents a repetition.

Path length (Figure 3-10)

- Clip application: Surgeon 2 shows overall shorter path lengths than surgeon 3.

- Dissection right hand: Surgeon 2 shows longer path lengths than the other surgeons.

- There seem to be no other clear difference in path length due to hardware setting, hand, or surgeon.

- The clip application task shows shorter path length compared to the dissection task.

3.4.1.5 Comparison Plots

In the previous plots, some interesting trends were noticed. We will examine these trends in further detail in the remainder of this section. The first plots (Figure 3-11Figure 3-12) show the differences in exerted forces between the left hand and the right (dominant) hand. Figure 3-13 shows the differences in completion times between the surgeons in both tasks. All plots show the 25th percentile, median and 75th percentile points in a rectangle, with whiskers extending to either the most extreme data points or to the most extreme point within this range if this would result in a whisker length greater than 3 times the inter-quartile range (75th percentile point – 25th percentile point). Data points outside of this area are treated as outliers, and their location is indicated by the '+' sign.
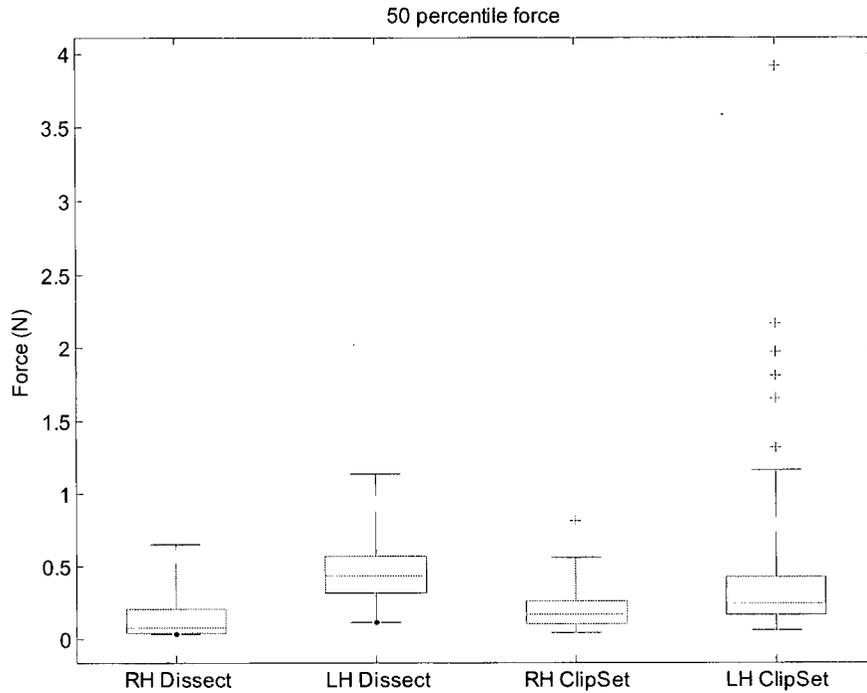
Figure 3-11 Comparison of the median in tip force exerted by the left hand versus the right hand. Since the bars represents data from all participants and all quality settings, the dissection bars represent 45 data points, and the clip application bars represent 90 data points each.

A Friedman test showed that the medians of the 50 percentile forces in the left hand are significantly higher than in the right hand for dissection (P=6E-9) and clip application (P=1E-5). It should be noted that participants and quality settings were grouped together and that these differences have added to the range of data.
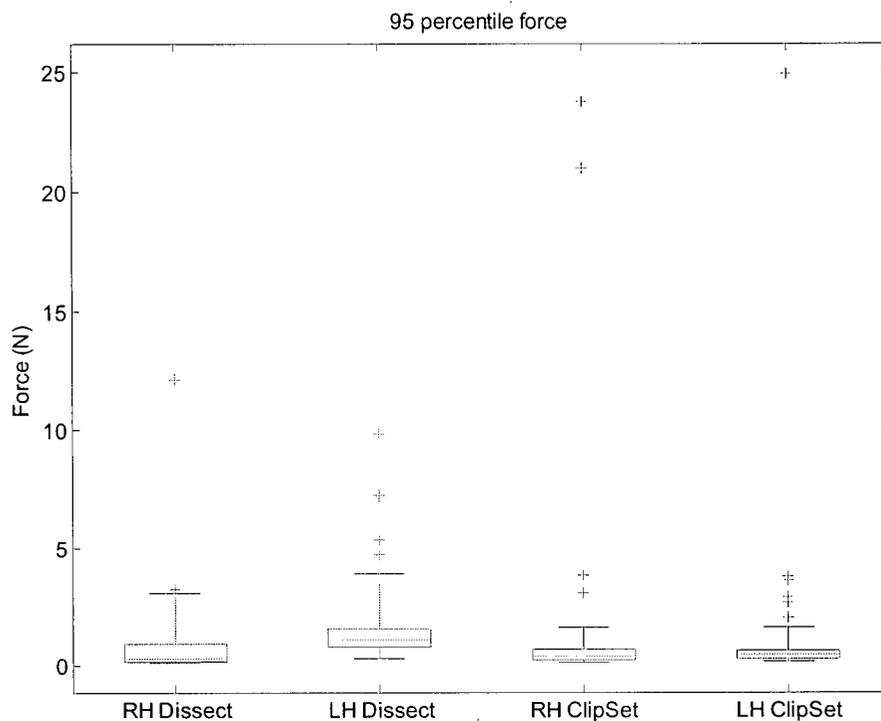
Figure 3-12 **Comparison of 95 percentile tip force exerted by the left hand versus the right hand. The dissection bars represent 45 data points, the clip application bars represent 90 data points each.**

In Figure 3-12, 95 percentile forces from surgeons at different hardware settings were added together. While the median value of the left hand is high compared to the right hand in dissection (P=4E-6, Friedman test), there was no significant difference in median of the data for the clip application task.
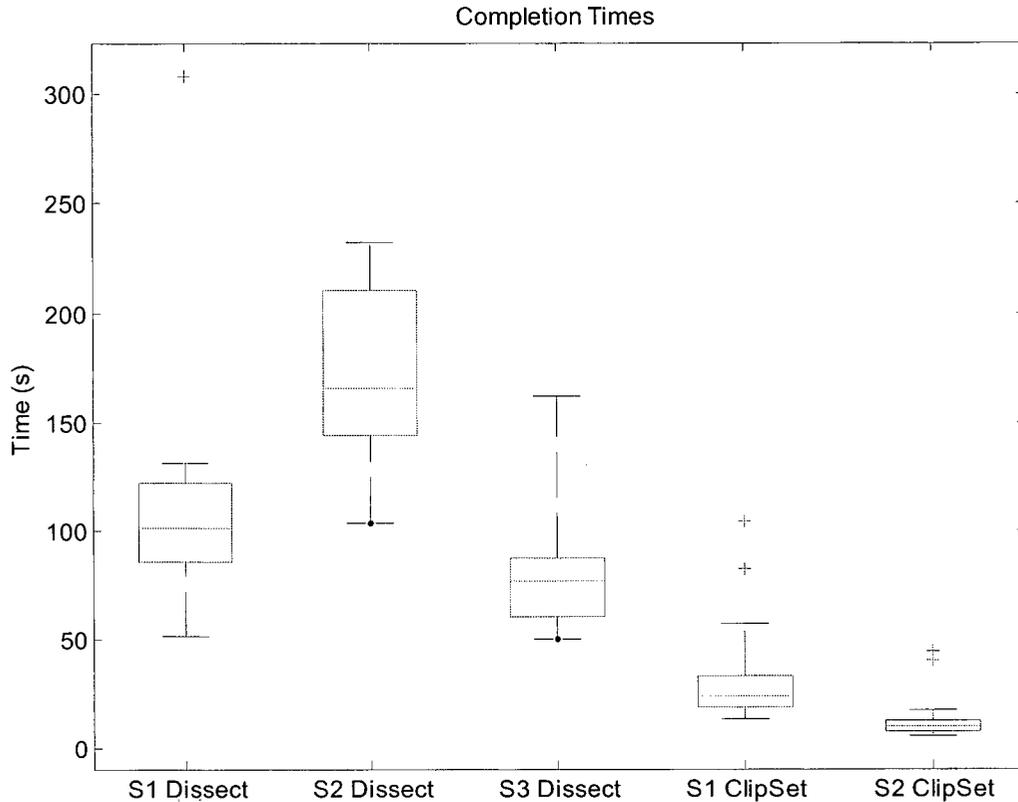
Figure 3-13 Completion times compared between surgeons. Each boxplot consists of 15 data points for dissection and 45 data points for the clip set task.

A second point of interest is the differences between surgeons. A Friedman test applied to the data from all 3 surgeons in the dissection task resulted in a significant difference in mean completion times between at least 2 surgeons ($P=9E-5$). Pair wise comparisons using the Friedman test showed that Surgeon 3 was statistically faster then Surgeon 2 ($P=1E-4$), but there was not statistical difference in median completion time between Surgeon 1 and 3 in dissection, while completion times of Surgeon 1 were statistically different from Surgeon 2 ($P=0.005$). The relative speed of Surgeon 1 to Surgeon 2 was opposite in the clip application task (there was a statistical difference in medians with $P=1E-9$).

3.4.1.6   Miscellaneous Observations on Task Execution

Surgeon 1 noted that even though he did not directly recognize the structure in the dissection task as an anatomical structure occurring in the human body, he would treat is as delicate tissue, because it most resembled a fluid filled sac.

Surgeon 3 used a slightly different approach than the other surgeons in dissection: he used the hook (Figure 3-14) upside down. He would do this to prevent the sharp point of the hook sticking into the delicate tissue below the fat. In the OR, he would either do this, or stick the hook into the fat, and pull it away from the under lying tissue before burning. Since this is not possible in the simulation, he used the 90 degree hook of the tool to burn tissue with.
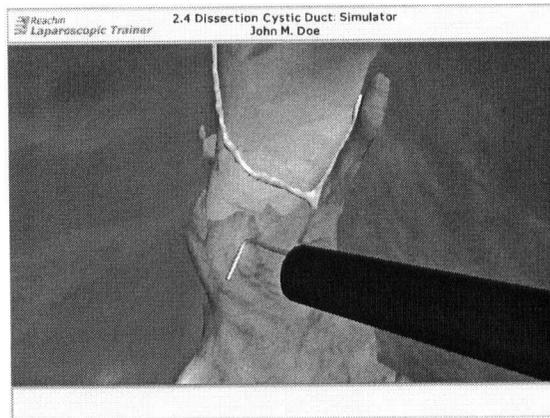
Figure 3-14   The hook of the cautery tool.

3.4.2 Statistical Analysis

3.4.2.1   *Friedman Test*

The Friedman test was first used to test the null-hypothesis that the hardware degradations did not have a significant effect on the performance measures for each of the individual participants. To test for this, a separate test was applied to data from each participant. A significance level of $\alpha = 0.05$ was chosen.

Table 3.4:   P values for intra-surgeon Friedman tests. A significant P-value indicates that it is likely that the performance measures have different medians.

	50p force	95p force	Time	50p velocity	Path length
Diss Surgeon1	0.10	**0.006**	0.7	0.86	0.11
Diss Surgeon 2	0.06	**0.003**	**0.001**	**0.007**	**0.003**
Diss Surgeon 3	0.57	**0.01**	0.06	0.6	0.17
Clip Surgeon 1	**0.001**	**6E-5**	0.16	**0.008**	0.07
Clip Surgeon 2	0.12	**0.002**	**0.0002**	0.35	**0.003**

For all participants, the null hypothesis was rejected based on the 95 percentile force measure. In the dissection task, completion time, median velocity, and path length of Surgeon 2 were affected by the hardware settings as well. In the clip application task, both Surgeon 1 and Surgeon 2 showed significant differences in performance measures other than 95 percentile force, but different performance measures affected each surgeon.

Subsequently, pair-wise comparisons are made to test which settings are significantly different from the High Fidelity setting. For any of the 7 pair-wise tests, the Friedman test was applied to all data available. The number in brackets behind the setting description indicates how many participants performed the task at the given setting. Table 3.5 lists the P-values for each comparison. To limit the probability of making a Type-I error to 0.05, only P-values smaller than 0.0073 are judged to be significant in each of the individual comparisons (see Methods section). The median force performance measure was not used on data from the dissection task since it did not show any overall differences for the individual surgeons in the previous test. Completion time could not be used for some of the settings since not enough data was available (while other performance measures have separate rows of data for left hand and right hand, this is not the case for completion time, thereby reducing the amount of data). In the clip application task some of the measures are not covered because no subject data was available due to Surgeon 3 only performing the dissection task.

Table 3.5 P values of pair wise comparisons between the High Fidelity setting and each of the other 7 settings. The data is from the dissection task. The number in brackets behind each setting indicates the number of surgeons the data is from. A bold face P-value indicates a significant difference (see text).

High Fidelity compared to:	P-value				
	50p Force	*95p Force*	*Comp. Time*	*Median Velocity*	*Path Length*
No Force Feedback *(3)*	N/A	**8E-6**	0.25	0.07	0.15
All Degradations (3)	N/A	0.21	0.52	0.48	0.72
Backlash (1)	N/A	1	*N/A*	*0.76*	0.12
Cogging (2)	N/A	0.28	*0.76*	*0.51*	1
Friction (1)	N/A	0.06	*N/A*	*0.53*	0.12
Inertia (1)	N/A	0.22	*N/A*	*0.53*	0.03
Saturation (1)	N/A	0.22	*N/A*	*0.76*	0.35

Table 3.6 P values of pair wise comparisons between the High Fidelity setting and each of the other 7 settings. The data is from the clip application task. The number in brackets behind each setting indicates the number of surgeons the data is from. A bold face P-value indicates a significant difference (see text).

High Fidelity compared to:	P-value				
	50 perc. force	*95 per. force*	*Comp. time*	*Median Velocity*	*Path Length*
No Force Feedback(2)	**0.0071**	**7E-6**	0.19	0.02	0.02
All Degradations (2)	0.48	0.35	0.71	0.72	0.60
Backlash (0)	N/A	N/A	N/A	N/A	N/A
Cogging (2)	0.40	0.79	0.32	0.90	0.11
Friction (1)	0.21	0.53	N/A	0.23	0.26
Inertia (1)	0.66	0.19	N/A	0.90	0.03
Saturation (0)	N/A	N/A	N/A	N/A	N/A

In the dissection task, the only significant difference that could be detected was in 95 percentile forces between High Fidelity and No Force Feedback. In the clip application task, a difference between these two settings is significant both in median forces as well as 95 percentile force

levels. Other performance measures did not detect a significant difference in pair wise comparisons.

3.4.2.2 Time-Dependent Effects

To make sure that the results are not biased due to significant time-dependent effects such as learning, fatigue, loss of interest, etc., the data was checked for a correlation between trial number and performance measure. The first trial the subject executes is designated Trial 1. Every following data segment that is treated separately during the analysis is given a successively higher trial number. During the clip application task, the subjects apply 3 clips on one 'duct'. Since each of these clips are treated separately in the analysis, they are given separate (and sequential) numbers. Using starting time of the trial, as opposed to trial number, gave similar results.

Correlation coefficients were computed for every surgeon, task, and hand (except for completion time). It is generally understood that when $|R| < 0.3$ the correlation is little or none, for $0.3 \le |R| < 0.7$ correlation is weak, while for $|R \ge 0.7|$ the correlation is strong. Only values with $|R| > 0.45$ are reported here. To interpret the correlation coefficient, its squared value R^2, called the coefficient of determination, expresses the proportion of variance in the dependent variable explained by the independent variable. Therefore a correlation coefficient of 0.7 between force and experiment number means that a little less than half of the variance in the force data can be explained with the experiment number. Since we are more used to thinking in standard deviations than variance we can also put is as follows: if R=0.7, then if we compute the standard deviation of force from the line that describes the correlation (see for example Figure 3-16), and we divide this standard deviation by the original (non-compensated) standard deviation of the force, the ratio will be 0.71. This means that the standard deviation of force is reduced by 29 percent after compensation for experiment number at the given value of R [Norman, 1994] (For R=0.5 and 0.9 the ratio is 0.87 and 0.43 respectively). Finally, the reported P-value indicates the probability of getting a correlation as large as the observed value by random chance, when the true correlation is zero.

The three strong correlations found are:

- 50 percentile force: Dissect, Surgeon 1, left hand: R=0,74 P=0.002.

- 95 percentile force: Dissect, Surgeon 1, left hand: R=0,70 P=0.003

- Path length: Dissect, Surgeon 2, left hand: R=-0.71, P=0.003

Figure 3-15-Figure 3-17 show the variables in these datasets plotted versus experiment number. A line indicates the trend. Markers indicate degradation setting.
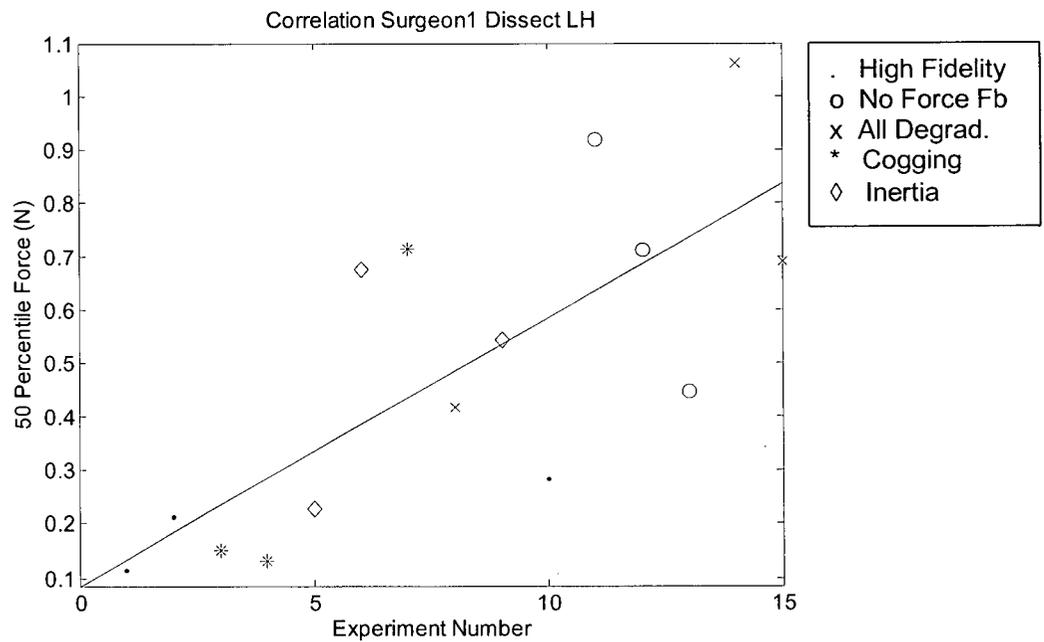
Figure 3-15: Mean forces versus experiment number: Surgeon 1, Left hand.
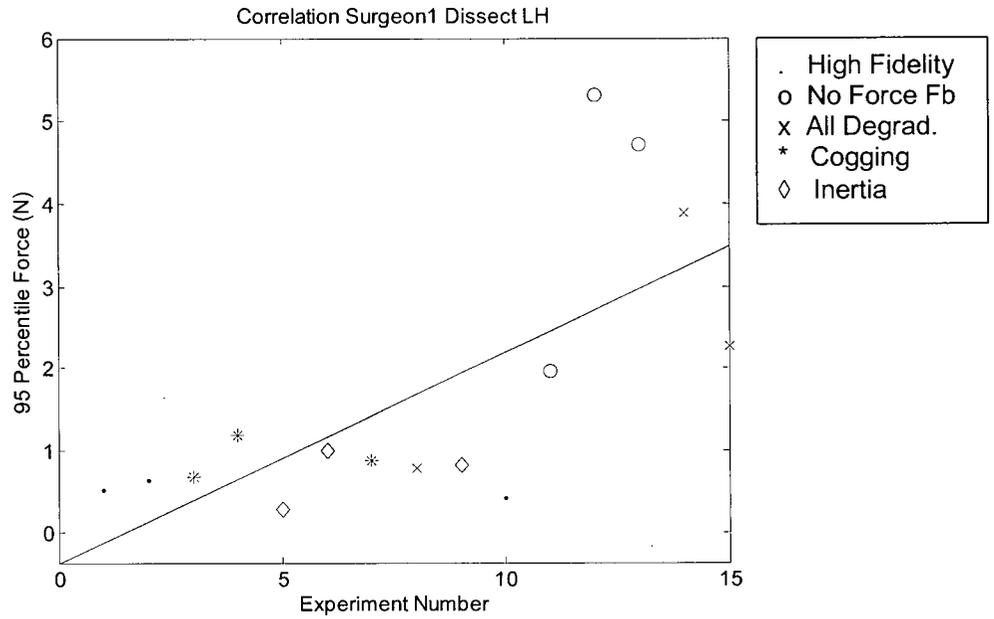
Figure 3-16   95 Percentile forces versus experiment number: Surgeon 1, Left hand.
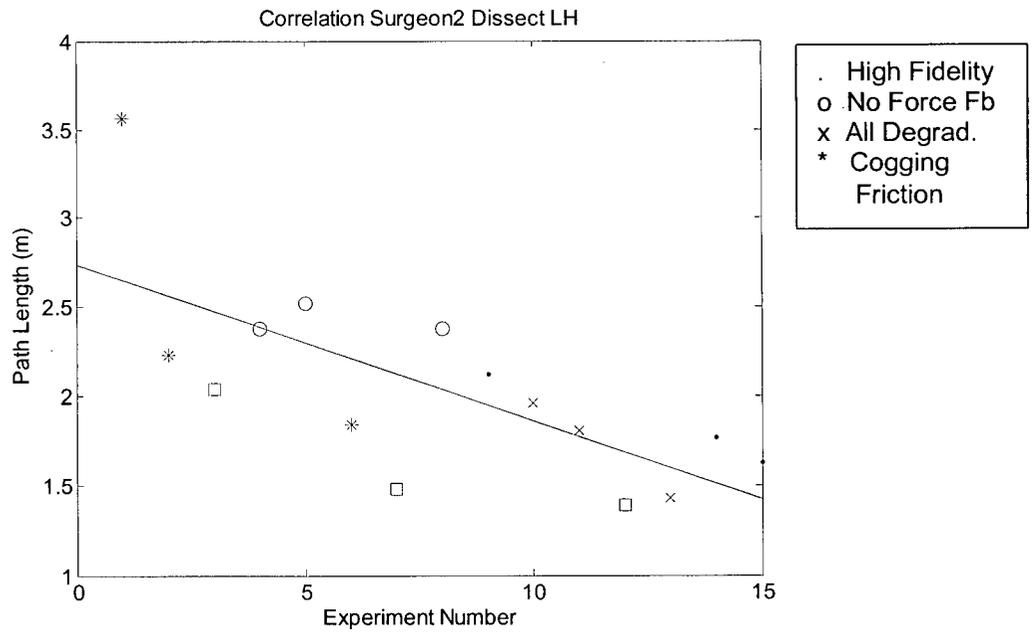
Figure 3-17   Path length versus experiment number: Surgeon 2, Left hand.

90

Weak correlations were found for the following cases:

Table 3.7: Weak correlations between performance measures and experiment number.

Perf. Measure	Task	Participant	R	P
Completion Time	Dissect	Surgeon 1	-0.51	0.05
Completion Time	Dissect	Surgeon 2	-0.53	0.04
Completion Time	Dissect	Surgeon 3	-0.63	0.01
Median Velocity	Dissect	Surgeon 3	0.61	0.02
Path Length	Dissect	Surgeon 2	-0.58	0.02
Path Length	Dissect	Surgeon 3	-0.53	0.04

3.5 Discussion

This section will discuss the implications of our results and specifics on task execution by the individual participants. This will be followed by comments on how the surgical environment, haptic hardware, and the strength of haptic degradations may have influenced our results.

3.5.1 Inter Setting Differences

Our first hypothesis was that the hardware settings caused no significant difference in the performance measures (H_0).

3.5.1.1 Force Measures

Based on the Friedman test, H_0 was rejected for all surgeons for both tasks for the 95 percentile force measure. This means that at least one of the hardware settings made a significant difference in exerted (peak) forces. Pair wise comparisons between the High Fidelity and the other 7 settings showed that only the No Force Feedback conditions caused significantly different 95 percentile forces across all surgeons and both tasks, while the No Force Feedback settings caused significantly different 50 percentile forces in clip application as well. An important result is that the force saturation at 1 N that is present in the All Degradations setting as well as the Force Saturation setting did not cause a significant difference in task performance.

The null hypothesis was rejected for Surgeon 1 in the clip application task using the 50 percentile force measure as well. Pair wise comparisons showed that the only hardware setting that resulted in significantly different median forces across all surgeons is the No Force Feedback setting.

The increase in exerted forces under no force feedback we found is in line with results found by Wagner in a simulated dissection task [Wagner, 2002], and by Moody in a VR suturing task [Moody, 2002].

3.5.1.2 Time, Median Velocity and Path Length Measures

In the dissection task, the null hypothesis had to be rejected based on the time, median velocity and path length measures for Surgeon 2 but not for the other two surgeons. Pair-wise comparisons over the data from all surgeons did not lead to a significant difference in these measures for any of the hardware settings compared to the High Fidelity setting. From Figure

3-8-Figure 3-10, it can be seen that Surgeon 2 stands out from the other two surgeons in this task by having high repeatability between repetitions in combination with larger differences between settings.

In the clip application task, data from Surgeon 1 showed significant differences between the different settings in median velocity. By inspection of Figure 3-9, the median velocities at the left hand High Fidelity setting stand apart (lower) from the other settings. Pair-wise comparisons over both surgeons between the High Fidelity setting and the 7 other settings did not reveal any significant difference.

There was a significant difference between settings in the data from Surgeon 2 in completion time and path length in the clip-setting task. Pair wise comparisons over both surgeons did not reveal significant differences between the High Fidelity setting and any of the other 7 settings. This is likely due to the fact that Surgeon 2 has a high repeatability with relatively large differences between settings compared to the other surgeon.

3.5.2 Other Observations

3.5.2.1 *Learning and Hurrying*

There are three possible explanations for the correlations found between repetition number and force or time:

1. Learning (completion time decreases, quality remains constant or goes up).
2. Hurrying (completion time decreases, quality as well).
3. None of the above, in the cases that show a correlation between trial number and performance measure, the particular order of the degradations influenced the outcome.

The only strong correlation between exerted forces and experiment number was found for Surgeon 1 in dissection. In this particular session, the last experiments were done under the No Force Feedback setting 3 times in a row followed by two repetitions under the All Degradation setting. While setting order was chosen at random, in this particular case the random choice was not well distributed. This could have influenced the outcome.

The 3 weak correlations between time and experiment number were negative. Since no reliable evidence of changes in forces over time were found, the most likely explanation is learning.

Correlations with time were found only in the dissection task. A possible explanation for why this did not appear in the clip application task is that it took all participants roughly 60 minutes to complete all repetitions of the dissection task, compared to only 30 minutes for the clip application task. Surgeon 1 first performed the clip application task, Surgeon 2 performed the dissection task first.

The strong negative correlation found between path length and experiment number may be due to learning since it is accompanied by a weak negative correlation in completion time: assuming no other differences between the task repetitions, a decrease in path length combined with an increase in completion time would suggest that more efficient movements are made but at a lower speed (An increase in path length with a decrease in completion time would suggest faster movements but more unnecessary movements as well)

3.5.2.2   *Inter-Hand Differences*

When data from different surgeons were grouped together to do hand-wise comparisons per task, the median values of all 50 percentile forces suggest that lower median forces are exerted by the left hand compared to the right hand in both tasks. This may reflect the functional differences between the left and right hand. In both tasks the left hand is used for holding the tissue in place and putting it under tension. The right hand is used to do more subtle manipulation and positioning. Also, the left hand applies a forces of longer duration compared to the right hand, affecting the median force applied.

In the median velocities, a clear distinction between the right hand and left hand reflects this difference in function as well. We will go into somewhat more detail on this below.

3.5.2.3 Inter-Surgeon Differences

The cumulative distribution plots of tool-tip forces in Figure 3-4 showed Surgeon 3 being remarkably less affected by the No Force Feedback and All Degradation settings than the other two participants in the dissection task. Surgeon 3 also logged the shortest median completion time. Surgeon 3's video gaming experience may explain this. Both laparoscopic literature [Grantcharov, 2003], and repeated unanticipated observations in our research group[9] suggest that gaming experience may influence task performance in virtual environments.

Figure 3-9 showed a smaller difference in velocity between right hand and left hand for Surgeon 3 compared to the other surgeons. This may show adaptation of that participant's technique to the virtual environment. In the simulator the user can dissect by grasping the tissue with the non-dominant hand, and dissect away all tissue in one stroke. While this is a highly unlikely, if not impossible, way to perform dissection in the OR, it is efficient on the simulator. If the non-dominant hand only grasps the tissue in one place, using the one-stroke technique in the dominant hand, its median velocity will be low. Therefore the technique used has a direct influence on the median velocities, and may be the cause of the differences between surgeons.

3.5.2.4 Variance between repetitions

There was a notable variance in all the performance measures between the different repetitions of the same task under the same setting. While there will always be some variation when a task is executed repeatedly; stance, posture, fatigue, and even day of the week may have an impact on task performance and repeatability.

3.5.3 Participants

This section will discuss how our results may be biased by the size and composition of our subject pool.

---

[9] Sensory Perception and Interaction Research Group, Computer Science Department, University of British Columbia. URL: http://www.cs.ubc.ca/labs/spin/
Related studies forthcoming

An implication of Grantchov's paper [Grantcharov, 2003] on hand dominance, gender, and video gaming experience on performance in laparoscopic surgery is that these factors should be taken into account when recruiting subjects. In our study, all subjects were male and right-handed which is representative of the current majority of surgeons. In Grantchov's study, male participants completed the tasks significantly faster while participants with right hand dominance performed fewer unnecessary movements. A larger group of participants would be needed to account for the variability between surgeons. In our experiment the surgeon with substantial computer gaming experience (Surgeon 3) seemed least affected by the different settings and showed high repeatability in general. Other factors that could influence individual performance are age, innate motor skills, the innate ability to internally represent models of spatial information, and the way people are trained. Because laparoscopy has only recently become widespread, standardization of curriculums is still underway. Surgeons with many years of experience, like the participants in our study, all had very different training experiences. Since we were able to test only up to 3 participants, our results may be skewed due to individual characteristics of the participants.

An unexpected result we found was the difference in completion time between the mechanical engineering students in a pilot test and the expert surgeons in the final study. We expected the expert to be able to do the tasks at least as fast as complete novices, while instead they took longer. From observation, it seems plausible that this was due to difference in techniques used. The surgeons performed the dissection task by making many small dissections bit by bit, in a generally more careful fashion. In contrast, the engineering students tended to 'sweep' the cautery tool over the tissue, dissecting large portions of the structure in one stroke: a way of dissection that is not possible on real tissue, but very efficient in the VE.

3.5.4 The VE Hardware and Software

This section will discuss how the choice of software and hardware may have influenced our results, and how future versions of VR simulators may give different results.

3.5.4.1 Level of Force Feedback

Surgeon 1 noted that he relied mainly on vision in the clip application task, due to the low stiffness of the tube that the clips are applied on. In his experience, manipulating the cystic duct in a real patient with standard instruments provides significantly more force feedback then in the task presented by the software. This was also reflected in comments by other surgeons. In fact when asked about the force feedback, none of the surgeons seemed to notice that force feedback was turned off completely at some of the repetitions. Two participants in the pilot study were so convinced that they did have force feedback in the No Force Feedback setting, that they insisted after completion of the experiment to do the task again in this setting. They were stunned to notice that there was not any force feedback present.

A plausible explanation for this is that the forces generated in the VE are too low. Both Kinnaird [Kinnaird, 2004] and Rosen [Richards, 2000] collected forces in the OR during laparoscopic cholecystectomy with standard surgical tools that are modified to measure forces. Force levels measured by both researchers in the OR are significantly higher than force levels in the Reachin VE. While we were not aware of this at the time we were conducting the experiments (Kinnaird obtained her results after the start of our experiments), our hardware degradation software could potentially be used to amplify the forces computed by the Reachin VE to higher levels.

As discussed in more detail in chapter 1, Stylopolous *et al.* suggest in [Stylopoulos, 2004] that laparoscopic surgeons derive how much force they are exerting on tissue also based on visual cues as tissue color, contour, and adjacent tissue integrity. While the more subtle visual force cues that a real environment gives like blanching of tissue under tension are not yet incorporated in VR surgical environments, changes in shape due to applied forces are obvious and could be the dominant factor in judging force feedback.

A third hypothesis to explain that subjects did not notice the absence of force feedback is that by repeating the same task over and over again the participants after a while strongly rely on motor memory to execute the tasks. Possibly haptic feedback was used to determine the properties of the environment during the exploring and learning phase, but after enough practice the participants were able to execute the task without paying too much attention to haptic cues. Such

a scenario is more likely to occur in the VE, where subsequent repetitions are very similar, than in the OR.

As a final remark on the force levels in the VE, the fact that our analysis shows that surgeons exerted significantly different forces on the tissue under the No Force Feedback setting compared to the High Fidelity setting, shows that in both tasks the settings did make a difference even though the participants may not have consciously noticed when force feedback was absent. It is important to notice that force saturation at 1 N did not result in a statistically significant difference.

3.5.4.2   *Other Aspects of the Tasks*

The dissection and clip application tasks in the experiment were not as natural to the surgeons as we expected. To minimize learning, the participants were instructed to perform the task as they would do during real surgery. However, the participants requested further instruction, for reasons such as the following:

- The structures that were subject to dissection or clip application were not similar enough to any corresponding structure in human anatomy to evoke an obvious approach of how to complete the task.

- In the dissection task, participants mentioned that they would have used blunt dissection instead of cautery because of the fragility of the underlying tissue. Blunt dissection by means of spreading the tissue by opening the forceps is not implemented on the simulator.

- When applying cautery in the VE, the covering tissue can be removed by a single sweep over the tissue, almost like applying paint on a surface with a brush. In the OR, it is not possible to burn away the tissue at such a fast rate. Instead many small movements need to be made while the cautery tool has to be used by pulling tissue away from the underlying structure instead of pushing it. This may have led to the difference between completion times of the surgeons and the graduate students: the graduate students may have dissected in the way that is most effective in the VE, while the surgeons were asked to dissect in a similar way to the OR, which required more time.

3.5.5 Strength of Haptic Degradations

The haptic degradations were chosen to be within the limits of what could be expected in a cost-effective hardware design. There were no complaints of any of the participants that the degradation settings were too strong. There may have even been room to make the effects more strong: Surgeon 1 said at one point that he found it hard to distinguish between the different settings.

Would the experiment have benefited from stronger degradations? This would be dependent on the results. It is plausible that any of the hardware degradations can affect task performance, if one simply increases the strength of the degradations enough. What we would like to know, is at which strength it starts to make an impact, something we would have tested for explicitly if we were not bound by the limited time available by our participants. Therefore if we would have executed the experiments with stronger degradations and had obtained the same results, we would know even more since our current settings would be a subset of these stronger settings (assuming that there is such a point in amplitude of hardware degradation above which surgical task performance is affected). On the other hand, if results of that experiment would show that the degradations did make a significantly difference, we would not have been able to identify levels of friction etc. that can be allowed in a design without affecting task performance.

Finally, if the force levels in the surgical environment were higher (which is probably more representative of real conditions), how would this have affected our results? It should be realized that the forces generated by the surgical environment are only present when a tool is in contact with tissue. The friction, cogging, backlash, and inertia forces from our hardware degradation software are present anywhere in the workspace when turned on. Since the hardware degradation forces can be seen as a nuisance factor, higher surgical environment forces would lead to a higher signal to noise ratio. Exactly how this would translate to task performance taking into account the difference between tissue contact and free movement, can only be determined by experiments.

3.5.6 Statistical Power

One last question that needs to be addressed is whether the Friedman test had enough power to draw any conclusions on the quality settings for which we did not find any significant difference. According to ([Zar, 1996], [Noether, 1959]): the power of the Friedman test is roughly $\frac{3k}{\pi(k+1)}$ time the power of the ANOVA (k is the number of fixed-effect factors). In our case, $k=5$, and thus our results have 80% of the power that we would have obtained from an ANOVA on the same data, were it appropriate. The power of the ANOVA test goes up with increasing distances in means between groups and decreasing variances of the groups. The variance of the 50 percentile and 95 percentile forces exerted in the No Force Feedback setting is large compared to the other settings, but nevertheless the Friedman test detected a significant difference between the No Force Feedback and the High Fidelity setting. Since in general power goes up with increasing differences in means and decreasing variance in the data, the smaller variance in data sets other than High Fidelity should lead to more power to detect an equal difference in means. This suggests that the Friedman test did have acceptable power to find differences if in fact there were any.

However, sufficient statistical power to detect differences for 1 to 3 participants does not mean that the results of this study are fully representative of the greater surgeon population.

Finally, bugs in the clip application tasks that were discussed in the results section led to a decrease in reliability of completion time as a performance measure. The extra noise introduced by these shortcomings has decreased the effectiveness of the tests to find inter-setting differences for this performance measure.

Chapter 4 Conclusions and recommendations

4.1 Main Findings & Thesis Contributions

4.1.1 Experimental Setup

We have modeled and implemented inertia, backlash, cogging, friction, and force saturation and shown that it is possible to implement degradation factors of haptic hardware under the following circumstances:

- Multiple models working simultaneously,
- in multiple degrees of freedom,
- some with non-linearities,
- on hardware with complex and unknown dynamic properties,
- at varying and low sample rates

While the friction, cogging, and force saturation are stable at practically any amplitude, stability issues limit the mass the inertia/backlash model can represent. When all factors are turned on at the same time, maximal parameters as shown in Table 4.1 can be simulated.

Table 4.1 Stable parameters of the hardware models

		Translation	*Rotation*
Inertia	*Mass*	$= 0.2\ kg$	$= 2gm^2$
	K	$= 600\ N/m$	$= 4\ Nm/rad$
	B	*0.3 critical damping*	*0.2 critical damping*
Backlash	*Gap Width*	$= 1\ mm$	$= 2^o$
Cogging	*Amplitude*	*0.6 N*	*0.04 Nm*
Friction	*Pre-sliding*	$= 100\ \mu m$	*=5 millirad*
	Stick Velocity	$= 5\ mm/s$	$= 0.1\ rad/s$

101

While the models can be further improved upon to either extend the range of model parameters (e.g. mass), or the fidelity of the effects, the current model parameters fit the range we need to test for and we feel that the fidelity is high enough for our purpose.

Occasional VE sampling rates as low as 200Hz made the creation of accurate models increasingly challenging. Especially our inertia/backlash model will benefit from either more computer power or VE environments that takes up less resources.

4.1.1.1 Contribution: Real-time Simulation of Inexpensive Component Substitution in a Haptic Interface

This is the first time a platform has been developed to simulate in real-time the result of introducing friction, cogging, force saturation, backlash and inertia in a haptic interface. This allows developers to test the effect of substituting expensive haptic hardware with inexpensive components by simply adding the custom software to the haptic loop, instead of having to go through an expensive iterative hardware-redesign process to obtain acceptable parameters. Because our software can be called from the device driver, a hardware developer can test the hardware settings with a variety of software applications without having to modify any of the applications themselves.

4.1.2 User Study

We tested task execution of expert surgeons in a dissection task and a clip application task under different haptic conditions. While our design was for 8 surgeons, a reasonably representative sample of the population, extreme difficulties in recruitment and hardware reliability constrained our observations to three individuals, with consequent impact on the type of analysis we could perform and its power. Three surgeons performed the dissection task and 2 surgeons performed the clip application task.

4.1.2.1 Did the Degradations in Haptic Quality Lead to Differences in Task Performance?

Our first goal was to test if degradations in quality of haptic feedback would lead to significant differences in task performance. We tested for 8 different settings:

- High Fidelity
- No Force Feedback
- Force Saturation
- Friction

- Inertia
- Inertia with Backlash
- Cogging
- Saturation+Cogging+Inertia+Backlash

The following performance measures were used:

- Median force
- 95 percentile force
- Completion time

- Median velocity
- Path length

A Friedman tests on data from individual surgeons showed that for each surgeon the 95 percentile force performance measure was significantly affected by the hardware settings. Other performance measures showed significant differences for individual cases only.

Seven pairwise comparisons between the High Fidelity settings and each of the other settings showed that over all participating surgeons, only No Force Feedback caused a significant chance in exerted forces (both median and at the 95 percentile level). In previous research, differences in exerted forces were found between force-feedback and no-force-feedback in a virtual dissection task and a tele-operated dissection task on artificial tissue. In this study no significant differences in completion time were found due to degraded hardware. It was found though that the completion time as a measure of performance in the clip application task had low reliability due to malfunctioning of the VE software. Also, none of the hardware degradations caused a significant difference in path length or median velocity compared to the High Fidelity setting.

4.1.2.2 Which Degradations Matter Most?

The second goal was to order conditions for which a significant effect was found according to impact. Since significant differences were found for only one single setting, no such ranking can be made with the current data. Experiments with stronger degradations are necessary to show such ranking.

4.1.2.3 Other observations

Some other practical observations were made that are important for future work. There was considerable variance in exerted forces between repetitions of the same task at the same setting. Variance between repetitions in a study using the same software and hardware was lower in [Kinnaird, 2004]. Kinnaird's study employed a different task with a longer duration; her experiment however was of much shorter duration, because she did not need to test for different hardware settings. She also recorded only 2 repetitions. These factors could all have contributed to the smaller variance in data.

Contrary to our expectations, graduate students were faster in executing the experiment than expert surgeons. In the VE, a dissection technique can be applied that is very fast but not realistic. While the expert surgeons were instructed to apply a dissection technique similar to the OR, graduate students who did not have OR experience dissected in the manner that was most efficient.

The Kolmogorov-Smirnov statistic in combination with bootstrapping was not able to detect any significant differences (introduction to the test and results are in Appendix I). This was partly due to large confidence intervals of the D-values, while non-overlapping 95% confidence intervals lead to a very conservative test: the chance of making a Type I error is very small, but it has low power. This method is expected to have higher power when the subjects form a better representation of the surgeon population.

In the dissection task all surgeons showed a weak negative correlation between completion times and trial number. For Surgeon 2 the path length of the left hand tool-trajectory showed a strong negative correlation with time. Both phenomena are most likely due to either learning or using a faster but unrealistic dissection technique that is possible in this particular virtual environment.

Force CDFs showed that surgeons performed differently, justifying our within-subjects design. Each surgeon maintained a consistent velocity profile within the same task, suggesting both consistency of performance by individuals, and an overall lack of influence of hardware

degradations on performance. The only exceptions to these are the left hand clip dissection velocities, and tool-tip velocity of Surgeon 1 under No Force Feedback in dissection.

4.1.2.4   Contribution: Degraded Haptic Quality Did Not Significantly Impact Performance

We found that as long as force feedback is turned on, the introduced hardware degradations did not make a significant impact on surgical task performance in a VR simulator. While our number of participants was limited, this suggests that haptic hardware for laparoscopic surgery does not need very expensive components and therefore could be made cheaper than is currently the case.

4.2 Recommendations for Future Work

4.2.1 Experimental Platform

4.2.1.1   Improvement and Extension of Models

Several aspects of the reported models are potentially improvable. Stiffening the virtual coupling will increase the fidelity of the inertia / backlash model; to do this stably requires application of more sophisticated filtering methods and a better understanding of the actual hardware dynamics. Also, a better velocity estimate for the friction model will make the transition from slip to stick state more reliable and possible at a lower velocity.

In this study it was necessary to limit the number of hardware degradations we tested, and consequently we implemented only those we felt were most important given our experimental goals. Models can be expanded and improved upon if there is no such burden. An extra model can be made to account for torque ripple. Cogging and torque ripple models can be adjusted to match specific motor designs, while the friction model can be adjusted to match specific friction characteristics of certain materials and contact situations.

4.2.1.2   Upgrading the 'Gold Standard' Software and Hardware

Shortly before the end of this project, Kinnaird published preliminary results of a validation study of the used VR environment. She measured tool-tip kinematics and exerted forces during a gall-bladder dissection task in the OR and in the VR environment by 2 expert surgeons doing 2

repetitions each. Her results suggested that the surgeons treated the VR differently from the OR. The most notable difference she found was the lower forces exerted in the VR environment. Therefore future versions of the VR environment may have stronger haptic cues. On the other hand, this thesis has presented arguments why surgeons may rely less on haptic cues in future simulation software that has improved graphics that show subtle force cues visually. When interpreting the results of this study, it should be taken into account that the software was the state of the art in laparoscopic simulation at the time of the experiment.

Occasional VE sampling rates as low as 200Hz made the creation of accurate models increasingly challenging. A faster computer that could provide update rates of at least 500 Hz, but preferably 1000 Hz, would allow stiffer models and therefore a more crisp backlash/inertia effect. We hope that this situation will improve with increased computational power, but such gains could be negated by corresponding increases in VE model complexity. Our results suggest that the value of maintaining a minimal update rate of 500-1000 Hz might outweigh the benefits of increasing model complexity.

As mentioned in the introduction, soft tissue environments are relatively easy to render haptically and are therefore less demanding of haptic hardware than stiffer environments. However, our case hardware degradation software would have benefited from better hardware. The friction model and backlash inertia model would have benefited from more accurate position measurement at high update rates. Both the low encoder resolution and noticeable backlash and flexibility in the instrument shaft when the instrument is extracted led to inaccurate position measurement in the current hardware. Also, as result of working with commercial hardware and software, we did not have access to details about software and hardware that would have helped us in increasing the stability of our own haptic simulation and we could not implement changes in for example device driver functionality that would have reduced computational overhead. Finally, development of the models and running the experiments was impaired by the significant down time of the hardware (shipping, repairs, and trouble shooting).

4.2.2 User Studies

4.2.2.1 Subject Pool

A user study with more participants could improve the significance of the results. Considering the difficulty of getting experienced surgeons to participate in the study, a small group of surgeons could be complemented with non-medical participants with significant gaming experience. The latter would be easier to recruit, have advanced skills in 3D VR environments, and if given proper instruction, may be able to duplicate the expert surgeons behaviour in basic skills tasks. Judging from our experience with engineering students and the software implementation of the surgical VE, proper instructions will be important to obtain similar task execution.

4.2.2.2 Experiment Design

In future studies that require repetitions of settings, it would be recommended to assign the settings in blocks of all k settings to the user, randomizing the order of the settings within the blocks. This would avoid having more than 2 subsequent tasks at the same setting, and would assure an even distribution of settings over the length of the experiment.

4.2.2.3 Next Step: Does Haptic Quality Impact Learning?

The current study investigates the immediate impact of haptic quality on an expert surgeon's task execution. A second important question is how haptic quality affects learning. If medical students would train in a VE with low quality haptic feedback, how would this affect their final performance? Results of this study suggest that subjects relied less on force feedback when repeating a task many times. This study would address the issue of force feedback reliance at different stages of training.

4.3 Hardware Quality in Future Haptic Interfaces

The market for haptic interfaces has been divided into two segments: cheap consumer haptic interfaces with low degrees of freedom, and expensive high fidelity complex interfaces for research and professional applications. This study has investigated if for one particular professional application, laparoscopic training, such a high-fidelity interface can be replaced by a less expensive one by using cheaper components. In a recent production launch a similar move has been made by a producer of low-volume high-end haptic interfaces that are used for 3D

design: interfaces that cost until recently $15,000 or more have been brought to market with lower torque output and relatively high friction for $1,950[10] (backdrive friction < 0.26N, cont. force >0.88N). Preliminary reports suggest that this device has discernibly lower performance (forces, workspace etc) than its higher-end counterparts, and yet we anticipate that much can be done with it. While more research is necessary, our own results suggest that it is possible to lower hardware costs of haptic interfaces for laparoscopy without affecting performance as long as a minimum amount of force feedback is still provided.

---

[10] Sensable Technologies PHANTOM® Omni™ haptic device (volume pricing below US$1000)

http://www.sensable.com/newsevents/pressreleases/pr-Devkit072104.asp

Appendix I: Data analysis with the Kolmogorov-Smirnov Test and Bootstrapping.

Introduction

The Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test (KS test) evaluates the hypothesis that two independent datasets are drawn from the same distribution. This test was developed by Smirnov [Smirnov, 1939] in 1939 and builds on a test previously developed by Kolmogorov for evaluating goodness-of-fit for a single sample [Kolmogorov, 1933]. Some of the interesting characteristics of the test are:

- The samples can be obtained from data from any kind of underlying distribution

- A difference measure obtained has no dimension.

The underlying idea of the KS test is that if the two datasets are drawn from the same population, one would expect their cumulative distribution functions to be similar. Therefore, if there is any significant difference at any point between the cumulative distribution functions of the two datasets, one can conclude that there is a high likelihood that the datasets were drawn from different populations. In addition, we can use the D-value (Figure 0-1) as a difference measure.
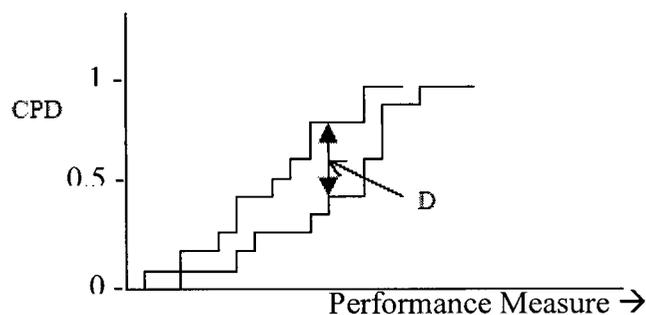
Figure 0-1   The cumulative distribution function for two finite data sets are represented as a staircase. The Kolmogorov-Smirnov statistic, D, is defined as the maximum vertical difference between two cumulative distribution functions.

The maximum vertical distance between the two distributions is indicated with the letter D (sometimes this value is referred to as M [Sheskin, 2000]). The null-hypothesis that the two

samples came from the same underlying distribution is rejected in a two-tailed test at significance level 5% when the D exceeds the value calculated in the following formula:

$$1.36\sqrt{\frac{n_1 + n_2}{n_1 n_2}} \qquad n_1, n_2 > 25 \qquad (1.4)$$

in which n_1 and n_2 are the number of independent samples in dataset 1 and dataset 2 respectively. Slightly modified versions of this formula are used for one-sided tests, and for testing at significance levels other then 5%, or with smaller numbers of samples.

Introduction to Bootstrapping

Resampling methods are used to calculate confidence bounds for statistical results when analytical methods are not available or are too complex. Effron introduced the bootstrap method in 1979 [Effron, 1979]. Bootstrapping is a method to estimate variability and bias in a statistic of a sample taken from a large population. While traditional parametric inference relies on assumptions about distribution of the population, the bootstrapping method assumes only that the samples' distribution is a good estimate of the actual population distribution; no assumptions of the shape of the distribution are made.

The Bootstrapping method works as follows. If γ is a parameter of the population P, and $\hat{\gamma}$ is the statistic that estimates this parameter based on a sample S drawn from P with $S = \{s_1, s_2, ..., s_n\}$, then in an ideal world, Monte Carlo sampling can be used to repeatedly draw a sample S and calculate the statistic $\hat{\gamma}$ for each S to create a distribution of $\hat{\gamma}$ to infer confidence bounds on the estimate. In the real world, only a limited number of samples can be taken. The bootstrapping method therefore applies Monte Carlo sampling to the original sample S to estimate confidence bounds on the statistic $\hat{\gamma}$ to make inferences about γ. S is resampled with replacement m times. Each S_i^* has n elements: therefore it is likely to contain some elements from S more than once, while sometimes not having any at all. The variability in S_i^* will lead to variation in $\hat{\gamma}$ as well. Therefore if the process is repeated many times, a population of $\hat{\gamma}$ is built from which a

110

confidence interval can be derived. To derive a confidence interval, m has to be a fairly large number, usually $m \geq 1000$.

The standard bootstrapping method assumes that the samples are independent. However in our study, force and kinematic data are temporally correlated: the position of the tooltip for example will be strongly correlated to its position in the immediately adjacent samples, especially at the higher update rates. Autoregressive analysis on multiple samples of data showed a high variance in the number of correlated samples. Therefore the moving block bootstrap is used which does not require any knowledge about the autoregressive model order [Lahiri, 2003].

Methods

We used the KS statistic in combination with bootstrapping to test the null hypothesis that there is no significant difference in D-values between repetitions of a task at the High-Fidelity setting and the other setting of interest. We replicate our experiment 1000 times by randomly resampling at all levels. We first randomly pick a surgeon. From the data of this surgeon we then randomly pick a repetition of the High Fidelity setting (HF1), a second repetition of the High Fidelity setting (HF2), and a repetition of the setting of interest (SI1). After resampling HF1, HF2, and SI1, the KS statistic is then applied between the HF1 and HF2 data sets to calculate a D_{intra} and between HF1 and SI1 to calculate D_{inter}. When repeated 1000 times, the range of D_{intra} values shows the variability within the High Fidelity setting, while the D_{inter} values show the range of D values between the High Fidelity and setting of interest. We reject the null hypothesis if 95 percent of the range of D_{intra} and D_{inter} are not overlapping.

Results

The plots on the following pages shows the KS statistic results for instrument-tissue interaction forces (Figure 0-2), and tool-tip velocity (Figure 0-3). Within each row, the black line represents the 95% range of the D-statistic within the High Fidelity repetitions (D_{intra}), and the red line represents the 95% range of the D-statistic between the High Fidelity setting and the degradation listed on the y-axis at this row (D_{inter}). One might expect the confidence intervals on D_{intra} to be the same within the clusters since they are comparisons within the High Fidelity setting. This is not the case because the D_{intra} and D_{inter} values are calculated by performing within-subject

comparisons: since not all surgeons participated in each setting, the D_{intra} values are extracted from different (combinations of) surgeons. For example in the dissection task, the D_{intra} values in the No Force Feedback and All Degradations settings reflect data from all 3 surgeons; in the Saturation setting the D_{intra} values represent data only from Surgeon 3. For settings in which the same surgeons participated, small variations in D_{intra} occur due to the bootstrapping process.

In both velocity and force KS plots, all red and black confidence interval pairs are overlapping, which means that no significant difference was found.

Force results show a clear distinction between the two tasks. In the dissection task the confidence intervals are small compared to the clip setting task. This is especially true for the right hand and confidence intervals on D_{intra}.

The confidence intervals on the velocity data are small compared to the confidence intervals on the force data.
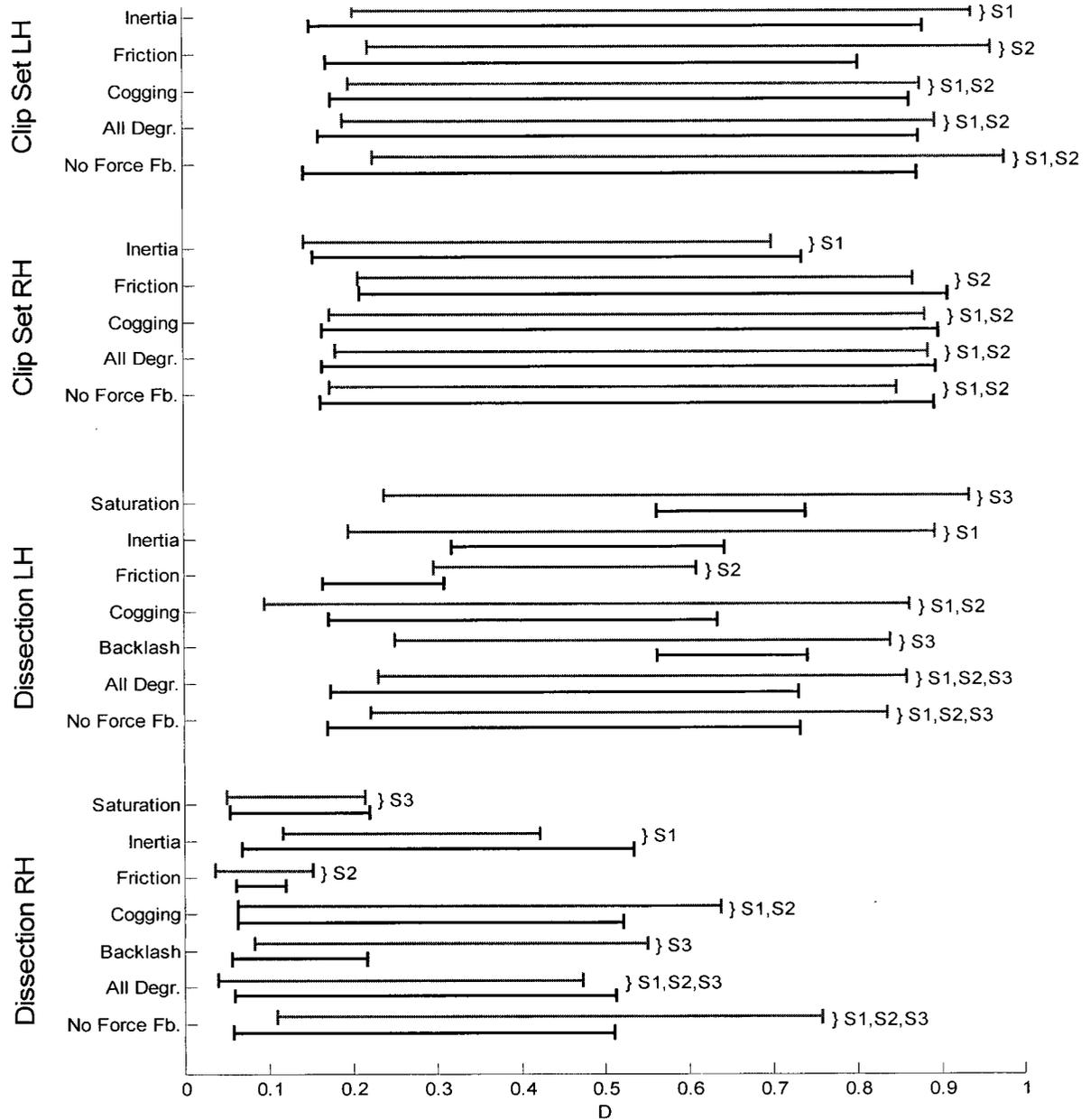
Figure 0-2 Results from the Kolmogorov-Smirnov analysis on force data. 95 percent confidence bounds for the D—value are given for intra-Hi Fidelity (black), and between High Fidelity and each setting (red). If the intervals indicated by the red and black line do not overlap, there is a high likelihood of a difference between the setting and the High Fidelity setting. S1, S2, and S3 indicate which surgeon (number) the data originates from.
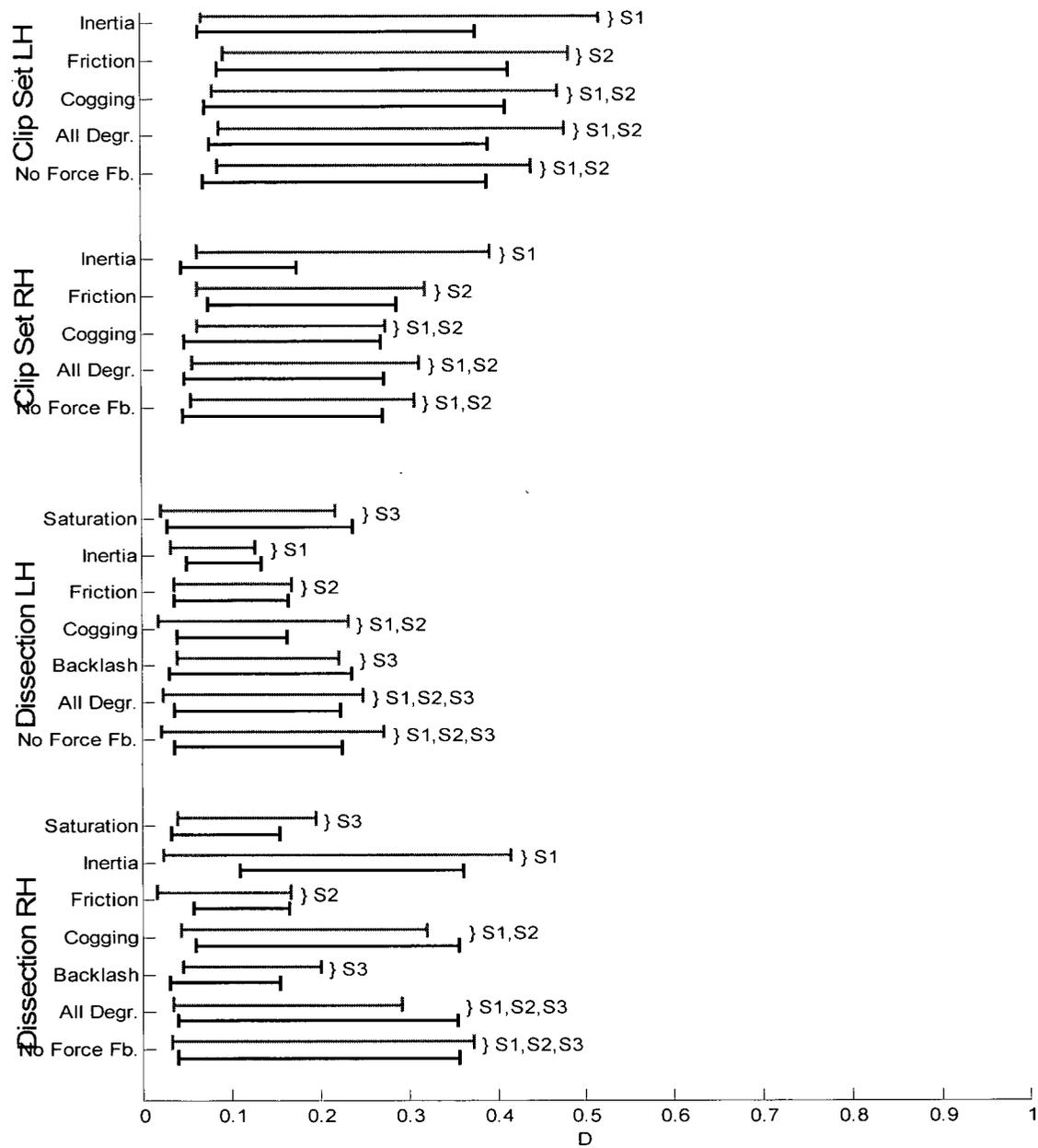
Figure 0-3: Results from the Kolmogorov-Smirnov analysis on velocity data. 95 percent confidence bounds for the D—value are given for intra-hifi (black), and between hifi and each setting (red). If the intervals indicated by the red and black line do not overlap, there is a high likelihood of a difference between the setting and the High Fidelity setting. S1,S2, and S3 indicate which surgeon (number) the data originates from.

114

Discussion

We found no significant differences between the High Fidelity setting and any of the other settings by applying the KS-test. Confidence bounds on the D-statistic were much larger than in a study performed by Kinnaird [Kinnaird, 2004] who found smaller differences between repetitions of a simulated dissection task. In contrast to this study, Kinnaird's study has:

- Different task: a more complicated gall-bladder dissection task
- Longer duration of the task
- Only 2 repetitions vs. 3 repetitions in this study
- Only high fidelity setting
- Different participants: same surgical experience level, different individuals.

These differences will be discussed in this section to shed some light on how they may have affected the outcome.

One observation that can be made from Figure 0-2 (which shows the confidence intervals on the D-value of the Kolmogorov-Smirnov statistic applied to force data) is that the confidence intervals in the clip application task are almost always larger than the confidence intervals in the dissection task. Average completion times for these tasks are roughly 1 minute for clip application, but 2 minutes for dissection. The dissection task in Kinnaird's requires dissecting a whole area instead of one line and took on average 5 minutes to complete. This suggests that there may be an increase of reliability of the D-value of force measurements with increasing task duration. The number of repetitions decreased with the increase of completion time; the clip application task has 9 repetitions, dissection 3, and Kinnaird's 5 minute dissection task was repeated twice. With an increase of repetitions it is more likely to obtain a rarely occurring value. By taking the 95 percentile of D-values, these outliers may not be filtered out since every point was in a way multiplied in it's nearest neighbors by bootstrapping resampled data.

The total experiment completion time was below 20 minutes in Kinnaird's study, while 1.5 to 3 hours in this study. Our longer experimental time may have also contributed to the larger variation between repetitions. The fact that the same surgeons work for at least 90 minutes in a

row in the OR during the real procedure makes fatigue a less likely explanation though it is hard to tell if monotony of the tasks and possible resulting boredom did contribute to the results.

Finally, the uneven number of populations that we draw from and the unequal variances in populations contribute to the wide range of 95% D-value range. We used two or three surgeons to represent the larger surgeon population, and 3 to 9 repetitions to represent the variability due to doing the experiment over with the same surgeon. We resampled the roughly 10,000 data points in a single repetition to represent doing the experiment again with the same surgeon under exactly the same circumstances that occurred when the dataset was originally recorded. As can be expected, variations due to resampling of the dataset are very small compared to differences between surgeons or repetitions. The range of the D-values was determined mostly by the differences between a small number of surgeons and a small number of repetitions. While the 95 percentile range is used to filter out outliers, outliers are multiplied by resampling.

Bibliography

Adams, R. J., Klowden, D. and Hannaford, B. (2001a), "Virtual Training for a Manual Assembly Task", Haptics-e, The Electronic Journal of Haptics Research (www.haptics-e.org), 2 (2), pp.

Adams, R. J., Klowden, J. and Hannaford, B. (2001b). Virtual training for a manual assembly task. 2, Nr. 2.

Adrales, G. L., Chu, U. B., Witzke, D. B., Donnelly, M. B., Hoskins, D., Mastrangelo, M. J., Jr., Gandsas, A. and Park, A. E. (2003), "Evaluating minimally invasive surgery training using low-cost mechanical simulations", Surgical Endoscopy, 17 (4), pp. 580-5.

Armstrong-Helouvry, B. (1991), "Control of Machines with Friction", pp.

Armstrong-Helouvry, B., Dupont, P. and Canudas De Wit, C. (1994), "A survey of models, analysis tools and compensation methods for the control of machines with friction." Automatica, 30 (7), pp. 1083-1138.

Armstronghelouvry, B., Dupont, P. and Dewit, C. C. (1994), "A Survey of Models, Analysis Tools and Compensation Methods for the Control of Machines with Friction", Automatica, 30 (7), pp. 1083-1138.

Baldwin, P. J., Paisley, A. M. and Brown, S. P. (1999), "Consultant surgeons' opinion of the skills required of basic surgical trainees", Br J Surg, 86 (8), pp. 1078-82.

Basdogan, C., De, S., Kim, J., Muniyandi, M., Kim, H. and Srinivasan, M. A. (2004), "Haptics in minimally invasive surgical simulation and training", IEEE Computer Graphics and Applications, 24 (2), pp. 56-64.

Batter, J. J. B., F. P. Jr. (1971). "GROPE-I: A computer display to the sense of feel." *International Federation of Information Processing*, Ljubljana, Yugosolavia,.

Baumann, R. and Clavel, R. (2001). "Haptic interface for virtual reality based minimally invasive surgery simulation". *IEEE International. Conference on Robotics.and Automation*

Beamish, T., Maclean, K. and Fels, S. (2004). "Manipulating Music: Multimodal Interaction for DJs". *ACM SIGCHI*, Vienna, Austria.

Ben Ur, E., "Development of a Force-Feedback Laparoscopic Surgery Simulator", Master's Thesis, Massachusetts Institute of Technology, 1999

Benarous, M. and Eastham, J. F. (1999). "The effect of the distribution of the magnetisation in brushless DC machines on cogging torques". *Ninth International Conference on Electrical Machines and Drives, 1999*

Berci, G., Margolin, D. A. and Marks, J. M. (1998). "Minimally Invasive Surgery: Principles and Outcome", Harwood Academic Publishers.

Berkley, J., Weghorst, S., Gladstone, H., Raugi, G., Berg, D. and Ganter, M. (1999), "Fast finite element modeling for surgical simulation", Stud Health Technol Inform, 62, pp. 55-61.

Betrisey, S. and Vollenweider, M., "Device for simulating a rod-shaped surgical instrument having a force feedback effect", Patent WO03023736, Xitact SA,

Bielser, D., "A framework for open surgery simulation", PhD thesis, Swiss Ferderal Institute of Technology Zurich, Switzerland, 2003

Bin, A. and Stiles, W. P., "Haptic feedback joystick", Patent US6429849, Microsoft Corp., USA,

Breedveld, P., Stassen, H. G., Meijer, D. W. and Jakimowicz, J. J. (1999), "Manipulation in laparoscopic surgery: overview of impeding effects and supporting aids", J. Laparoendosc. Adv. Surg. Tech. A, 9 (6), pp. 469-480.

Breedveld, P., Stassen, H. G., Meijer, D. W. and Jakimowicz, J. J. (2000), "Observation in laparoscopic surgery: overview of impeding effects and supporting aids", J. Laparoendosc. Adv. Surg. Tech. A, 10 (5), pp. 231-241.

Bridges, M. and Diamond, D. L. (1999), "The financial impact of teaching surgical residents in the operating room", Am J Surg, 177 (1), pp. 28-32.

Brouwer, I., Maclean, K. E. and Hodgson, A. J. (2004). "Simulating cheap hardware: a platform for evaluating cost-performance trade-offs in haptic hardware design". *Robotics and Automation, 2004. Proceedings. ICRA '04. 2004 IEEE International Conference on*

Brouwer, I., Ustin, J., Bentley, L., Sherman, A., Dhruv, N. and Tendick, F. (2001). "Measuring in vivo animal soft tissue properties for haptic modeling in surgical simulation". *Stud.Health Technol.Inform. (Medicine Meets Virtual Reality 2001)*

Chan, C. S., Tong, Z., Dang, A. and Qian, J. (2003). "Virtual Reality Modeling of Traditional Chinese Architecture". *Virtual Systems and MultiMedia*

Chen, J., Dimattia, C., Taylor, R. M. I., Falvo, M., Thiansathaporn, P. and Superfine, R. (1997). "Sticking to the point: A friction and adhesion model for simulated surfaces". *Proceedings of the Sixth Annual Symposium on Haptic Interfaces for Virtual Environment and Teleoperator System*, Dallas, TX, USA.

Chung, J. Y. and Sackier, J. M. (1998), "A method of objectively evaluating improvements in laparoscopic skills", Surgical Endoscopy, 12 (9), pp. 1111-6.

Coleman, R. L. and Muller, C. Y. (2002), "Effects of a laboratory-based skills curriculum on laparoscopic proficiency: a randomized trial", Am J Obstet Gynecol, 186 (4), pp. 836-42.

Colgate, J. E. and Schenkel, G. (1994). "Passivity of a Class of Sampled-Data Systems: Appliation to Haptic Interfaces". *American Control Conference, 1994*

Colgate, J. E., Stanley, M. C. and Brown, J. M. (1995). "Issues in the haptic display of tool use". *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. 'Human Robot Interaction and Cooperative Robots'*, Pittsburgh, PA, USA.

De, S. and Srinivasan, M. A. (1999), "Thin walled models for haptic and graphical rendering of soft tissues in surgical simulations", Stud Health Technol Inform, 62, pp. 94-9.

Den Boer, K. T., De Wit, L. T., Davids, P. H., Dankelman, J. and Gouma, D. J. (2001), "Analysis of the quality and efficiency in learning laparoscopic skills", Surg Endosc, 15 (5), pp. 497-503.

Deodhar, R. P., Staton, D. A., Jahns, T. M. and Miller, T. J. E. (1996), "Prediction of cogging torque using the flux-MMF diagram technique", IEEE Transactions on Industry Applications, 32 (3), pp. 569-576.

Derossis, A. M., Fried, G. M., Abrahamowicz, M., Sigman, H. H., Barkun, J. S. and Meakins, J. L. (1998), "Development of a model for training and evaluation of laparoscopic skills", 175 (6), pp. 482-487.

Dunnett, C. W. (1964), "New Tables for Multiple Comparisons with a Control", Biometrics, 20, pp. 482-491.

Effron, B. (1979), "Bootsrap methods: Another look at the jackknife." The Annals of Statistics, 7, pp. 1-26.

Ernst, M. O. and Banks, M. S. (2002), "Humans integrate visual and haptic information in a statistically optimal fashion", Nature, 415 (6870), pp. 429-433.

Eubanks, T. R., Clements, R. H., Pohl, D., Williams, N., Schaad, D. C., Horgan, S. and Pellegrini, C. (1999), "An objective scoring system for laparoscopic cholecystectomy", J. Am. Coll. Surg., 189 (6), pp. 566-574.

Fried, G. M., Derossis, A. M., Bothwell, J. and Sigman, H. H. (1999), "Comparison of laparoscopic performance in vivo with performance measured in a laparoscopic simulator", 13 (11), pp. 1077-1081.

Friedman, M. (1937), "The use of ranks to avoid the assumption of normality implicit in the analysis of variance." Journal of the American Statistical Association, 32, pp. 675-701.

Gerdes, J. C. and Kumar, V. (1995). "An impact model of mechanical backlash for control system analysis". *Proceedings of the American Control Conference*

Gillespie, B. and Cutkosky, M. (1996). "Stable user-specific haptic rendering of the virtual wall". *Proceedings of the ASME Dynamic Systems and Control Division*

Gosline, A., "bla", EECS University of British Columbia, 2003a

Gosline, A. H. (2001). Evaluation of Friction Models with a Haptic Interface. SPSC515 Course Report, University of British Columbia.

Gosline, A. H., "Simulation of Linear Elastic Media with Fluid Inclusions", M.A.Sc. thesis, Department of Electrical and Computer Engineering UNIVERSITY OF BRITISH COLUMBIA, 2003b

Gosline, A. H., Salcudean, S. E. and Yan, J. (2004). "Haptic simulation of linear elastic media with fluid pockets". *Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2004. HAPTICS '04. Proceedings. 12th International Symposium on*

Grantcharov, T. P., Bardram, L., Funch-Jensen, P. and Rosenberg, J. (2003), "Impact of hand dominance, gender, and experience with computer games on performance in virtual reality laparoscopy", Surg Endosc, pp. Online First.

Grantcharov, T. P., Rosenberg, J., Pahle, E. and Funch-Jensen, P. (2001), "Virtual reality computer simulation", Surgical Endoscopy, 15 (3), pp. 242-244.

Grewal, M., S. and Andrews, A. P. (2001). "Kalman Filtering Theory and Practice Using MATLAB", John Wiley & Sons, Inc.

Hacksel, P. and Salcudean, S. E. (1994). "Estimation of Environment Forces and Rigid-Body Velocities Using Observers". *Proceedings of the IEEE International Conference on Robotics and Automation*, San Diego, USA.

Haessig, D. A. and Friedland, B. (1991), "On the Modeling and Simulation of Friction", Journal of Dynamic Systems Measurement and Control-Transactions of the Asme, 113 (3), pp. 354-362.

Hanselman, D. C. (1994). "Brushless permanent-magnet motor design", McGraw-Hill, New York.

Hayward, V. and Armstrong, B. (2000). A new computational model of friction applied to haptic rendering. Experimental Robotics. P. Corke and J. Trevelyan. VI: 403-412.

Heijnsdijk, E. A., Pasdeloup, A., Van Der Pijl, A. J., Dankelman, J. and Gouma, D. J. (2004), "The influence of force feedback and visual feedback in grasping tissue laparoscopically", Surg Endosc, 18 (6), pp. 980-5.

Heilig, M. L., "Stereoscopic-television apparatus for individual use", Patent U.S.,

Heilig, M. L., "Sensorama simulator", Patent US3050870; G09B5/06; G02B27/22S,

Immersion Corporation (2003). Press Release: Immersion Launches Gaming Software That Immediately Enables Force Feedback for the Vast Majority of PC Games; TouchWare Gaming Turns Sound Waves into Touch Sensations. San Jose, CA.

Ishii, M. and Sato, M. (1993). "A 3D interface device with force feedback: a virtual work space for pick-and-place tasks". *IEEE Virtual Reality Annual International Symposium, 1993*

Ishikawa, T. and Slemon, G. R. (1993), "A method of reducing ripple torque in permanent magnet motors without skewing", IEEE Transactions on Magnetics, 29 (2), pp. 2028-2031.

James, D. L. P., D. K. (1999a). "ArtDefo, Accurate Real Time Deformable Objects". *Proceedings of Computer Graphics ACM SIGGRAPH 99 Conference*

James, D. L. P., D. K. (2002). "DyRT: Dynamic Response Textures for Real Time Deformation Simulation with Graphics Hardware". *Proceedings of ACM SIGGRAPH 2002, Computer Animation*, San Antonio, Texas.

James, D. L. P., D.K. (1999b). "ArtDefo, Accurate Real Time Deformable Objects". *Computer Graphics: ACM SIGGRAPH 99 Conference Proceedings*

Jensen, R. F. J. S. H. W. (2001). "Visual Simulation of Smoke". *Proceedings of SIGGRAPH*

Johnson, C. T. and Lorenz, R. D. (1992), "Experimental identification of friction and its compensation in precise, position controlled mechanisms", IEEE Transations on Industry Applications, 28 (6), pp.

Jukic, T. and Peric, N. (2001). "Model based backlash compensation". *American Control Conference*, Arlington, VA, USA.

Kapuscinski, C. L., "Motor Selection and Damper Design for a Six Degree of Freedom Haptic Display", M.A.Sc. thesis, Department of Mechanical Engineering Northwestern University, 1997

The 'Karlsruhe Endoscopic Surgery Trainer' http://www-kismet.iai.fzk.de/TRAINER/mic_trainer1.html.

Karnopp, D. (1985), "Computer simulation of stick-slip friction in mechanical dynamic systems", ASME Journal of Dynamic Systems and Control, 107, pp. 100-103.

Kawamura, S. and Ito, K. (1993). "A new type of master robot for teleoperation using a radial wire drive system". *Proceedings of the 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems '93, IROS '93*

Kilchenman, M. and Goldfarb, M. (2001). "Force saturation, system bandwidth, information transfer, and surface quality in haptic interfaces". *IEEE International Conference on Robotics and Automation*

Kim, J., De, S. and Srinivasan, M. A. (2002a). "Computationally Efficient Techniques for Real Time Surgical Simulation with Force Feedback". *10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, Los Alamitos, CA.

Kim, S., Hasegawa, S., Koike, Y. and Sato, M. (2002b). "Tension based 7-DOF force feedback device: SPIDAR-G". *Virtual Reality, 2002. Proceedings. IEEE*

Kinnaird, C., "A Multifaceted Quantitative Validity Assessment of Laparoscopic Surgical Simulators", M.A.Sc. thesis, Mechanical Engineering University of British Columbia, 2004

Kolmogorov, A. N. (1933), "suula determinazione emperica di una legge di distribuzione", Giorn dell'Inst. Ital. degli. Att., 4, pp. 89-91.

Krijn, M., Emmelkamp, P. M., Olafsson, R. P. and Biemond, R. (2004), "Virtual reality exposure therapy of anxiety disorders: A review", Clin Psychol Rev, 24 (3), pp. 259-81.

Kwon, D.-S. and Woo, K. Y. (2000). "Control of the haptic interface with friction compensation and its performance evaluation". *Intelligent Robots and Systems, 2000. (IROS 2000). Proceedings. 2000 IEEE/RSJ International Conference on*

L. Villegas, B. E. S., M. P. Callery, D. B. Jones (2003), "L. Villegas1, B. E. Schneider1, M. P. Callery1 and D. B. Jones", Surgical Endoscopy, Online First, pp.

Lahiri, S. N. (2003). "Resampling Methods for Dependent Data", Springer-Verlag.

Lajoie-Mazenc, M., Nogarede, B. and Fagundes, J. C. (1989). "Analysis of torque ripple in electronically commutated permanent magnet machines and minimization methods". *Fourth International Conference on Electrical Machines and Drives*

Larsen, O. V. H., J.;Østergaard,L.R.;Hansen,K.V.;Nielsen,H (2001). "The Virtual Brain Project : development of a neurosurgical simulator." *Medicine Meets Virtual Reality*, Newport Beach, California, USA.

Lawrence, D. A., Pao, L. Y., White, A. C. and Xu, W. (2004). "Low cost actuator and sensor for high-fidelity Haptic interfaces". *Proceedings 12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*

Lirici, M. M. (1997), "Editorial - New techniques, new technologies and educational implications." Minimally Invasive Theapies & Allied Technologies, 6, pp. 102-104.

Maclean, K. E., "Emulation of Haptic Feedback for Manual Interfaces", Ph.D. thesis, Mechanical Engineering MIT, 1996

Mahvash, M. and Hayward, V. (2004), "High-fidelity haptic synthesis of contact with deformable bodies", Computer Graphics and Applications, IEEE, 24 (2), pp. 48-55.

Massie, T. H. and Salisbury, J. K. (1994). "The PHANToM haptic interface: a device for probing virtual objects". *the Third Ann. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, ASME/IMECE*

Maxwell, S. E. D., H.D. (2003). "Designing Experiments and Analyzing Data. A model comparison perspective."

Mayrose, J., Chugh, K. and Kesavadas, T. (2000), "Material property determination of sub-surface objects in a viscoelastic environment", Biomed Sci Instrum, 36, pp. 313-7.

Moody, L., Baber, C. and Arvanitis, T. N. (2002). "Objective surgical performance evaluation based on haptic feedback". *Medicine Meets Virtual Reality Conference*, Newport Beach, CA, USA.

Moore, D. F., Schena, B. M. and Shahoian, E. J., "An input device with force feedback", Patent GB2350170, Immersion Corp.,

Moy, G., Wagner, C. and Fearing, R. S. (2000). "A compliant tactile display for teletaction". *Proceedings of the IEEE International Conference on Robotics and Automation*

Nahvi, A., Hollerbach, J. M., Freier, R. and Nelson, D. D. (1998). "Display of friction in virtual environments based on human finger pad characteristics". *American Society of Mechanical Engineers, Dynamic Systems and Control Division*, Anaheim, CA, USA.

Noether, P. V. E. G. E. (1959), "The asymptotic efficiency of the X^2 -test for a balanced incomplete block design." Biometrika, 46, pp. 475-477.

Norman, G. R. and Streiner, D. L. (1994). "Biostatistics, the bare essentials", Mosby Year Book Inc.

O'malley, M. and Goldfarb, M. (2002). "Comparison of Human Haptic Performance in Real and Simulated Environments." Proceedings of the IEEE 10th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems.

Payandeh, S. and Li, T. (2003), "Toward new designs of haptic devices for minimally invasive surgery", International Congress Series, 1256, pp. 775-781.

Reznick, R., Regehr, G., Macrae, H., Martin, J. and Mcculloch, W. (1997), "Testing technical skill via an innovative "bench station" examination", Am.J.Surg., 173 (3), pp. 226-230.

Richard, C., Cutkosky, M. and Maclean, K. (1999). "Friction Identification for Haptic Display". *the 8th Ann. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, ASME/IMECE*, Nashville, TN.

Richard, C. and Cutkosky, M. R. (2002). "Friction modeling and display in haptic applications involving user performance". *Proceedings of the IEEE International Conference on Robotics and Automation*

Richards, C., Rosen, J., Hannaford, B., Pellegrini, C. and Sinanan, M. (2000), "Skills evaluation in minimally invasive surgery using force/torque signatures", Surgical Endoscopy, 14 (9), pp. 791-798.

Rosen, J., Hannaford, B., Richards, C. G. and Sinanan, M. N. (2001), "Markov modeling of minimally invasive surgery based on tool/tissue interaction and force/torque signatures for evaluating surgical skills", IEEE Trans.Biomed.Eng, 48 (5), pp. 579-591.

Rosenberg, L. B., "Force feedback device with microprocessor receiving low level commands", Patent US2002033841, Immersion Corp., USA,

Rosenberg, L. B., Moore, D. F., Bruneau, R. D., Schena, B. M. and Martin, K. M., "Mechanical and force transmission for force feedback devices", Patent US6400352, Immersion Corp., USA,

Rosser, J. C., Rosser, L. E. and Savalgi, R. S. (1997), "Skill acquisition and assessment for laparoscopic surgery", Arch Surg, 132 (2), pp. 200-4.

Salisbury, K., Conti, F. and Barbagli, F. (2004), "Haptic rendering: introductory concepts", IEEE Computer Graphics and Applications, 24 (2), pp. 24-32.

Satava, R. (2003). "Keynote talk". *Medicine Meets Virtual Reality 2003*, Newport Beach, CA, USA.

Schijven, M. J., J (2003), "Virtual Reality Surgical Laparoscopic Simulators", Surgical Endoscopy, Online First, pp.

Scott-Conner, C. E. H. (1999). "The SAGES Manual: Fundamentals of Laparoscopy and GI Endoscopy." Springer, New York.

Sensable Technologies (2004). http://www.sensable.com/products/3ddesign/index.asp.

Shaver, M. J. http://www.cs.ubc.ca/~mshaver/ToolProject/ToolProject.html.

Sheskin, D. J. (2000). "Handbook of parametric and nonparameteric statistical procedures." Chapman & Hall /CRC.

Siegel, M. (2002). "Tactile display development: the driving-force for tactile sensor development". *IEEE International Workshop on Haptic Virtual Environments and Their Applications*

Smirnov, N. V. (1939), "Estimate of deviation between empirical distribution functions in two independent samples. (Russian)", Bull Moscow Univ, 2, pp.

Soper, N. J. (2004). Surgical Education, Talk at the Centre of Excellence for Surgical Education and Innovation. Vancouver, BC, Canada.

Sozer, Y. and Torrey, D. A. (1998). "Adaptive torque ripple control of permanent magnet brushless DC motors". *Proceedings of the Thirteenth Annual Applied Power Electronics Conference and Exposition*

Srinivasan, M. A. and Basdogan, C. (1997), "Haptics in virtual environments: Taxonomy, research status, and challenges", Computers & Graphics, 21 (4), pp. 393-404.

Srinivasan, M. A., Beauregard, G. L. and Brock, D. (1996). The Impact of Visual Information on the Haptic Perception of Stiffness in Virtual Environments. Proc. of the 5th Ann. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems. Atlanta, GA, IMECE.

Studer, C., Keyhani, A., Sebastian, T. and Murthy, S. K. (1997). "Study of cogging torque in permanent magnet machines". *IEEE Industry Applications Conference, 1997. Thirty-Second IAS Annual Meeting, IAS '97.,*

Stylopoulos, N., Cotin, S., Maithel, S. K., Ottensmeye, M., Jackson, P. G., Bardsley, R. S., Neumann, P. F., Rattner, D. W. and Dawson, S. L. (2004), "Computer-enhanced laparoscopic training system (CELTS): bridging the gap", Surg Endosc, 18 (5), pp. 782-9.

Sutherland, I. (1963). "Sketchpad: A man-machine graphical communication system". *IFIPS Proceedings of the Spring Joint Computer Conference*, Detroit, Michigan.

Tan, H. Z., Srinivasan, M. A., Eberman, B. and Cheng, B. (1994). "Human Factors for the Design of Force-Reflecting Haptic Interfaces". *the 3rd Ann. Symp. on Haptic Interfaces for Virtual Environment and Teleoperator Systems, ASME/IMECE*, Chicago, IL.

Taylor, R. H., Funda, J., Eldridge, B., Gomory, S., Gruben, K., Larose, D., Talamini, M., Kavoussi, L. and Anderson, J. (1995), "A telerobotic assistant for laparoscopic surgery", Engineering in Medicine and Biology Magazine, IEEE, 14 (3), pp. 279-288.

Tendick, F., Downes, M., Goktekin, T., Cavusoglu, M. C., Feygin, D., Wu, X. L., Eyal, R., Hegarty, M. and Way, L. W. (2000), "A virtual environment testbed for training laparoscopic surgical skills", Presence-Teleoperators and Virtual Environments, 9 (3), pp. 236-255.

Traverso, L. W., Koo, K. P., Hargrave, K., Unger, S. W., Roush, T. S., Swanstrom, L. L., Woods, M. S., Donohue, J. H., Deziel, D. J., Simon, I. B., Froines, E., Hunter, J. and Soper, N. J. (1997), "Standardizing laparoscopic procedure time and determining the effect of patient age/gender and presence or absence of surgical residents during operation. A prospective multicenter trial", Surg Endosc, 11 (3), pp. 226-9.

Vassallo, S. P., Bailey, D. W. and Gregorio, P., "Interface Apparatus with Cable-Driven Force Feedback and Four Grounded Actuators", Patent EP1417547, IMMERSION CORP,

Villegas, L., Schneider, B. E., Callery, M. P. and Jones, D. B. (2003), "Laparoscopic skills training", Surgical Endoscopy, 17 (12), pp. 1879 - 1888.

Vlachos, K., Papadopoulos, E. and Mitropoulos, D. N. (2004). "Mass/ inertia and joint friction minimization for a low-force five-dof haptic device". *Proceedings of the IEEE International Conference on Robotics and Automation*

Wagner, C. R., Stylopoulos, N. and Howe, R. D. (2002). "The role of force feedback in surgery: analysis of blunt dissection". *10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002.*

Waldron, K. J. and Tollon, K. (2003). Mechanical Characterization of the Immersion Corp. Haptic, Bimanual, Surgical Simulator Interface. Springer Tracts in Advanced Robotics. 5: 106-112.

Wolfe, B. M., Szabo, Z., Moran, M. E., Chan, P. and Hunter, J. G. (1993), "Training for minimally invasive surgery. Need for surgical skills", Surg Endosc, 7 (2), pp. 93-5.

Wright, P. K. and Dornfeld, D. A. (1998), "CyberCut: A Networked Manufacturing Services", Transactions of NAMRC/SME, 16, pp.

Zar, J. H. (1996). "Biostatistical analysis", Prentice Hall, Upper Saddle River, New Jersey.

Zhang, J., Payandeh, S. and Dill, J. (2002). "Haptic subdivision: an approach to defining level-of-detail in haptic rendering". *Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2002*