AD_____

Cooperative Agreement Number DAMD17-95-2-5023

TITLE: ENT Surgical Simulator

PRINCIPAL INVESTIGATOR: David Heskamp

CONTRACTING ORGANIZATION: Lockheed Martin Tactical Defense Systems
Akron, Ohio 44315-0001

REPORT DATE: November 1998

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 1998	Final (15 Aug 95 - 14 Oct 98)

4. TITLE AND SUBTITLE
ENT Surgical Simulator

5. FUNDING NUMBERS
DAMD17-95-2-5023

6. AUTHOR(S)
David Heskamp

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lockheed Martin Tactical Defense Systems
Akron, Ohio 44315-0001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The Madigan Endoscopic Sinus Surgery Simulator is a state-of-the-art procedural simulator designed to train residents and staff surgeons in the skills necessary for endoscopic sinus surgery. An earlier Midterm Report documented progress in the development and evaluation of the simulator during the first two years of the cooperative agreement, DAMD17-95-2-5023. This report covers progress made during the last year of this three-year cooperative agreement. Researchers on a multi-institutional development team have upgraded the simulator in response to suggestions from a separate multi-institutional evaluation team. The focus of the evaluation effort evolved to characterize the degree and type of transfer of training from the simulator to actual operating room performance.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Endoscopic Sinus Surgery, Virtual Reality, Simulation, Training, and Evaluation	111
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

DA Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

____ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

X For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

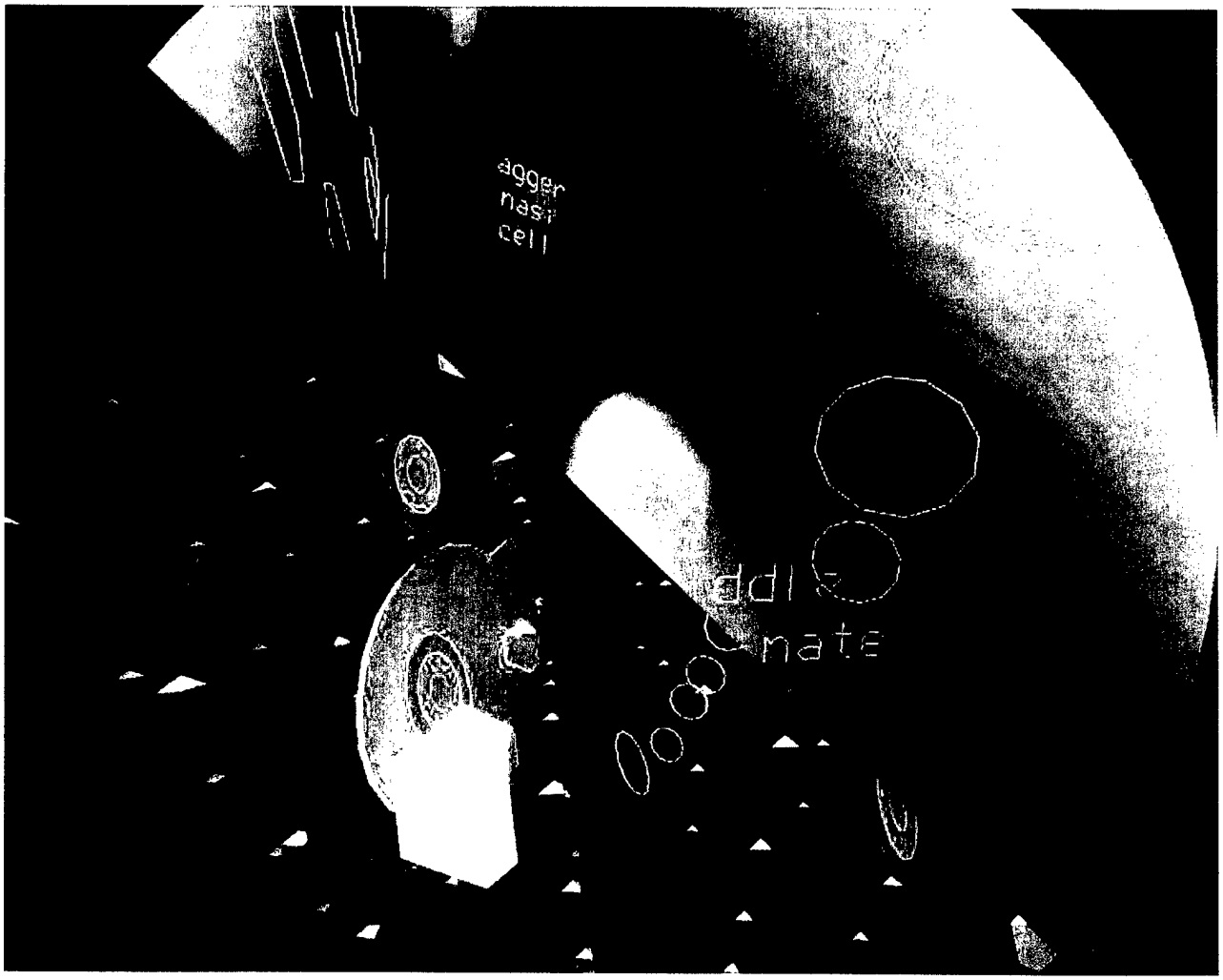
D 3 H-G 22 Oct 98

PI - Signature Date

Final Report
Cooperative Agreement No. DAMD17-95-2-5023

November, 1998

ENT SURGICAL SIMULATOR



Principal Investigators:
Charles Edmond, M.D.
David Heskamp
Major Glen Mesaros, M.D.

Contracting Organization:
Lockheed Martin Tactical Defense Systems
Akron, OH 44315

Prepared for:
U.S. Army Medical Research and Materiel Command
Fort Detrick, MD 21702-5012

TABLE OF CONTENTS

INTRODUCTION

Originally, this cooperative agreement, DAMD17-95-2-5023, was a two-year program running from September 1995 through October 1997. Lockheed Martin applied for and the Army granted an extension to that program. The extension began in October 1997. Because of the uncertainty of receiving funds for the extension, Lockheed Martin prepared and delivered a "final report" in October 1997. Ultimately, the Army extended additional funding to Lockheed Martin and now refers to the "final report" from October 1997 as the "Midterm Report."

The authors attempted to make this a standalone report yet minimize duplication of information contained in the Midterm Report. To gain a complete appreciation for the entire project, it is best to first review the Midterm Report that is in the Defense Technical Information Center's Technical Reports database. This, the final report, covers only the period of the extension to the cooperative agreement, from October 1997 through October 1998. The authors anticipate no further extension to this contract.

Subject

Through cooperative agreement DAMD17-95-2-5023, the research team sought to build and evaluate an endoscopic sinus surgery simulator.

Purpose

The purpose of the extension was to build upon the original work as described in the Midterm Report, addressing recommendations made during the evaluation phase of the original contract. The statement of work for the extension to the cooperative agreement specified the following:

- Expand the simulator evaluation to address the question: "What is the impact of experience with the simulator on actual operating room performance?"
- Upgrade existing computer hardware to double its processing power for sinus surgery simulation.
- Port the simulation real-time software to the new computer and tailor its operation for that machine.
- Continue to refine the original polygonal patient model and the process for model development by:
  - Increasing the number of polygons to take advantage of the new hardware.
  - Segmenting additional features including the sphenoid ostium and sphenoid sinus cavity.
  - Creating polyps and mucocels that can be inserted into the standard patient model.

- Evaluate the National Library of Medicine (NLM) visible female data set. Considering the outcome of the evaluation, use the NLM data or collect data from a suitable cadaver, and create a second patient model.
- Make software revisions that the evaluation phase indicates. Dr. Edmond will prioritize requirements. Implement the new requirements based upon those priorities.
- Evaluate the haptic system. If the evaluation outcome shows that it would be better to apply forces to the endoscope than to the dissection instrument, modify the hardware accordingly.

Note that these items from the statement of work were generated well before completing the original contract. The authors of the proposed statement of work submitted that proposal without the benefit of the knowledge gained from the then ongoing study phase of the original cooperative agreement. Hence, the actual work performed on the extension contract differed slightly from that anticipated, as follows:

- Prior to the extension phase of the cooperative agreement, investigators implemented the sphenoid sinus, sphenoid ostium, and polyps. As described later in this report, patient model development in the extension far exceeded their expectations.
- Regarding the haptic system, the simulator evaluation team garnered sufficient experience with the haptic device to know that the forces were indeed better placed on the dissection instrument. Thus, the effort shifted to improving the haptic device by increasing processing power, and to delivering a stiffer sensation through the haptic tool while improving its reliability and maintainability.

Scope

The scope of research comprised two parts. First, the team made technological advances in response to recommendations from the Midterm Report and in response to the ongoing evaluation of the simulator by clinical experts. These advances involve computational capacity, haptic technology, patient models, and simulation models.

Second, there was the question of the efficacy of augmented virtual reality devices in training surgical procedures. Specifically, the extension allowed investigators to expand the simulator evaluation to determine the impact of experience with the simulator on actual operating room performance. The hypothesis is that the implementation of a minimally invasive prototype surgical simulator enhances graduate medical education and reduces patient-oriented risk through standardized basic training and computer-assisted instruction. Furthermore, residents in training and experienced practitioners will be able to achieve expertise through repetitive proctored challenges, without risk to patients.

Background

The success of computer simulation in flight training and the high cost of medical education motivate the use of computer simulation as a training tool for surgical procedures. In recent years, advances in interactive graphics and virtual reality technology have greatly enhanced the arsenal of instructional tools (1), moving these systems toward more common commercial graphics platforms.

Prior research in surgical simulation includes abdominal laparoscopy (2,3,4,5); limb surgery (6,7); eye surgery (8,9); plastic surgery (10); gastrointestinal endoscopy (11,12,13,14); anesthesiology (15); epidural anesthesiology (16,17); and interventional radiology procedures (18).

Wickham (19) summarizes the need for these novel and extensive training techniques for endoscopic surgery skills:

> Evaluation of new operative competence is urgently needed because of the rapidity of changes in interventional treatment. Training programmes must be established so that interventionists' training is similar to that of airline pilots. A surgeon or radiologist should not be allowed to treat patients with sophisticated and potentially dangerous instruments without the experience of simulated operations and closely supervised procedural training. Fully equipped training centers should be established with simulator laboratories where interventionists can develop the different hand-eye coordination required for the transition from open to endoscopic techniques . . . The need is urgent: the traditional methods of "see one, watch a video, do one" are completely inadequate preparation for minimally invasive techniques . . . A theoretical evaluation of competence by written or oral examination is totally insufficient to determine whether a clinician has gained the manual ability to carry out complex open or endoscopic surgery.

The military has historically been at the forefront of training and evaluation via simulation. In the 1930s, the Army purchased Ed Link's first flight simulator, initiating what is now a billion-dollar simulation industry supporting various defense and space-related activities. Savitsky and LeDonne state, "readiness is the state of being immediately available and capable of performing the mission" (20). Readiness has been military medicine's very reason for existence. Experience in support of the Persian Gulf War casts significant doubts on the capability of the Army medical to fully comply with their mission statement (21,22). The Government Accounting Office (GAO) cited that, in support of the Persian Gulf War, medical personnel "were not qualified in the specialties to which they were assigned, were not physically able to perform their jobs, did not have proper medical credentialing documents, or had not completed training" (23). In response to these deficiencies, the Department of Defense (DOD) enacted a policy regarding medical readiness (24). The DOD directed each of the military services to utilize readiness training programs that include "realistic and skills

training and maximize the use of emerging technology such as distance learning, computer simulation, and virtual reality" (25). In recent years, the Defense Advanced Research Projects Agency (DARPA) played a leading role in funding medical simulation research (26).

Functional Endoscopic Sinus Surgery (FESS) simulation is, arguably, the most practical place to start. It has less rigorous tissue interaction requirements than most other surgical procedures. The anatomy of the sinus region is primarily rigid to a good approximation making FESS simulation less dependent upon the immature technology of modeling deformable objects. Furthermore, the sinuses encompass a small volume enabling the development of an effective low-inertia haptic device. Endoscopy itself limits surgeon interaction to an endoscope and another surgical tool. Instrumentation of these devices is more practical than instrumentation of the hand as required to fully simulate open-air surgery.

During the first two years of the cooperative agreement, investigators installed the prototype device and upgraded it as the result of the evaluation effort. Figure 1 shows the device in operation. At the start of the extension the device comprised the following components (refer to figure 2):

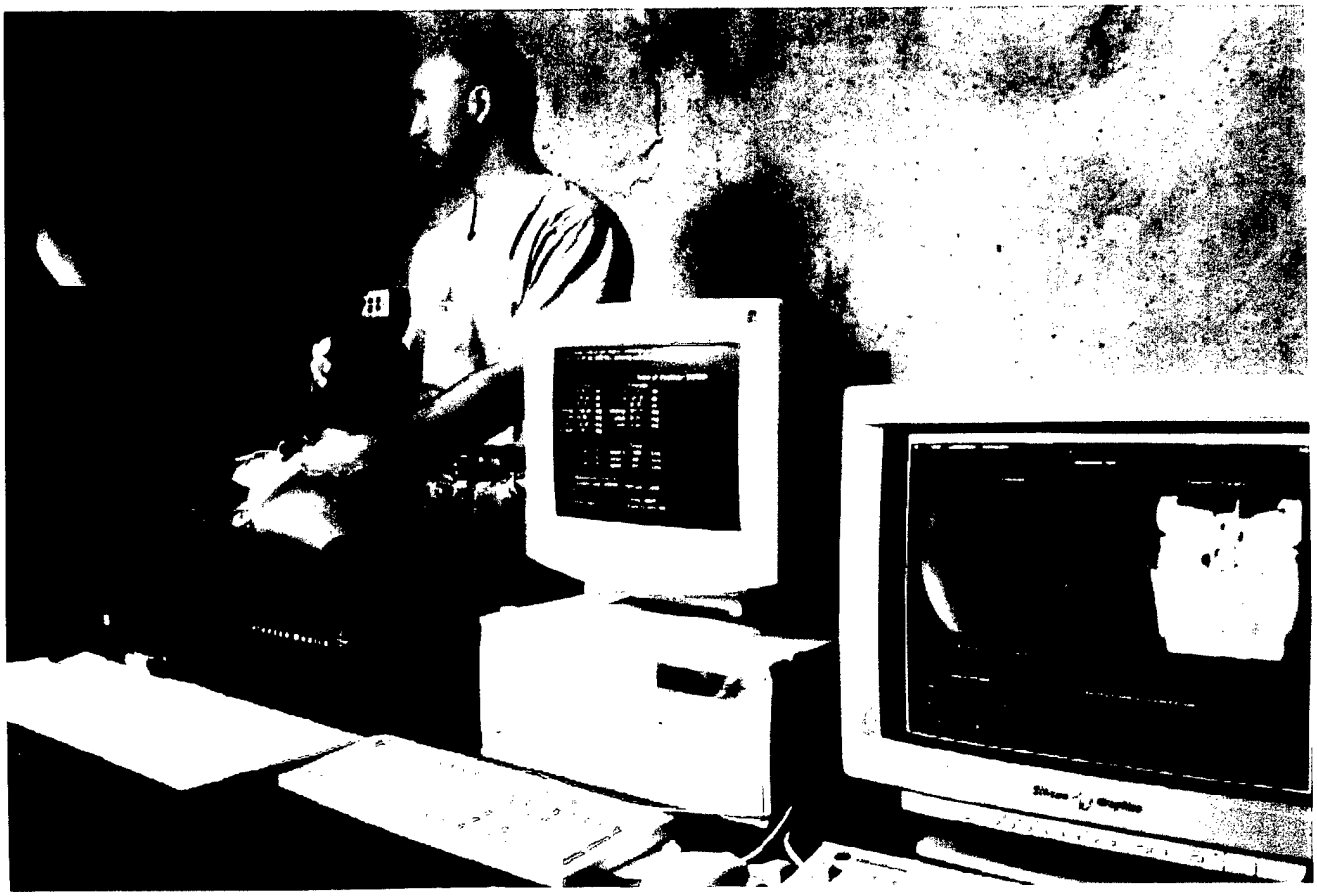


Figure 1. Prototype Device in Operation

- A haptic system consisting of a PC and associated electro-mechanical hardware tracks both an endoscope and a surgical instrument in six degrees of freedom. The surgical instrument has three degrees of haptic feedback. The haptic system also tracks the state of the surgical instrument's scissors-like grip and the state of a foot switch.
- A simulation computer renders real-time simulated endoscopic images. Simulation software in the simulation computer conducts the simulation. Polygonal patient models derived from the NLM Visible Male simulate the patient.
- The 21" monitor presents Computed Tomography (CT) images and endoscopic images and allows proctor control of the simulation through a user-friendly Graphical User Interface (GUI).
- A 20" monitor presents an endoscopic display, similar to that viewed by surgeons in the operating room. External speakers augment the training with audio feedback.

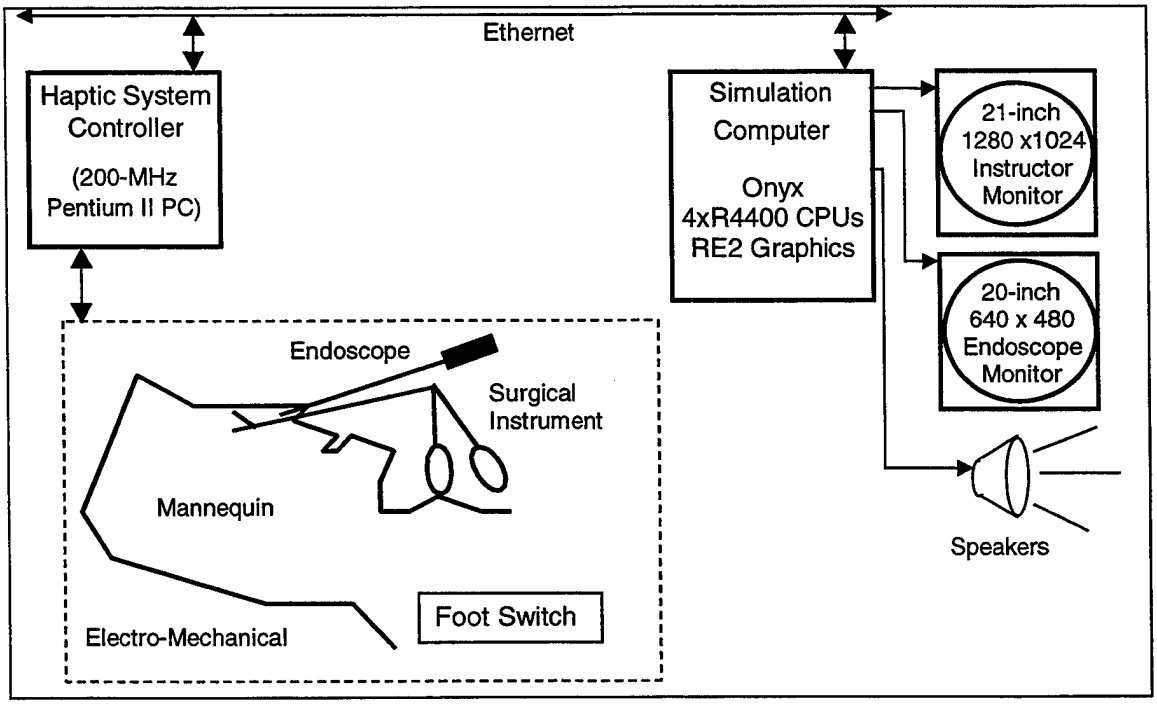


Figure 2. System Components at the Start of the Extension

The simulator allows performance of endoscopic sinus surgery on a virtual patient using replicas of an endoscope and other surgical instruments. The student surgeon freely explores the virtual anatomy by manipulating the simulated endoscope. Using a simulated needle, the student injects a vasoconstrictor into sinus tissue, causing it to blanch and reducing future bleeding. The student displaces the middle turbinate toward the septum with a simulated freer, gaining access to deeper recesses of the sinus cavity. The student dissects the ethmoid bulla, middle turbinate, anterior ethmoid cells, uncinate process, sphenoid ostium and maxillary ostium with any of the modeled dissection instruments. All instruments except the

endoscope provide force feedback in three axes, allowing the student to feel virtual tissue through the instrument's handle.

The student starts training in an abstract novice environment instead of with a virtual patient. Using this environment, the student performs tasks and becomes familiar with the suite of surgical tools. This novice-level training improves hand-eye coordination through immersive experience. Upon achieving an appropriate score at the novice level, the student moves to intermediate-level training where abstract training aids appear in the context of sinus anatomy. Aids for navigation and injection are identical to their novice-level counterparts. They prompt the student to perform the appropriate tasks while learning the anatomy by interacting with labeled anatomical structures. At the advanced level, the student performs surgical procedures without the benefit of abstract computer-based aids.

One of the most critical issues to address in this research is determining the efficacy of simulators in medical training. Lessons learned from the initial effort can be extrapolated to other areas of surgical simulation and simulation in general: model complexity, real-time training requirements, haptic requirements for effective training, novice through advanced training requirements, etc. In addition, this effort focuses on endoscopic surgery, a technique that could prove to be the primary modality for the majority of surgeries, including trauma. Recent advances in endoscopic techniques have extended to minimally invasive neurosurgical procedures, facial reconstructive and plastic procedures and oral maxillo-facial conditions.

Prior to the initiation of the extension of the cooperative agreement, the evaluation team concluded that the simulator was valid for the ESS domain based on the following findings:

- ENT subjects performed significantly better than non-physician subjects did on both the novice (abstract) and intermediate (anatomical with aids) models.
- Initial performance on the novice model correlates with residency level and degree of prior ESS experience.
- Patterns of difficulty for asymptotic performance on the simulator seem to match the typical pattern of subtask difficulty in the operating room.
- Subject ratings of the realism of the virtual anatomical model were consistently high on the post-session questionnaire and in open-ended comments.
- Post-training questionnaire responses confirm that the ENT subjects perceived the simulator as valid and useful for ESS training.

Now, the extension to this cooperative agreement, the subject of this report, takes the simulator and its evaluation to the next step by studying simulator experience relative to operating room performance. The evaluation study described here has been guided by the work of many others in the field. As Hoffman et al. (27) suggest, the end users have been included in the formative stages of the simulator design, a vital step in establishing

educational goals and curriculum design. In developing evaluation criteria the evaluation team took into account both objective and subjective considerations (28), and relied heavily on the basic surgical proficiency measures of time and accuracy (29). The development team incorporated much of the evaluation protocol into the simulator itself (30,31).

BODY

Experimental Methods

The research team for the extension of the cooperative agreement comprised the following individuals and institutions, each with a unique purpose as given in table 1.

Table 1. Team Assignments

TEAM MEMBER	RESPONSIBILITIES
Lockheed Martin	Lockheed Martin, prime contractor for the cooperative agreement: • Performed the simulator systems engineering task • Defined the software architecture • Developed software for rendering, contact detection, dissection, tissue deformation, CT display, student evaluation, voice recognition, and computer-aided instruction • Developed the instrument models • Integrated the major components, including the simulation computer, its software, tactile feedback and patient models
The Madigan Army Medical Center (MAMC)	MAMC provided residents and staff surgeons who participated as subjects in the evaluation of the efficacy of the device. MAMC also provided clinical experts for rating operating room performance. Major Glen Mesaros MD coordinated the trials and provided additional clinical expertise regarding new simulator features.
Charles Edmond MD	Dr. Edmond directed the evaluation effort and the simulator development.
Ohio Supercomputer Center (OSC)	OSC enhanced the standard virtual patient model, adding numerous anatomical components. OSC created a new model based on the NLM female data set.
Human Interface Technology Laboratory (HIT Lab) at the University of Washington	Working with MAMC, the HIT Lab conducted the simulator evaluation, and the "summative" evaluations at project completion. Because the HIT Lab provided no material component to the simulator itself, it could evaluate the simulator without bias.
Immersion Corporation	Immersion Corporation modified the tactile feedback hardware for improved reliability and maintainability.

In this report the term "development team" refers to those institutions that focused primarily on simulator development, including Lockheed Martin, the Immersion Corporation, and OSC. The phrase "evaluation team" includes those individuals and institutions that focused primarily on the simulator evaluation, i.e., Dr. Charles Edmond, MAMC, and the HIT Lab.

During this phase of the program, the development team continued development in parallel with the evaluation. A complete simulator stayed at Madigan to serve the needs of the evaluation team. Lockheed Martin used its own capital equipment simulator to serve the needs of the development team at its facility in Akron. The Immersion Corporation loaned the OSC a haptic device that combined with OSC computers to form a complete simulator for use in Columbus. Thus, for the duration of the extension, three devices allowed work to proceed in parallel.

The extension to the contract began with version 1.2 of the simulation software. As the evaluation team generated new requirements, the development team incrementally implemented them and released new software so that the evaluation team could work with the new software and provide feedback. An incremental software development process describing the approach governed the software development and passed an Underwriter's Laboratory audit for ISO 9001 accreditation. The C++ simulation software, developed according to current object-oriented design principles, created a robust foundation for iterative refinement in response to simulator evaluation. Dr. Charles Edmond played a pivotal role in the process by establishing requirements and, along with Dr. Mesaros, testing them for adequate implementation.

Evaluation

Evaluation efforts for this project fall into two general categories:

- "Formative" evaluation, which attempts to provide design specification input to the development team during the development process
- "Summative" evaluation, which assesses the success of that effort by formally analyzing the effectiveness of the system

Details of the initial formative and summative evaluations are in the Midterm Report. Specifically, the evaluation team presented evidence for the validity of the system as an endoscopic sinus surgery simulator.

The evaluation team focus during the current phase was on the following three questions:

- QUESTION 1 - What is the impact of experience with the simulator on actual operating room performance?
- QUESTION 2 - How does experience with the simulator affect individual components of surgical competence?

- QUESTION 3 - What is the effectiveness of the various training aids and protocol methods incorporated into the simulator?

The first of these questions appears in the statement of work for the extension to the cooperative agreement.

While previous evaluation efforts focused on simulator *validation*, the current phase focuses on *transfer of training* issues. Evaluation studies used version 2.0 of the simulator. To evaluate effectiveness of the simulator as a training tool, investigators systematically measured several possible indicators of surgical skill:

- Real-time performance in the OR as assessed by an attending staff ENT proctor
- Blind analysis of endoscopic video acquired during these procedures
- Independent paper-and-pencil tests of several components of surgical competency
- Simulator scores on various training exercises
- Time spent on the simulator
- Endoscopic sinus surgery experience

Ten junior and senior ENT residents at MAMC served as subjects, some of whom had prior training with the simulator. Since the number of available residents was limited researchers performed quasi-experimental studies rather than full factorial design experiments. As such, these should be considered "pilot" studies which provide some potentially useful methods for evaluating surgical simulators, as well as some suggestive findings for further exploration.

In addition to measures of transfer of training, researchers continued assessing the usefulness of various components of the simulator. Because of the number of features of interest, systematic controlled variable studies of utility were not possible within the scope of this project. Instead the evaluation included subjective assessments by experienced staff consultants, by the resident trainees, and by the technical proctor. These assessments may provide fruitful hypotheses for future studies.

The evaluation team collected several measures to determine whether there was a statistical correlation, including simulator scores, operating room performance ratings, ratings of videotaped operating room procedures, and surgical competency ratings. The following sections describe the methods used to capture them.

Evaluation – Simulator Protocol

The simulator score protocol began with the proctor introducing each subject to the overall configuration of the simulator. The subjects familiarized themselves with the voice recognition software, audibly selecting some of the available menus. The proctor executed

commands at the system console if subjects found use of the voice recognition software difficult.

The proctor then introduced the subject to the instrumented endoscope and its 0, 30, 45 and 70 degree software configurations. Subjects learned to rotate the angled scopes axially, by rotating the shaft of the instrumented endoscope, and to rotate the image by speaking to the voice recognition software.

The proctor introduced the subjects to the instrumented forceps and its ability to simulate all the virtual dissection tools and injection needles. The proctor described the mechanics of the instrumented forceps (open and closed) as simulating the syringe plunger during injection and opening and closing the jaws of the dissection tools.

Subjects received a brief verbal description of the three subtasks (navigation, injection and dissection) and what would be required of them during the trial. Each subject watched the simulator recreate or play back a procedure performed by Dr. Mesaros, MAMC staff otolaryngologist and co-investigator. During the playback, subjects stepped up to the mannequin and became familiar with the instrumentation. The proctor described the subtasks in more detail, including requirements for completing the trial. The playback depicted blood affecting the scope and the proctor described reasons for it, along with how to remedy the problem by wiping the instrumented scope on the foam pad located on the mannequin's forehead. The proctor encouraged subjects to ask questions and "think aloud" during the procedure.

After completing the novice model, subjects proceeded to the intermediate model. In this model, subjects attempted to perform a total ethmoidectomy, including the navigation, injection and dissection subtasks, using training aid labels embedded in the virtual paranasal anatomy. Again, the simulator first played a recorded procedure while the proctor explained requirements for completing the trial. A subject's familiarity with the procedure determined the need for further instruction by the proctor during the trial.

Subjects then proceeded to the advanced model. This model required subjects to perform a total ethmoidectomy and remove three polyps lateral to the middle turbinate without the help of the training aids. During this trial, the subjects had to remember the navigation paths, injection sites and anatomical components to dissect. The subject's familiarity with the procedure determined the need for further instruction by the proctor during the trial.

Evaluation – Operating Room Performance Ratings

An experienced staff otolaryngologist proctored each of four subjects as they performed their first surgeries. All subjects performed a routine endoscopic sinus surgery procedure. Two of the subjects had trained on the simulator. Two had no simulator training. Proctors evaluated the residents on 10-point scale where 1 is inadequate and 10 is perfect for each the following aspects of the procedure:

- Patient Preparation
    a. Turn the bed
    b. Patient positioning
    c. Intranasal preparation
        1) Placement of pledgets with topical vasoconstrictor and local anesthesia
    d. Facial skin prep
    e. Surgical draping of patient
    f. Monitor positioning
    g. Set-up of endoscopic equipment (suction, endoscopes, microdebrider, Mayo stand, etc.)
    h. CT scans on view box
- Navigation (Left and Right)
    a. Inferior pass
    b. Intermediate pass
    c. Superior pass
- Injection (Left and Right)
    a. Uncinate
    b. Root of middle turbinate
    c. Lateral aspect of middle turbinate
    d. Sphenopalatine artery
- Medialization of middle turbinate (Left and Right)
- Uncinectomy (Left and Right)
- Maxillary antrostomy (Left and Right)
- Anterior ethmoidectomy (Left and Right)
- Posterior ethmoidectomy (Left and Right)
- Sphenoidotomy (Left and Right)
- Overall rating

Evaluation – Ratings of Videotaped Procedures

In addition to the proctor's ratings of resident operating room performance, a videocassette recorder captured the endoscope video of each procedure for further analysis. A blind panel of four experienced sinus surgeons rated five of these videotapes (four first time surgeries from above and one operation performed by an experienced staff surgeon). The panel rated each videotape on the same 10-point scale for the following performance criteria:

- Navigation
- Injection
- Uncinectomy
- Anterior ethmoidectomy
- Maxillary antrostomy
- Tool-tool dexterity
- Tissue respect
- Proper depth of image for task
- Image-task alignment
- Tool manipulation
- Orientation of video image
- Tool selection
- Surgical confidence
- Case difficulty

Evaluation – Surgical Competency Ratings

A total of nine otolaryngology residents from MAMC completed the surgical competency ratings. Residents represented each of the five residency years. The subjects viewed a videotape recording of a simulated procedure generated from the Madigan ENT surgical simulator. Researchers stopped the tape periodically and subjects answered multiple choice questions. Subjects completed other components of the exam while viewing a series of still images also from the simulator. The exam attempted to ascertain the subjects' surgical competency through five dependent measures: anatomical knowledge, endoscope angle awareness, procedural knowledge, instrument knowledge, and spatial awareness. Six of the nine subjects in the study group had experience with the simulator prior to this testing.

Results and Discussion

The Simulator – Overview of Improvements

Refer to the Midterm Report for a detailed description of the simulator. Refer to the introduction in this report for an abbreviated overview of the simulator at the beginning of the contract. Figure 2, from the introduction, depicts the simulator hardware at the inception of the extension to the cooperative agreement. Figure 3 below reflects the simulator after upgrades specified in the statement of work and upgrades requested by the evaluation team.

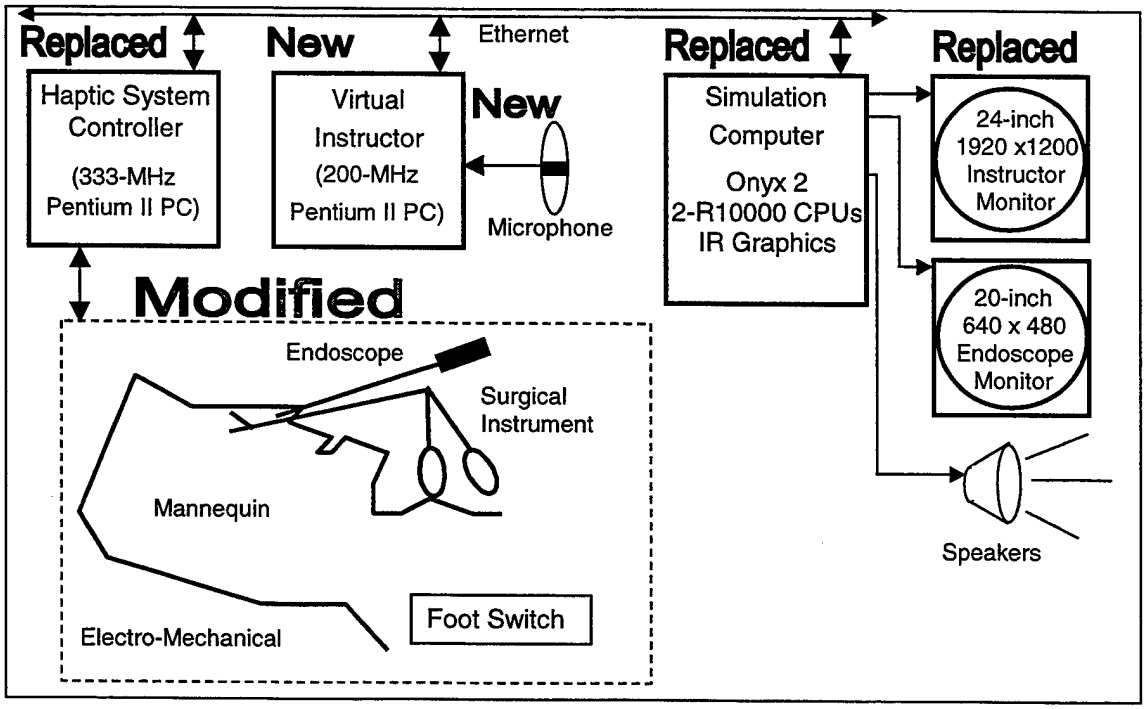


Figure 3. Simulator Components after Upgrade

Engineers from Immersion Corporation modified the haptic system electro-mechanical system hardware to improve its maintainability. The development team replaced the previous 200-megahertz (MHz) haptic system controller with a faster 333 MHz computer and used the old computer as a voice recognition computer now known as Martin, the Virtual Instructor. Martin listens for and responds to spoken commands, allowing hands-free control of the simulator. Most importantly, the development team replaced the SGI Onyx and its monitor with a new Onyx 2 and monitor that markedly increased the power available for simulation. Engineers developed new software for the virtual instructor and modified the software residing in the simulation computer and the haptic controller. The endoscope monitor is similar to operating room equipment and did not change. The following sections discuss changed components in detail.

The Simulator – Haptic System

The haptic system comprises a high-speed Pentium computer and an electro-mechanical component. It tracks the position and orientation of the endoscope through a mechanical apparatus external to the mannequin's head, and tracks another surgical instrument via a second mechanical apparatus inside the mannequin's head. The haptic system also monitors the state of the instrument's scissors-like grip and of the foot switch. As the user manipulates these physical replicas of an endoscope and surgical instrument, the connected electro-mechanical hardware senses their position and orientation in all six degrees of freedom. The haptic system PC reads and transfers the complete state of both tools to the simulation computer. The haptic system applies force in three axes to the distal tip of the surgical tool, simulating haptic cues associated with surgery. The haptic system does not apply force to the endoscope replica.

Prior to the start of the extension of this cooperative agreement, project clinical experts identified several drawbacks to the current design. Among the comments were lack of force on the endoscope, lack of torque on the surgical instrument and insufficient cues for palpation and tip constraint. While the following sections address the findings relative to these concerns, the reader should not draw the conclusion that the forces generated by the haptic device lack value. To the contrary, the investigators found that forces applied by the haptic device are very useful for augmenting depth perception and they enhance the immersive experience.

Haptic System – Reliability and Maintainability

Just prior to the completion of the Midterm Report the Immersion Corporation modified the electro-mechanical component (figure 4) of the haptic system to improve its reliability. The development team monitored these changes in Akron and at MAMC for several months to determine the effectiveness of the modification. They proved highly successful reducing the mean time between failures from a couple of days to over a month. When combined with an appropriate preventative maintenance program the haptic device should be reliable for installation in a commercial facility.

During this phase of the contract Immersion engineers focused on maintainability. They replaced keyless connectors with keyed connectors. They labeled all connectors so that the system could be disassembled for repair on-site. They replaced miniature set screws, which required fragile Allen wrenches that were not holding up, with screws that use a 3/64-inch or larger Allen wrench. These changes allowed the common repairs to be made at the installation site instead of at the factory.

Haptic System – Lack of Force on the Endoscope

When examining patients, clinicians tend to rest the shaft of the scope against structures in the nose. On the simulator, the haptic system applies no haptic feedback to the endoscope, forcing clinicians to control the scope visually. Some find it difficult to keep the simulated eyepoint inside the simulated anatomy using only visual cues. Until adequately trained to

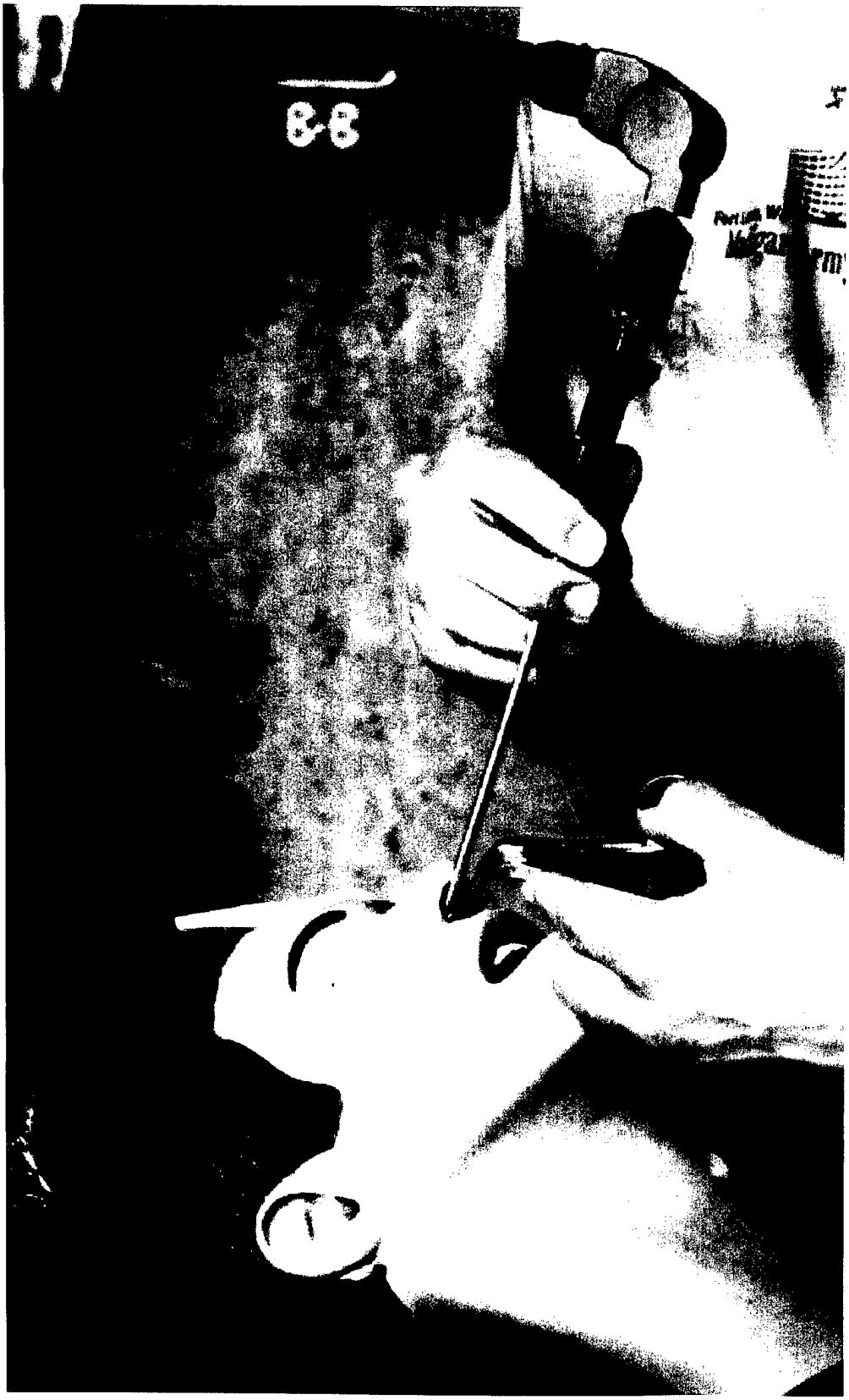


Figure 4. Electro-Mechanical Haptic System Component

control the scope entirely by sight, surgeons find their attention consumed by maintaining the endoscope's tip in an appropriate position and cannot adequately concentrate on dissection with the surgical instrument. Endoscope control is a skill that the student must master before proceeding to the surgery itself on the simulator. At the time of writing of the proposal for this extension to the contract, team members viewed this as a liability. Upon further evaluation, the team began to view this as an asset. Although more difficult, simulator users learn to control the endoscope without benefit of resting it against the tissue. This is particularly important during endoscope insertion where contact with the walls of the nasal passage tends to scrape the mucus membrane causing bleeding that at best makes the operation more difficult.

At the recent American Academy of Otolaryngologists meeting, approximately a hundred experienced sinus surgeons spent 10 to 20 minutes on the simulator. The majority became proficient enough at controlling the endoscope in that time period that they successfully attempted a modest dissection of the uncinate and ethmoid bulla. Others quickly mastered control of the endoscope and performed the difficult dissection of the frontal recess or maxillary ostium using an angled scope. Still others required more time than was available to master control of the scope and perform limited surgery.

The statement of work for this extension states that: "If the evaluation outcome shows that it would be better to apply forces to the endoscope than to the dissection instrument we will modify the hardware accordingly." Clinicians nearly all felt that forces applied to the surgical instrument are more beneficial than forces applied to the scope. Reflecting upon these results, developers believe that since the endoscope provides monocular (i.e., depthless) vision, it is imperative that the simulator user be able to feel the tip of the tool as it interacts with the anatomy. Investigators observed students and experienced surgeons pushing the needle far into the tissue, sometimes beyond the anatomical component itself. Once sensitized to the depth cues afforded by haptic forces, students gained better control of the instrument.

Thus, it became clear that the force should remain on the surgical tool and investigators turned their attention to the other weaknesses in the haptic system.

Haptic System – Lack of Torque on the Surgical Instrument

The current haptic system provides forces along three axes at the tool's tip. It provides no torque on the shaft. The forces enhance the student experience, making the surgery easier by guiding the tip of the surgical instrument within the patient anatomy. The student can "feel" the anatomy as the tool's simulated tip contacts the simulated surfaces of the patient anatomy. The haptic cue augments the student's depth perception in an environment where monocular vision provides little depth perception. However, the tool is approximately 6 inches long and forces applied to its distal tip feel unnatural when sensed by the student from the grip at its proximal end. Additionally, without torque feedback, it is possible for the student to maneuver the tool into positions and orientations that are unattainable in the real world. For example the student can operate on the left nostril while the shaft of the tool is

through the right nostril. The student, therefore, can develop improper tool manipulation technique and receive negative training.

Unfortunately, the non-recurring cost and design time to create a device that applies six degrees of force made it impractical within the funding and schedule of this cooperative agreement. Engineers did, however, design a plug that inserts easily into the opposite nostril limiting the proximal end of the shaft to the appropriate nostril. Combining that limitation with the force applied by the electro-mechanical system at its distal end applies some torque to the shaft and constrains the student to more natural positioning of the tool.

Haptic System – Insufficient Force Cues

The problem of applying force is the inability to simulate rigidity. For example, when the tool contacts the skull, one should feel an extremely hard surface that the tool cannot penetrate without undue force. Three years ago at the inception of this program, investigators imagined the ability to use haptic forces to simulate the force of palpation enabling a doctor to tap on an anatomical structure and determine its makeup among bone, soft tissue and something in between.

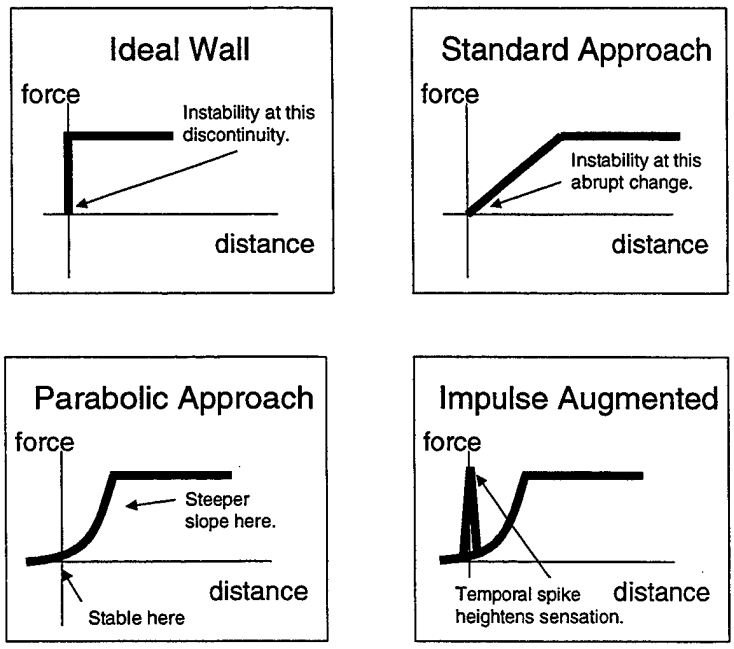


Figure 5. Various Force Curves

For soft tissues, the solution is simple. However, for hard bone-like structures the haptic problem is substantial. Investigators experimented with the shape of the force curve used to simulate contact with bone. The curve shown in figure 5 shows the industry-standard approach to simulating a wall (32). Changing the slope of the line improves the sensation of rigidity but introduces instability in the system. Investigators experimented with a parabolic approach as shown in figure 5. In this approach the force becomes a function of the square of

the distance, creating a steeper slope and making the wall feel more rigid. However, this yields only a marginal improvement and the sensation does not approach that of tapping on bone. Further augmentation of the force curve through use of a temporal spike substantially increases the user's awareness of contact. Humans are more sensitive to force transients than steady state forces (32). However, investigators did not have time to explore whether altering the amplitude and duration of the temporal spike based on tissue type would be effective in the simulation of the cues of palpation.

To further increase the force applied at the distal tip of the tool, investigators replaced the previous 200 MHz haptic controller PC with a new 333-MHz PC. Taking advantage of the increased power and restructuring software so that it ran evenly from frame to frame, developers increased the frame rate from about 3000 to 4080 hertz. This allowed an increase of 18 percent in the slope of force curve without introducing instability.

While developers made progress toward the goal of simulation of palpation cues, the simulator continues to lack the ability to provide haptic palpation cues of bone, soft tissue and something in between.

The Simulator – Simulation Computer

Unlike simulating terrain on a flight simulator, the close proximity of the anatomical structures to the viewpoint causes a high percentage of these polygons to fall simultaneously inside the viewing frustum. This places a heavy load on the simulation computer and graphics hardware. At the completion of the Midterm Report, the heart of the simulator was an SGI Onyx with four R4400s and Reality Engine II graphics. The statement of work for this extension to the cooperative agreement states that the development team would:

- Upgrade the existing computer hardware at least doubling its processing power for the application.
- Port the real-time software to the new computer and tailor its operation for that machine.

After running benchmark tests on several computers, Lockheed Martin purchased and delivered a new SGI Onyx 2 computer with two R10000s and an Infinite Reality Graphics pipeline. The development team ported the real-time software to take advantage of the new computer using the latest compilers and operating system. Simulator latency, a key measure of the computer's effectiveness, varies widely depending on the number of central processing units (CPUs) and frame rate. Frame rate in turn depends upon the number of CPUs, system architecture, graphics speed, and CPU speed. The real-time simulation software senses the computer configuration, dynamically configuring itself to take full advantage of its hardware. For example, upon sensing four CPUs, the real-time simulation software spawns four parallel processes, one for each CPU. The table below compares key performance characteristics between the new system and the old system illustrating that while nearly tripling the number of polygons, the new computer processes them at a faster rate and cuts the latency in half.

Table 2. Performance Characteristics

	Old System	New System
Hardware	4xR4400 RE II Graphics	2xR1000 IR Graphics
Patient Model Size (~ polygons)	25,000	69,000
Update Rate (hertz)	20	30
Latency (~ milliseconds)	200	100

Based on simulator benchmarks, the new computer nearly quadrupled the processing power of the old Onyx. That is, a patient model with nearly four times as many polygons would process without increasing latency. Dr. Edmond preferred a reduction in latency to an increase in polygons. The new 100-millisecond latency is in line with flight simulator response times yielding better response to operator input.

The Simulator – Virtual Instructor

Reviewers will not find any reference to this new component in the proposal or statement of work. It falls under the area in the statement of work wherein the development team makes "...revisions that the evaluation phase indicates." Billinghurst (33) preceded this work by demonstrating the viability of voice recognition and audible feedback integrated into a virtual environment.

Developers created a new unit for the simulator and named it Martin, the virtual instructor. Martin provides hands-free control of the simulator so that a simulator user can train without assistance from another person. Before Martin, the student surgeon would ask someone to change tools using the graphical user interface on the instructor monitor. The following dialog illustrates how Martin functions.

Surgeon : "Martin."
Martin : "Yes?"
Surgeon : "Go."
Martin : "Go."
Surgeon : "Help."
Martin : "Complete the examination. Proceed inferior to the inferior turbinate. Stop at the nasopharyngeal outlet. Turn hoops on before proceeding."
.
. (... After 10 seconds)
.
Martin : "I'm sleeping."
.
.
.
Surgeon : "Wake up."

Martin	: "Yes?"
Surgeon	: "What's this?"
Martin	: "You are pointing to the agger nasi cell."
Surgeon	: "Help."
Martin	: "Wipe the endoscope to improve visibility."
Surgeon	: "Help."
Martin	: "Dissect the uncinate to expose the maxillary ostium. Use the left-biter or the down knife. Remove exposed bone fragments."
Surgeon	: "Knife."
Martin	: "Sickle Knife."
.	
.	(... After 10 seconds)
.	
Martin	: "I'm sleeping."
.	
.	
.	
Martin	: "Uncinate dissected."
Martin	: "Congratulations. You have completed the exercise."

Martin recognizes a wide variety of voices and responds to over sixty commands. It has a vocabulary of over 180 words and recognizes multiple words for the same command. For example, one might say "hummer" or "microdebrider" to select the same tool. Training Martin to recognize the voice of each new student is not feasible because student surgeons seldom have the extra time it would take to do so. Given these constraints, investigators determined that the following characteristics provide the best voice recognition system for the sinus surgery simulator:

- The vocabulary must be limited to the commands that Martin recognizes. Without a limited vocabulary, the voice recognition system would require extensive training for each individual that uses it.
- The vocabulary must also have an extensive garbage collection so that Martin ignores invalid utterances. For example, students often exclaim, "Nice!" If the word "nice" is not a member of the garbage collection vocabulary, Martin selects a sickle knife. Placing the word "nice" into the vocabulary and ignoring it solves that problem. There are approximately 30 such words including "da," "ah," "hi," and "oh."
- Martin must have a "sleep" mode so that the student surgeon can carry on a conversation without Martin misinterpreting what is said and accidentally altering the simulator's state. So, after 10 seconds, Martin announces, "I'm sleeping" and responds to only two commands: "Wake up" and "Martin."

Developers experimented with several voice recognition software packages. They found that voice recognition is computationally intensive, requiring its own CPU. They eliminated

packages that reside on the Onyx because that would have forced the purchase of an additional CPU at great expense. Ultimately developers selected a PC as the host because of price. The old haptic controller computer became the virtual instructor computer. Developers chose Dragon Dictate and Visual Basic as the development software.

While the PC performs speech recognition, the Silicon Graphics Incorporated (SGI) computer performs speech generation tasks by playing back pre-recorded files in a background server. This gives closed loop control and avoids conflict on the PC's sound card by allowing it to function only as Martin's ears. The sequence of events that causes the simulator to react to spoken commands is as follows. The student surgeon issues a command into the Virtual Instructor PC microphone. The PC recognizes the command as valid and transmits a corresponding string over Ethernet to the simulation computer. A background task in the simulation computer decodes the string, places the command into shared memory and forwards the string to the audio server. The real-time software acts on the command. The audio server fetches the appropriate file and plays it through the speakers. The student hears and sees the response, closing the loop. All of this takes 2 to 3 seconds.

The response to Martin was overwhelmingly positive. Martin recognizes men's and women's voices well. Even children can effectively issue commands using Martin.

The Simulator – Real-Time Software

The statement of work states that developers will "... make software revisions that the evaluation phase indicates. Dr. Edmond will prioritize requirements." The development team implemented new requirements based upon those priorities. This section discusses changes made to the real-time software beyond those already discussed regarding software in the haptic controller computer and virtual instructor computer. Table 3 lists improvements to the real-time software beyond those already discussed.

Several of the changes involve negative scoring. The evaluation team found that the simulator, at the completion of the Midterm Report, had no features that encouraged students to be careful during dissection. Particularly in the novice model, students who were reckless in the dissection task would often score as well as or better than students who exercised more care. Students at that time dissected balls placed randomly in space. For this phase of the cooperative agreement, developers created green, tool-specific, dissection objectives, positioned and oriented to align with appropriate anatomical structures for each tool. Developers also included red-colored obstacles that force the students to use proper tool orientation. The simulator penalizes students who accidentally dissect portions of the red obstacles (figure 6). This change mirrors the patient anatomy; the obstacles now function in a manner similar to arteries, nerves, and fragile bone. For example, if the student dissects the anterior ethmoid artery, the student's score is reduced and a warning given. It is now possible to dissect the lamina papyracea exposing the periorbital fat and reducing the student's score. Figure 7 illustrates these features. The authors completely removed the ethmoid cells that would have been present in figure 7 in order to clarify the new features. Students now dissect cautiously, improving the overall training that the simulator provides.

Table 3. Simulator Improvements

The uncinate is now malleable so that when pushed, it yields a crease for guiding the incision.
The sickle knife now cuts in a downward direction.
The endoscope camera and lens now roll independently.
The instructor can reset the orientation of the endoscope light cable feed.
Obstacles now provide negative scoring in the novice model.
New patient components warn the student and subtract from his score.
The simulator computer aided instruction contains enhanced suggestion features and a new feature that allows the student to point to an anatomical component and receive an audible identification.
New tools include a round burr, aggressive router, curved router, bayonet, and seeker.
A time stamp, added to frame-based recorded records, allows more thorough analysis of recorded data.
Simplified calibration of the forceps.
Labels now identify every segmented patient component.
Placement of novice-level objectives is identical to the anatomical counterparts.
Dissection of arteries causes massive bleeding.
The student can now pack the nasal cavity after the operation is complete.
The student scorer timer now stops when scores are satisfied allowing the student to dissect additional structures for practice without penalizing the score.
A new 45-degree scope reflects a recent product announcement from the Karl Storz Company.
The simulator announces the injection and dissection progress once, at completion of each component.
The computer aided instruction software now plays an unlimited number of suggestions.
The middle turbinate now dissects with more realism.

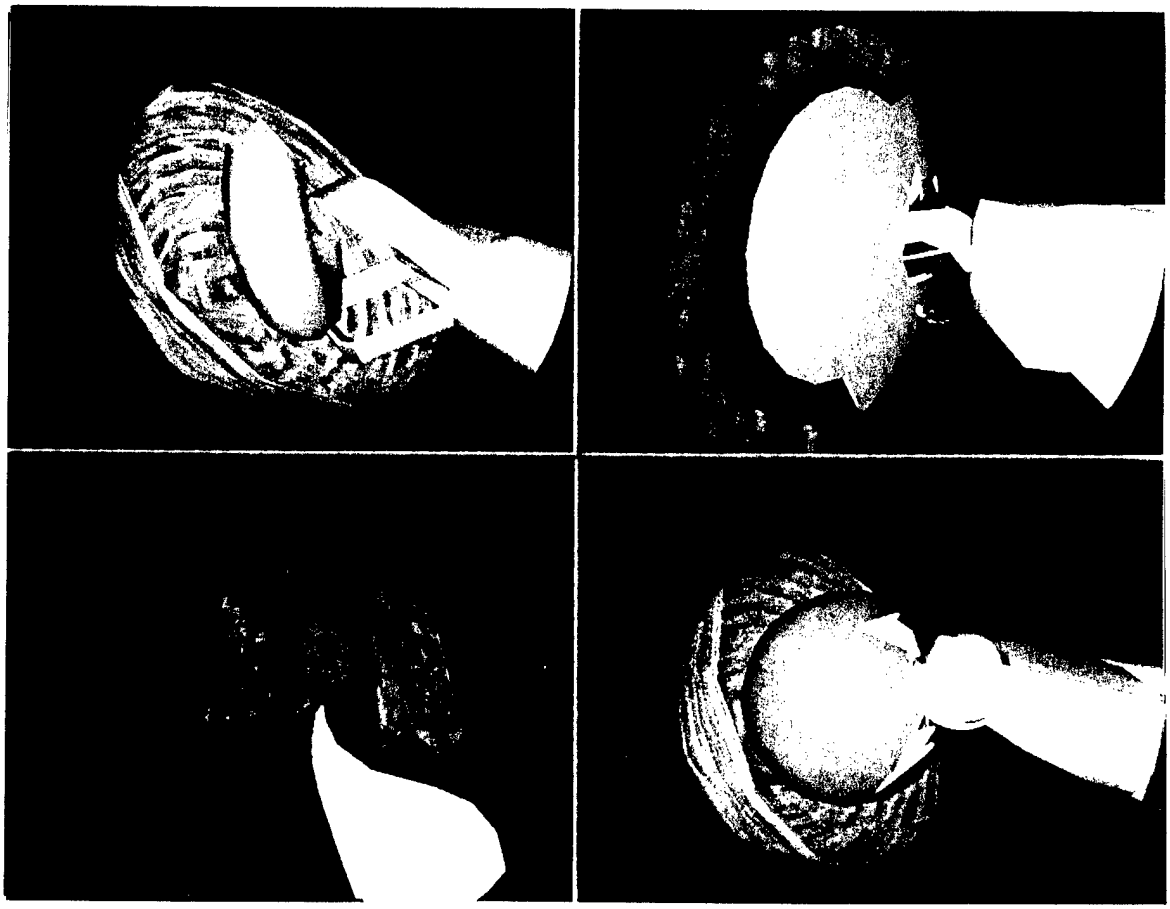


Figure 6. Tool-Specific Objectives (Green) and Obstacles (Red)

The Simulator – Patient Model

Patient models prior to this phase of the cooperative agreement utilized the complete capacity of the original Onyx computer. The capacity of the new Onyx 2 computer allows for substantially upgrading the patient model through additional detail (more polygons) and additional patient components. In their evaluation of the simulator, Dr. Edmond and others defined the need for additional patient models and a library of pathological patients (34). Investigators believe that by providing patient variance within the simulator, residents would learn to deal with the variability encountered in clinical practice.

The statement of work for this extension to the cooperative agreement states that the development team would:

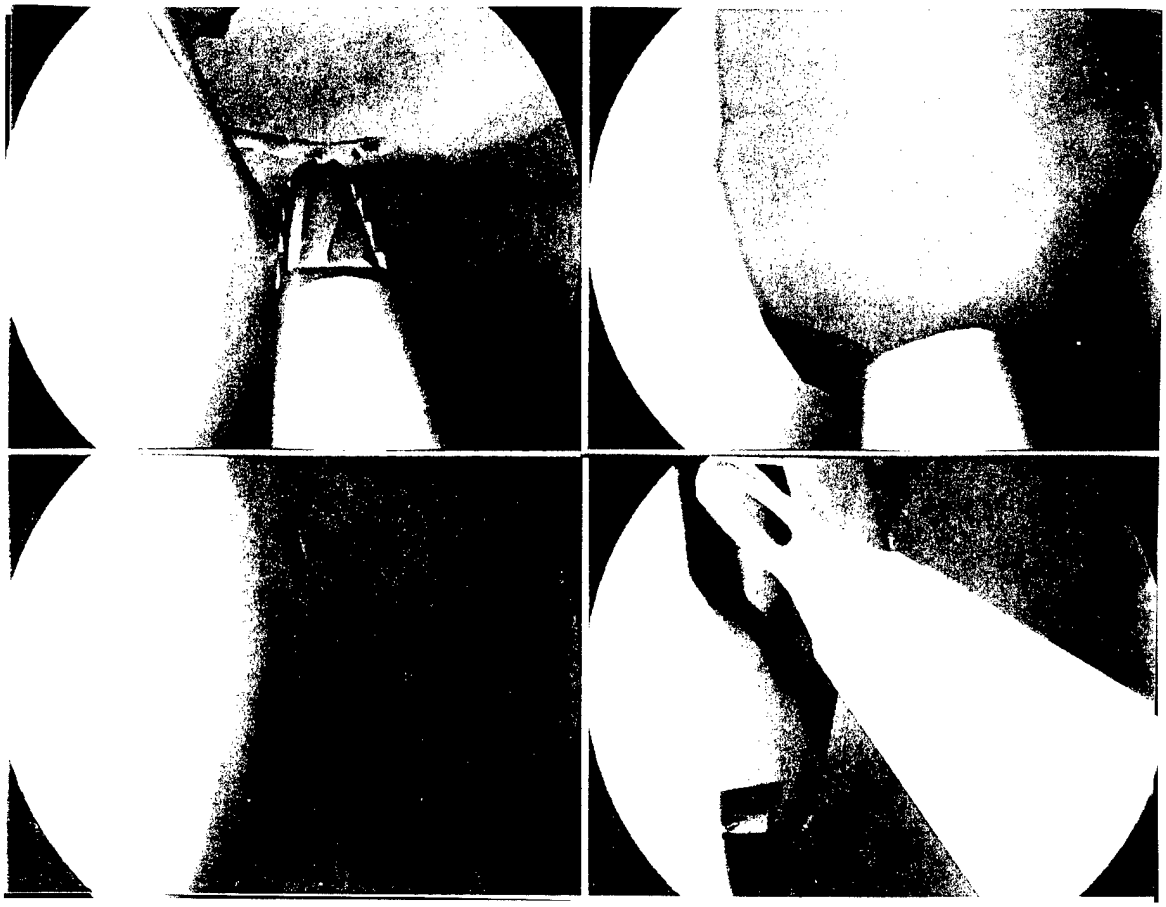


Figure 7. Top: Dissection of Anterior Ethmoid Artery (before and after)
Bottom: Dissection of the Lamina Exposes Periorbita and Fat (before and after)

- Under the direction of Dr. Edmond, continue to refine the current polygonal patient model and the process for model development by:
  - Increasing the number of polygons to take advantage of the new hardware.
  - Segmenting additional features including the sphenoid ostium and sphenoid sinus cavity.
  - Creating polyps and mucocels that can be inserted into the standard patient model.
- Evaluate the NLM visible female data set. Considering the outcome of the evaluation, the development team will use the NLM data or collect data from a suitable cadaver to create a second patient model.

Because investigators submitted this proposal prior to completing the original contract, the statement of work specifies some activities that were accomplished before the start of this extension. Specifically, models of the sphenoid sinus, sphenoid ostium, and polyps were part of the simulator previous to the extension. Investigators continued the evolution of the patient model, generating nearly twenty additional patient components. These new

components expand the simulator's capability to simulate sinus surgery from performance of a limited anterior ethmoidectomy to more comprehensive surgical procedures that include posterior ethmoidectomy and dissection of the frontal recess.

Patient Model – The NLM Male

During the first quarter of the contract extension, investigators at the OSC concentrated on generating the additional components and other revisions to the male patient model. Additional components include the lamina papyracea, periorbita, nasolacrimal duct, agger nasi cell, pituitary gland, frontal recess, frontal ostium, frontal sinus, anterior ethmoid artery, choanae, eustachian tube, adenoid pad, nasopharynx, ophthalmic artery, carotid artery, optic nerve, superior oblique muscle, medial rectus muscle, skull base and periorbital fat. These new components, along with the components from the original cooperative agreement, allow the students to fully explore and more fully interact with the virtual patient. The original goal of allowing the students to perform a limited anterior ethmoidectomy has been fully realized and extended to allow dissection of the frontal recess and posterior ethmoid sinuses. New components such as the skull base and the lamina papyracea provide boundaries that confine the operation. Other more delicate structures, such as the periorbita, arteries, and nerves, replicate perils found in humans, exposing the students to situations that teach avoidance in the operating room.

The simulator now has a library of pathologies, including a mucocele, polyps, deviated septum, concha bullosa, and bullet wound. Combined with the standard male and the new female models, these cases provide a rich variety of training scenarios.

To create additional patient model components, developers at the OSC identify components in the cryosections and refine the color mask images using existing *Editmask* software. *Editmask* provides a robust graphical user interface for loading and editing multiple color sections. The user delineates regions of interest in image space by manually painting over the cryosections with a unique color for each patient component. *Editmask* produces segmented color image masks as shown in figure 8. Color masks form a volume that feeds the Visualization Tool Kit (VTK) pipeline, generating isosurfaces in the form of Inventor™ files for use on the simulator.

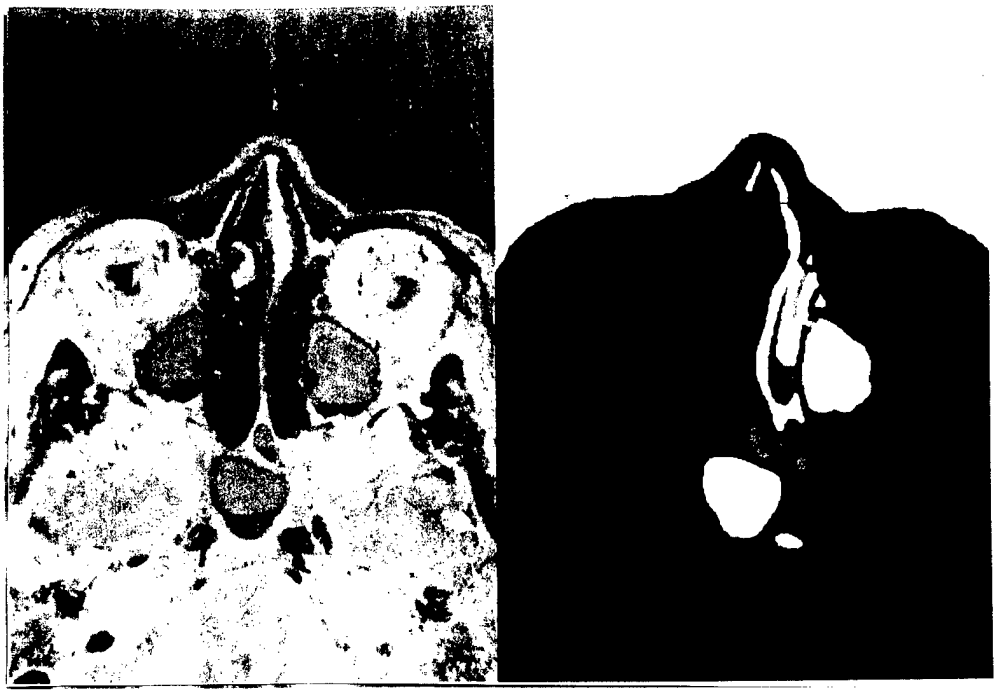


Figure 8. Left: Cropped Image from Original Cryosection Image: Level 1405
Right: Color Mask Image of Segmented Structures Found at Level 1405

Construction of small arteries, such as the anterior ethmoid artery, proved problematic because of the slice resolution of the male image set. Clinicians and anatomical experts could not identify them in the cryosections. Arteries with a diameter of 1mm or less could fall between adjacent slices. The investigators' original approach was to paint these structures by hand using the cryosections as a guide. This approach yielded jagged and flat structures.

Subsequently, investigators at OSC improved the results by constructing arteries using control points found in the cryosections and a spline, producing a smoother representation. This technique is similar to the approach that generated the wound path in the trauma model in the earlier phase of the cooperative agreement. After identifying the appropriate control points by hand, custom software interpolates a curve connecting these points to create a spline. The software then creates a tube around the spline using Catmull-Rom spline approximation.

The algorithm works in three dimensions. Given the control points, the algorithm first identifies all voxels that lie on this curve. To create the tube, all voxels in a sphere centered at each voxel on the spline are assigned to the artery segment. The algorithm ensures that the resultant artery has 26-connectedness, so there are no holes or kinks. This is especially useful for sharp turns, such as when the artery becomes tortuous.

The method that identifies the voxels on the curve is not incremental, as found in the literature but results from a parameterized form of the curve equation at a set number of

sample points. The accuracy of the curve increases with more sample points. A discrete 26-connected spline algorithm (35) connects these curve points. The user has control over the number of sample points and the radius of the arteries.

To further take advantage of the new more-powerful computer, investigators adjusted parameters in the patient model development process creating a patient model with more resolution. Prior to the extension, the male patient model contained approximately 25,000 polygons. The current patient model is composed of approximately 69,000 polygons, taking full advantage of the new simulation computer.

Patient Model – The NLM Female

Investigators at OSC evaluated the NLM female data set and found that it was somewhat better than the male data in that there were three times as many slices. The female image resolution is isotropic, i.e., .33mm x .33mm (in image resolution) with a slice thickness of .33mm. Recall that although the male images were similar to the female, 1 millimeter separated the slices. While the increased resolution makes the female data better than the male, the female data exhibits many of the same problems caused by preparation, as does the male. Swelling due to freezing closed the nasal passage, necessitating extensive manual alteration of the data. Nonetheless, investigators chose to use the female data set because the cost and schedule risk to collect computer tomography (CT) images at a similar resolution was too great. Furthermore, color images provide better differentiation among internal organs than does CT making it easier for developers to identify structures such as the muscles and fat around the eye.

In a manner similar to development of the male model, investigators created the female model from the high-resolution (2048x1216) cryosection images. These images provide the highest spatial resolution, as compared to the magnetic resonance (MR) or CT images. Investigators converted the input images to SGI Red Green Blue (RGB) format. Using SGI's *subimg* utility, they cropped away a large portion of the surrounding background and extraneous anatomy yielding more manageable 350 x 500 images. Images from level #1250 (just above the frontal sinuses) to #1530 (just below the hard palette) encompass the nasal cavity and surrounding paranasal sinuses. This isotropic data set precludes the need to interpret between slices as is necessary in processing the male data. Figure 9 shows the female isosurface model. With the increase in spatial resolution, the female image set is larger than the male data set.

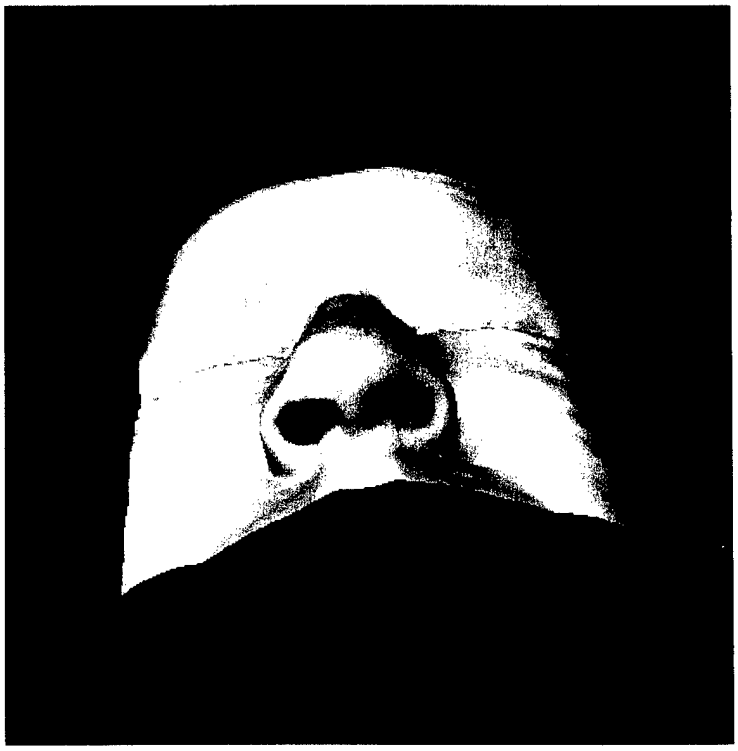


Figure 9. Female Skin Isosurface Model

To create the female model investigators used the VTK Version 2.1. New implementation of scripts drastically reduces the amount of disk space and processing time needed to generate the surface-based patient model. Version 2.1 introduces multiprocessing. The pipeline operations execute using up to 8 processors at the same time. It now takes around 5 hours to process the entire female model on a 24-processor Origin2000 with 250 MHz R10000s. Also, for the female there are only two pre-processing steps. One converts the image masks from SGI RGB format to portable any map (PNM) format, which VTK uses. The other reorders the image slices from top to bottom. Figures 10 and 11 illustrate how VTK components combine to form the image-processing pipeline.

Using the process described in figures 10 and 11, investigators created the female patient model comprising the uncinate, maxillary ostium, ethmoid bulla, anterior ethmoid artery, carotid artery, ophthalmic artery, optic nerve, sphenoid sinus, skull base, lamina papyracea, nasal passage, ethmoid cells, frontal recess, middle turbinate, frontal sinus, maxillary sinus, skin, agger nasi cell, periorbita, adenoid pad, choanae, eustachian tube, inferior turbinate, medial rectus muscle, nasolacrimal duct, nasopharynx, periorbital fat, pituitary gland, septum, sphenoid ostium, superior oblique muscle, and superior turbinate (figure 12).

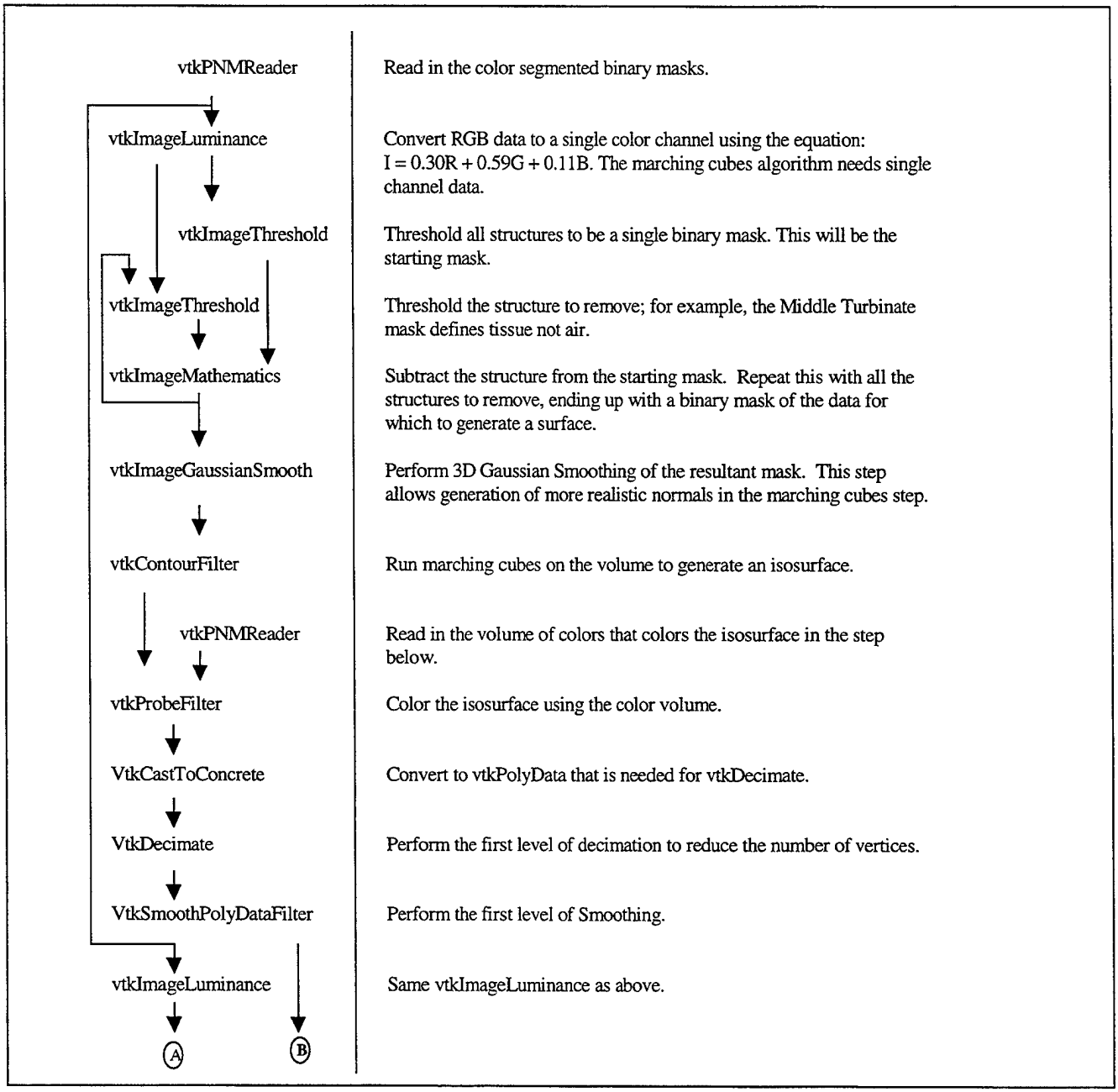


Figure 10. First Half of VTK Pipeline

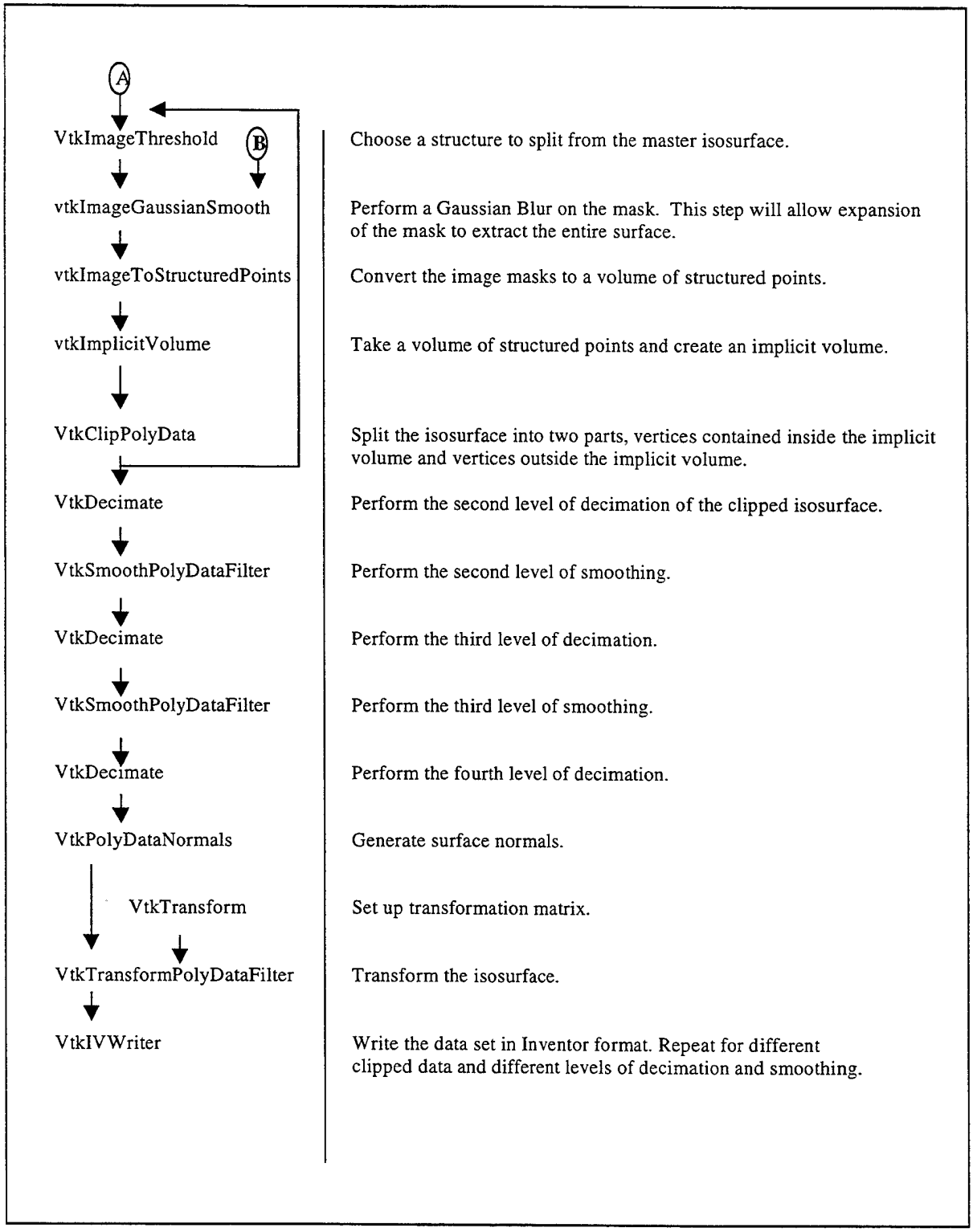


Figure 11. Second Half of VTK Pipeline

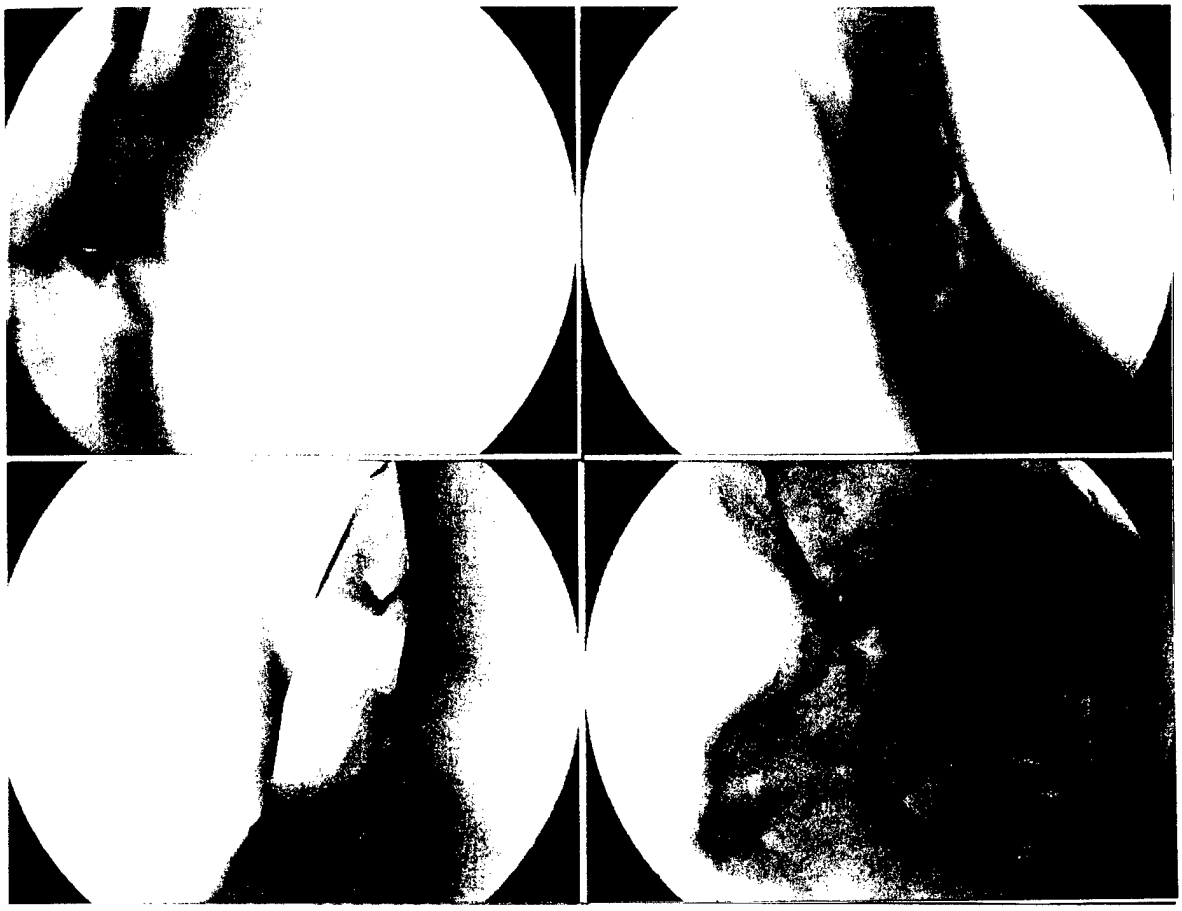


Figure 12. Left Top: Sphenoid Ostium; Right Top: Middle Turbinate Left Bottom: Ethmoid Bulla Behind Uncinate; Right Bottom: Eustachian Tube, Adenoid Pad, and Nasopharynx

Although the cryosection resolution was high (.33mm), the task of visualizing the nasal cavity directly through an endoscope required extremely high resolution to provide the appearance of smooth, contiguous surfaces as found in the operating room. Even with an acquisition of .33 mm, by providing the capability of viewing structures at extreme magnification during the simulation, discontinuities in the surface were visible. Figure 13 shows one such discontinuity along the base of the inferior turbinate.

To mitigate the visual artifacts presented in the simulation, both volume and surface techniques require an increased amount of data. In order to assure real-time performance, algorithmic techniques smooth irregularities and decimate the surface to maintain a tractable number of polygons. However, these polygons create a simplified surface description that has limited surface detail compared to human sinuses.

Figure 13. Base of Inferior Turbinate

Ideally, to provide patient variance it would be useful to employ a direct route for patient modeling (36). Inclusion of clinical data would expand the utility of the system from resident training to preoperative assessment and on to surgical planning. However, typical imaging acquisitions do not provide an isotropic acquisition and rarely achieve a resolution as precise as .33mm. The time required to preprocess patient-specific data prevents the common use of this method in clinical practice. However, new technologies continue to emerge. These technologies will provide the increased image resolution that simulation requires. New imaging techniques will become a source of investigation for improved realism and patient-specific data.

The Simulator – Hybrid Rendering

Today, simulator manufacturers virtually all choose to render polygonal models. Stredney and others recently compared the emerging volume rendering approaches with surface rendering (37). Surface representation provided superior visual cues for tasks such as navigation through the nasal cavity and positioning for surgical interaction with target areas. Volume rendering continues to be computationally intensive causing excess latency and a

lack of visual clarity, which are severe impediments in surgical training simulations. However, volume-based rendering has a distinct advantage in dissection, particularly in the case of ethmoid cells where the cells appear to float in space with a surface model. The fuzzy appearance of volume rendered images, although a liability in most surfaces, would not be a liability in dissected structures since dissection tends to leave a bloody surface anyway. Volume rendering of the entire sinus area from a perspective that covers the screen (up close) is not possible in real time with even the most powerful image generators. Lockheed Martin commissioned Ohio State University to study a hybrid approach that would, for example, use voxels to represent the ethmoid air cells and surfaces to represent structures that do not require extensive dissection. Combining volume and surface rendering might yield a hybrid approach that employs the strengths of both.

Appendix C contains the complete report. The study found that current volume rendering techniques are pixel fill limited and even small volumes when viewed up close, as through an endoscope, can exhaust the capabilities of the image generation hardware. Simulation could not tolerate the resultant loss of real time.

Evaluation

This section summarizes the findings of the simulator evaluation in an abbreviated fashion. Appendix D discloses the complete evaluation results.

To evaluate the effectiveness of the simulator as a training tool, investigators systematically measured several possible indicators of surgical skill:

- Operating room performance assessed by an attending staff proctor
- Blind analysis of endoscopic video acquired during these procedures
- Independent paper-and-pencil tests of components of surgical competency
- Simulator scores on various training exercises
- Time spent on the simulator
- Endoscopic sinus surgery experience

Ten junior and senior residents at Madigan Army Medical Center served as subjects, some of whom had prior training with the simulator. Limited by the number of available residents, researchers performed quasi-experimental studies rather than full factorial design experiments. As such, these should be considered "pilot" studies that illustrate potentially useful methods for evaluating surgical simulators, as well as suggest possibilities for further exploration.

Evaluation – Operating Room Performance Ratings by Staff Proctors

The operating room performance ratings address question 1 in the proposal: What is the impact of experience with the simulator on actual operating room performance? Four residents from the Madigan Army Medical Center Otolaryngology program participated in this study. Two had extensive simulator exposure; two had none. None had prior hands-on operating room sinus surgery experience. The evaluation of the operating room performance for the residents consisted of two components: real-time ratings by the attending staff proctor, and blind ratings of videotapes captured through the endoscope. This section discusses the ratings by staff proctors. The following section discusses the videotape ratings.

Although researchers attempted to select equivalent cases, the normal variability across patients is such that some of these individual components were not relevant to all cases. In addition, the attending proctor would sometimes perform the procedure on one side or would sometimes take over the instruments during particularly troublesome tasks. Consequently, investigators normalized for these variations by averaging the scores across relevant items for that particular case.

Three senior staff physicians who had been working together on training at Madigan Army Medical Center and who had similar approaches to teaching sinus surgery procedures served as proctors. Although informal comparisons indicated high inter-rater reliability among the

three proctors, use of three different raters on such a small group of subjects may increase variability in the study and should be formally assessed in future studies.

Examination of bivariate correlation between simulator time and proctor ratings reveals no significant findings. Researchers found that the students' scores from proctor ratings do not vary across students as much as blind videotape ratings of the same procedures (next section). This may be due to two influences. Proctors may perceive that ratings of their own students reflect upon themselves and hence they may be less objective; or, proctors may fear that a critical rating would be discovered by the student, adversely affecting the student-teacher relationship.

Evaluation – Ratings of Videotaped Procedures by Experienced Staff

Ratings of videotaped procedures also address the question: What is the impact of experience with the simulator on actual operating room performance? A panel of four, experienced sinus surgeons blindly evaluated five videos. Four videos captured first-time surgeries for the four junior residents. These were the same surgeries as those rated by proctors in the previous section. The fifth captured an experienced surgeon. The panel rated each video relative to fourteen items. Researchers looked at correlations among the 14 rating measures, and found two sets of variables highly correlated. They appear to represent separate factors as shown in table 4.

Table 4. Correlated Videotape Measures

Factor 1 – Scope and Instrument Control	Factor 2 – Dissection Skill
Image-task alignment	Surgical confidence
Orientation of image	Anterior ethmoidectomy
Proper depth of image	Tool manipulation
Tool-tool dexterity	Maxillary antrostomy
Tissue respect	Uncinectomy
Injection	

The three remaining variables fall out as follows:

Tool selection -- correlated with both of the above factors
Case difficulty -- not correlated with any item
Navigation -- negatively correlated with everything, but no significant correlation

The elements of Factor 1 relate to scope and instrument control, while dissection skill is a common thread among measures in Factor 2. It is interesting to note that surgical confidence is most highly correlated with this second factor, indicating perhaps that confidence comes from experience in successful dissection.

Figure 14 illustrates the mean across-rater ratings for each of these measures organized by subject. Although the chart is an unconventional use of a line graph, it nevertheless clearly reveals relative performance among the subjects. Each line in the chart represents the

performance of each subject across the measures. To help visualize the two major item groupings, the elements of Factor 1 appear on the left of the graph (1 through 6) and Factor 2 on the right (8 through 12) as follows:

1. Injection
2. Orientation of video image
3. Image-task alignment
4. Proper depth of image for task
5. Tool-tool dexterity
6. Tissue respect
7. Tool selection
8. Uncinectomy
9. Anterior ethmoidectomy
10. Maxillary antrostomy
11. Tool manipulation
12. Surgical confidence

Note: Case difficulty and navigation are not plotted.

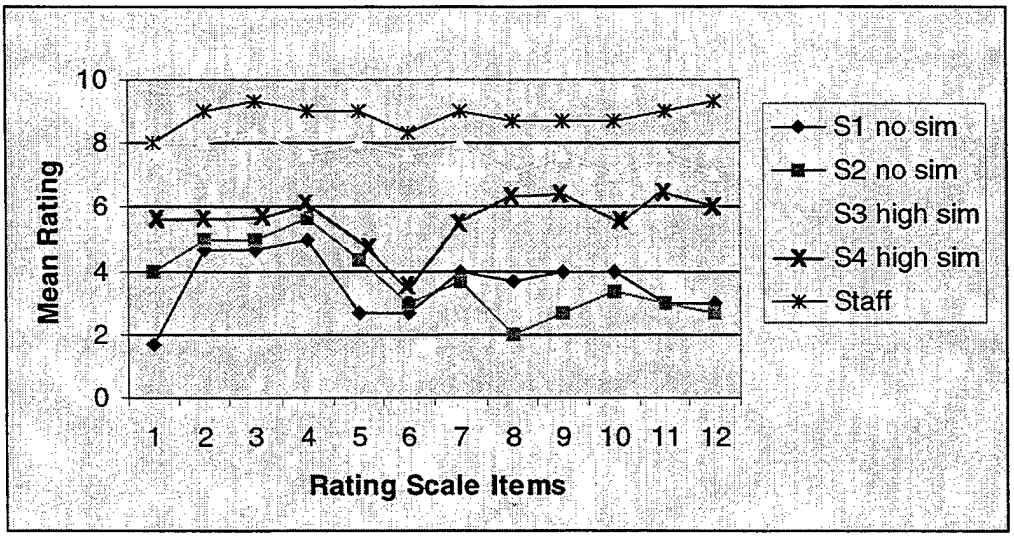


Figure 14. Mean Ratings for the Videotape Measures

The bivariate correlation matrix for these video ratings reveals that all of the above measures are partially intercorrelated except for case difficulty and navigation. This finding for case difficulty is expected, since the raters attempted to normalize their ratings across cases.

The finding of no correlation for navigation is perhaps due to the extreme variability across cases in navigability of the anatomy, and a subsequent confusion by the raters as to what they should rate. Supportive of this conclusion is the finding that there is a significant between-groups analysis of variance effect by rater for navigation only ($p < .009$), out of all the rating measures, indicating low inter-rater reliability for that measure.

While mean rating across items was, as expected, correlated with all of the other components, the strongest predictors of mean videotape rating were tool manipulation ($r = .962$), tool selection ($r = .961$), maxillary antrostomy ($r = .950$), and surgical confidence ($r = .948$).

One can see that simulation training appears to positively affect operating room performance ratings. The figure shows that the two simulation-trained residents performed consistently better than the control group across all measures. However, most likely due to the small number of subjects, these differences were statistically significant for only two items: tool manipulation ($t = 6.00$, $p = .027$) and surgical confidence ($t = 6.96$, $p = .020$). These results are, nevertheless, very promising and suggest that simulation training does indeed have a positive transfer of training effect.

Analysis of variances in the videotape ratings reveals significant differences among the subjects for all but proper depth of image, case difficulty, navigation, and orientation of video image. In other words, these four measures did not significantly discriminate among subjects, but all of the other measures did.

Evaluation - Relationships Among Operating Room Performance Scores

Table 5 presents a matrix of bivariate correlations among the primary measures of interest for the four residents for whom researchers recorded and rated their first operations. The measures are Prior Simulation Time, Mean Operating Room Rating (by the staff proctor), and Mean Video Item Rating (for initial procedure). Table 5 suggests that simulator experience could be a strong predictor of first-time operating room performance as determined by rating videos. This result approaches but does not achieve significance ($r = .911$, $p < .1$) probably due to the small number of subjects. This result clearly points the way to future research, indicating a need for additional subjects.

That the operating room rating does not show a strong correlation with either of the other measures may be due to variance in rating among the three proctors. Only the attending proctor rated each subject. Future studies should include a proctor who is unaware of the student's simulation experience. Additionally, due to the pace of student development and the availability of suitable cases, the proctor ratings were taken at various intervals spanning one year.

Table 5. Correlations among Measures Frozen at the Time of a Resident's First Surgery

	Prior Simulation Time	Mean Operating Room Rating	Mean Video Item Rating
Pearson Correlation			
Prior Simulation Time	1.000	.689	.911
Mean Operating Room Rating	.689	1.000	.804
Mean Video Item Rating	.911	.804	1.000
Significance (2-tailed)			
Prior Simulation Time	.	.311	.089
Mean Operating Room Rating	.311	.	.196
Mean Video Item Rating	.089	.196	.

Evaluation - Transfer of Training to Components of Surgical Competency

The primary goal of the surgical competency study addresses question 2 in the proposal: How does experience with the simulator affect individual components of surgical competence? The evaluation team, led by Dr. Edmond, determined these components to be:

-
- Paranasal sinus anatomical knowledge (basic, complex, and anomalous)
- Endoscopic view identification using different scope angles, including recognition that a particular view is ambiguous
- Procedural knowledge (both general and detailed for all simulated procedures)
- Instrument knowledge (identification, function, and appropriate selection)
- Visual spatial orientation (endoscopic)

No standards exist for evaluating components of surgical competency. To develop these measures, the evaluation team consulted experienced surgeons, who judged them valid for measuring essential skills needed to perform endoscopic sinus surgery.

Investigators summed each resident's component scores to calculate a cumulative competency measure. Of particular interest to this evaluation is that the three subjects with the highest overall scores on the competency evaluation also had 3 of the 4 highest cumulative simulation times. These results suggest a positive impact of simulator experience on surgical competency, as well as a need for further, more extensive research.

Though these studies are a valuable contribution to the field of surgical competency evaluation, future studies should incorporate test materials generated from real endoscopic

footage rather than simulated footage, to more directly address transfer of training to the clinical scenario.

Evaluation - Relationships Among Simulator Measures and Surgical Competency

Upon examination of the bivariate correlations among surgical competency measures and the subjects' scores on the simulator, one finds that overall competency test scores are significantly correlated with overall novice trial scores ($r = .875$, $p < .01$), but *not* with intermediate or advanced model total scores ($r = .228$ and $.123$). Table 6 provides these correlations.

In addition, several competency test sub-scores are significantly correlated with novice trial score: anatomical recognition ($r = .915$, $p < .01$), endoscope selection ($r = .769$, $p < .05$), and procedural awareness ($r = .956$, $p < .01$). This is perplexing since the novice model contains no anatomy to teach anatomical recognition and endoscope selection. Furthermore, the abstract novice model teaches only a high-level outline of the procedure.

Tool identification, while not significantly correlated with *overall* novice trial score, was significantly correlated with some of its components: novice navigation accuracy ($r = .774$, $p < .05$), novice dissection time ($r = -.799$, $p < .05$), and novice dissection score ($r = .805$, $p < .05$). Spatial awareness was not significantly correlated with any of these simulator performance measures, but it is interesting to note that is was more strongly correlated with scores on the intermediate and advanced models than on the novice model. This makes sense because the intermediate and advanced training scenarios contain anatomical models.

The intercorrelations among simulator scores for this group show that novice scores were predictive of intermediate scores, but *not* of advanced scores. Intermediate scores were predictive of advanced scores. The best novice trial predictor of both intermediate and advanced scores was novice *hazard* score, even though novice hazard scores were not significantly correlated with any other novice score or with any competency score. This perhaps suggests that the most discriminating performance factor in the novice model is the ability to avoid hazards, a skill that experienced surgeons would have acquired over time in the operating room.

Table 6. Correlations Among Competency Measures Taken Near the End of the Study

	Sinus Surgery Exp.	Total Sim. Time	Nov. Trial Score	Int. Trial Score	Adv. Trial Score	Comp. Score
Pearson Correlation						
Sinus Surgery Experience	1.000	.157	.698	-.188	-.264	.655
Total Simulator Time	.157	1.000	**.810**	.729	.624	.561
Novice Trial Score	.689	**.810**	1.000	.681	.592	**.875**
Intermediate Trial Score	-.188	.729	.681	1.000	**.966**	.228
Advanced Trial Score	-.264	.624	.592	**.966**	1.000	.123
Competency Score	.655	.561	**.875**	.228	.123	1.000
Significance (2-tailed)						
Sinus Surgery Experience	.	.665	.054	.721	.613	.056
Total Simulator Time	.665	.	**.015**	.100	.186	.116
Novice Trial Score	.054	**.015**	.	.136	.215	**.010**
Intermediate Trial Score	.721	.100	.136	.	**.002**	.712
Advanced Trial Score	.613	.186	.215	**.002**	.	.844
Competency Score	.056	.116	**.010**	.712	.844	.
Number of Subjects						
Sinus Surgery Experience	10	10	8	6	6	9
Total Simulator Time	10	10	8	6	6	9
Novice Trial Score	8	8	8	6	6	7
Intermediate Trial Score	6	6	6	6	6	5
Advanced Trial Score	6	6	6	6	6	5
Competency Score	9	9	7	5	5	9

Evaluation - Effectiveness of Simulator Features

The simulator feature study addresses question 3 in the proposal: What is the effectiveness of the various training aids and protocol methods incorporated into the simulator? Seven of the 10 residents and 3 of the 5 staff completed a post-test questionnaire following their simulation training. The Midterm Report contains a copy of the questionnaire. Its general objective was to assess the perceived usefulness of each of the components of the system, as well as the perceived training utility of the system as a whole. In addition, those subjects who had participated in the Phase 1 evaluation of Version 1.2 evaluated the changes to the system incorporated in Version 2.0.

Although the small pool of available residents did not permit a full factorial design analysis of each component of the training system, residents once again provided subjective evaluations of their pedagogical utility. This contingent of residents indicated the following general assessments:

- Overall, participants found the simulator extremely useful for surgical training.
- They found the anatomical model highly realistic, but several subjects suggested that the major anatomical landmarks and dissectable regions should be more distinct for training purposes.
- Anatomy of lamina papyracea and ethmoid cells needs improvement.
- Movement of the scope and surgical instruments was not as smooth as in the real procedure.
- Haptic feedback could be better and would also be useful on the endoscope itself, but participants preferred to keep the haptic feedback on the surgical instrument if given a choice.
- Some participants preferred to turn forces off during dissection of the anterior ethmoid cells, due to restricted access (a problem, subsequently corrected, with the software caused inappropriate forces in the region of the ethmoid cells).
- Participants found the simulator most useful for teaching anatomy, and for training hand-eye coordination, navigation skills, elementary dissection skills, and selection of appropriate instruments.
- Participants desired a greater depth of field.
- Participants preferred a seated operative posture to the current standing posture.
- Participants found the ability to slip through the virtual anatomy with the endoscope disconcerting.
- Participants judged voice recognition to be useful, but some had difficulty activating the simulator with their voice inputs.
- Participants experienced a slight delay in the update of the graphics after moving the instrumented endoscope.
- Some expressed the desire to move the endoscope closer to the tool to get a more comfortable view of the anatomy.

These results are consistent with findings in the Midterm Report and are not discussed in further detail here.

CONCLUSIONS

Developers from the Ohio Supercomputer Center, Immersion Corporation and Lockheed Martin substantially improved the sinus surgery simulator. The new computer processes nearly three times as many polygons in half the time. Patient models contain numerous additional anatomical components and comprise almost three times the number of polygons. New pathological patient models along with a new virtual patient, based on the NLM female data set, increase the richness of training scenarios. The new virtual instructor software recognizes student voice commands and provides expanded guidance to the students. The improved haptic system is more reliable, maintainable, and delivers a modestly improved haptic sensation.

The evaluation team, from Madigan Army Medical Center and the Human Interface Technology Laboratory, addressed the transfer of training issue. They developed a set of assessment methods, including real-time attending proctor rating scales, endoscopic video assessment scales, and several novel methods of evaluating components of surgical competence. In assessment of the impact of simulation experience on operating room performance, researchers encountered moderate difficulty generating the appropriate cases for evaluation of junior residents, for whom the biggest impact would be expected. In addition, control of operative exposure and experience across residents proved to be an unexpected challenge. Nonetheless, a number of promising trends emerged from the data:

- Two somewhat independent factors emerged from the rating scale used to perform blind assessments of initial operating room performance: scope and instrument control, and dissection skill. Surgical confidence correlates highly with the second factor, indicating perhaps that confidence comes with successful dissection experience.
- Simulation training appears to positively affect initial operating room performance as judged by senior surgeons rating anonymous videotapes of those procedures. The strongest effects are for tool manipulation and surgical confidence.
- Researchers developed several new assessment techniques to measure components of surgical competency. These are a potential contribution to the field of surgical competency evaluation. Future studies, however, should use test materials generated from real endoscopic footage rather than simulated footage, to more directly address transfer of training.
- Results from a group of nine residents followed a consistent trend that suggests transfer of training from the simulator to components of surgical competency. The component most strongly correlated with prior simulation experience is procedural awareness.
- Novice trial score on the simulator is a significant predictor of overall competency. The novice dissection score is the strongest predictor of surgical competency.

Given the limited availability of residents, these results establish only a framework for evaluating transfer of simulation training to operating room performance and to aspects of surgical competency as it relates to endoscopic sinus surgery. Objective measures, generated for evaluation of transfer and competency, and protocols developed to control for extraneous variables, may prove valuable to future studies. In addition to the findings derived from the pilot studies detailed in this report, these methods offer a contribution to the field of medical simulator evaluation.

Researchers will continue to seek answers to complex questions related to transfer of training. These evaluations are important in that they focus the development of this and future simulators to best serve the training needs of the medical community, and they document the value of the simulator relative to its purpose. However, the harshest judgment of value comes at the hands of the marketplace. The sinus surgery simulator generates considerable interest and favorable reaction when shown in venues, such as the annual meeting of the American Academy of Otolaryngologists/Head and Neck Surgeons, where clinical experts experience its capabilities. Yet, there is no more piercing question than the one posed by senior management and oft heard on the exhibit floor, "Have you sold one?" The Lockheed Martin team had hoped, by the close of this cooperative agreement, to answer that question with an unequivocal "Yes!" Nevertheless, team members report that a commercial company has agreed in principle to purchase licenses from Lockheed Martin and build several simulators for augmentation of continuing education courses held at major universities across the country. In the very near future, the Lockheed Martin team hopes to move on to the next marketplace evaluation question, "How many have you sold?"

REFERENCES

(1) Caird J, "Persistent Issues in the Application of Virtual Environment Systems to Training," *IEEE Computer Society: Human Interaction with Complex Systems* Vol. 3 (1996) 124-132.

(2) Cover S, et al, "Interactively Deformable Models for Surgery Simulation," *IEEE CG&A*, 13, 6, (1993) 68-75.

(3) Hon D, "Ixion's Laparoscopic Surgical Skills Simulator," *Medicine Meets Virtual Reality 2* (1994).

(4) Merril J, "Virtual Reality in Surgery and Medical Education," *Medicine Meets Virtual Reality 2* (1994).

(5) McGovern KT and McGovern LT, "The Virtual Clinic, A Virtual Reality Surgical Simulator," *Medicine Meets Virtual Reality 2* (1994).

(6) Chen D, Rosen J, Zeltzer D, "Surgical Simulation Models: From Body Parts to Artificial Person," *Proc* IMAGE VI Conference (July 1992).

(7) Pieper S, et al, "A Virtual Environment System for Simulation of Leg Surgery," *Stereoscopic Displays and Applications II, SPIE*, (1991).

(8) Peifer J, et al, "Virtual Environment for Eye Surgery Simulation," *Medicine Meets Virtual Reality 2* (1994).

(9) Sagar MA, Bullivant D, Mallinson GD, Hunter PJ, "Virtual Environment and Model of the Eye for Surgical Simulation," *Computer Graphics Proceedings* (1994) 205-212.

(10) Pieper S, "CAPS: Computer-Aided Plastic Surgery," Doctoral Thesis, MIT (1989).

(11) Baillie J, et al, "The Future of Endoscopy Simulation: A Duke Perspective," *Endoscopy* (1992) 24.

(12) Barde C, "Simulation Modeling of the Colon," *First International Symposium on Endoscopy Simulation*, World Congresses of Gastroenterology, Sydney (1990).

(13) Gillies D, Haritsis A, Williams C, "Computer Simulation for Teaching Endoscopic Procedures," *Endoscopy* (1992) 24.

(14) Poon A, Williams C, Gillies D, "The Use of Three-Dimensional Dynamic and Kinematic Modelling in the Design of a Colonoscopy Simulator," *New Trends in Computer Graphics*, Springer Verlag (1988).

(15) Good ML, Gravenstein JS, "Training for Safety in an Anesthesia Simulator," *Seminars in Anesthesia* (1993) 12, 4.

(16) Bostrom M, Singh SK, Wiley CW, "Design of An Interactive Lumbar Puncture Simulator with Tactile Feedback," *IEEE Annual Virtual Reality Symposium* (1993) 429-435.

(17) Stredney D, et al, "A Virtual Simulation Environment for Learning Epidural Anesthesia," *Medicine Meets Virtual Reality 4*, IOS Press (1996).

(18) Dawson J, et al, "An Interactive Training/Simulator or Interventional Radiology," *Vascular and Interventional Radiology*, 7 (1, Suppl 2): (1996) 132-136.

(19) Wickham JEA, "Minimally Invasive Surgery: Future Developments," *The British Medical Journal*, Vol 308(6922) (1994) 193-196.

(20) Savitsky MS, LeDonne DM, "Maximizing the Mission of Medical Readiness in a Joint Environment: A Systems Model," NMED (May-June 1995).

(21) Koehler RH, Smith S, et al, "Triage of American Combat Casualties: The Need for Change," *Military Medicine* Vol 159 (August 1994).

(22) Nguyen D, "Mass Casualties on the Modern Battlefield: A View from the 1st Armored Division (US)," *Military Medicine* Vol 159 (November 1994).

(23) General Accounting Office, "Health Care: Readiness of U.S. Contingency Hospital Systems to Treat War Casualties," GAO/T-HRD-92-17 (March 25, 1992).

(24) Department of Defense, DoD Instruction 1322:24 Military Medical Readiness Skills Training (December 20, 1995).

(25) Office of The Chief of Staff, Army; Director, Louisiana Maneuvers Task Force, Force XXI, America's Army of the 21ststst Century (1995).

(26) Satava R, "Doorway to the Future," *Medicine Meets Virtual Reality: Global Healthcare Grid* (1997).

(27) Hoffman HM, Murray M, Irwin A, McCracken T, "Developing a Virtual Reality-Multimedia System for Anatomy Training," *Medicine Meets Virtual Reality 4* (1996).

(28) Robb RA, "Virtual Endoscopy: Evaluation Using the Visible Human Datasets and Comparison with Real Endoscopy in Patients," *Medicine Meets Virtual Reality: Global Healthcare Grid* (1997).

(29) Johnson R, Bhoyrul S, Way L, Satava R., McGovern K, Fletcher JD, Rangel S, Loftin RB, "Assessing a Virtual Reality Surgical Skills Simulator," *Medicine Meets Virtual Reality 2* (1994).

(30) Hon D, "Medical Reality and Virtual Reality," *Medicine Meets Virtual Reality 4* (1996).

(31) Kaufmann DM, Bell W, "Teaching and Assessing Clinical Skills Using Virtual Reality," *Medicine Meets Virtual Reality: Global Grid* (1997).

(32) Rosenberg L, "A Force Feedback Programming Primer," Immersion Corporation (June 1997).

(33) Billinghurst M, et al, "Expert Surgical Assistant, An Intelligent Virtual Environment with Multimodal Input," *Health Care in the Information Age* (1996).

(34) Edmond CV, Heskamp D, Sluis D, Stredney D, Sessanna D, Wiet GW, Yagel R, Weghorst S, Oppenheimer P, Miller J, Levin M, and L Rosenberg (1997), "ENT

Endoscopic Surgical Training Simulator," *Proceedings of Medicine Meets Virtual Reality*, KS Morgan, et al (Eds.) IOS Press, Amsterdam (1997) 518-528.

(35) Kaufman A, Cohen D, and Yagel R, "Volumetric Graphics," *IEEE Computer* Vol. 26 No.7 (July 1993) 55-64.

(36) Edmond CV, Wiet GJ, and B Bolger, "Virtual Environments: Surgical Simulation in Otolaryngology" *Computers in Otolaryngology, Otolaryngologic Clinics of North America* Vol. 31 No. 2 (April 1988) 369-381.

(37) Stredney D, Wiet GJ, Yagel R, Sessanna D, Kurzion Y, Fontana M, Shareef N, Levin M, Martin K, and Okamura A, "A Comparative Analysis of Integrating Visual Representations with Haptic Displays," *Proceedings of Medicine Meets Virtual Reality*, KS Morgan et al (Eds.) IOS Press Amsterdam, 20-26.

APPENDIX A

RELATED PUBLICATIONS

and PRESENTATIONS

(October 1997 through October 1998)

Related Publications and Presentations

Who	When	Where	Title/Description
Edmond, C	October 1997	American College of Surgeons Meeting Chicago, IL	Presented progress on the ENT Surgical Simulator.
Edmond, C	November 1997	Society of University Otolaryngologists Annual Meeting, Washington, DC	Presented progress on the ENT Surgical Simulator.
Heskamp, D Rosemarino, A Miller, J Gurcak, A	November/ December 1997	Radiological Society of North America Chicago, IL	Simulator demonstrated in the Silicon Graphics Booth.
Weghorst, S	January 1998	Medicine Meets Virtual Reality San Diego, CA	Validation of the Madigan ESS Simulator
Stredney, D	January 1998	Medicine Meets Virtual Reality San Diego, CA	A Comparative Analysis of Integrating Visual Representations with Haptic Displays
Edmond, C Heskamp, D	February 1998	Computers & Otolaryngology: Practical Integration of Advanced Technologies. Baylor College of Medicine Houston, TX	Simulator used as an integral part of the continuing medical education didactic session.
Heskamp, D Edmond, C	March 1998	Nationwide Television	Video recorded at Baylor (above) transmitted by satellite to 300 television markets. Aired to potential audience of 30 million viewers.
Edmond, C	January 1998 - March 1998	University of Washington Seattle, WA	1) Presentation to the School of Education. 2) Presentation to the Department of Surgery.
Edmond, C Wiet, G Bolger, B	April 1998	"Otolaryngology Clinics of North America" Volume 31, #2	Virtual Environments: Surgical Simulation in Otolaryngology.
Edmond, C Heskamp, D	May 1998	American Pediatric Society and Society for Pediatric Research Annual Meeting New Orleans, LA	Live demonstration / explanation of simulator at the state-of-the-art Plenary Session.
Wiet, GJ Rudman, DT Stredney, D Sessanna, D Yagel, R Heskamp D Edmond, CV	May 1998	101st Annual Meeting of the Triological Society Palm Beach, FL	Functional Endoscopic Sinus Surgery Training Simulator
Heskamp, D Gurcak, A Stredney, D Sessanna, D	June 1998	Ohio State Annual Department of Otolaryngology Alumni Symposium Columbus, OH	Demonstration of simulator at the conference.
Heskamp, D Edmond, C Weghorst, S	July 1998	National Board of Medical Examiners OTTAWA Conference Philadelphia, PA	Simulation in Medicine: Experience from an Endoscopic Sinus Surgery Simulator

Kuppersmith, R	September 1998	American Academy of Otolaryngologists Annual Meeting San Antonio, TX	Presentation on simulation including video of simulator.
Heskamp, D Miller, J Gurcak, A Edmond, C Mesaros, G	September 1998	American Academy of Otolaryngologists Annual Meeting San Antonio, TX	Demonstration of simulator at the conference.
Heskamp, D Gurcak, A Miller, J	October 1998	Conference on Technology in Trauma, Technology in Surgery – Management in the Information Age Cleveland, OH	Demonstration of simulator at the conference.

The authors anticipate that additional papers and presentations are forthcoming utilizing results published herein.

APPENDIX B

PERSONNEL RECEIVING PAY

FROM THIS EFFORT

Consultants

Charles Edmond Jr. MD

Lockheed Martin Tactical Defense Systems

David B. Heskamp, Jeffrey G. Miller,
J. Prulhiere, R. Baldwin, M. Slevin, S. Watson, A. Gurcak, S. Smeiles-Xenias, T. Stockdale,
B. Miner, J. Strott, J. Hough, R. Scott, R. Grimaldi, L. Carroll, V. Pearson, A. Angel, H. Jordan

Ohio Super Computer Center and The Ohio State University

Don Stredney, Dennis J. Sessanna,
Gregory J. Wiet MD, Roger Crawfis PhD, Naeem Shareef

Human Interface Technology Laboratory

Suzanne Weghorst, Chris Airola, Peter Oppenheimer

Immersion Corporation

Michael D. Levin,
Ken Martin, Dave Bailey, Ryan Bruneau, Sam Serna

APPENDIX C

HYBRID RENDERING APPROACH STUDY

Merging Surfaces with Volume Data

Don Stredney,[1] Dennis J. Sessanna,[1] Gregory J. Wiet,[2,3] Roger Crawfis,[4] and Naeem Shareef [4]

[1] Ohio Supercomputer Center, Columbus, Ohio

[2] Department of Otolaryngology, The Ohio State University Hospitals, Columbus, Ohio

[3] Pediatric Otolaryngology-Head and Neck Surgery, Children's Hospital, Columbus, Ohio

[4] Department of Computer and Information Science, The Ohio State University, Columbus, Ohio

Introduction/Background

From the beginning of this project, we decided to build the anatomical model from a surface representation, i.e., polygon mesh, for the Functional Endoscopic Sinus Surgery (FESS) simulation system. The model was based on two requirements for the system: real-time rendering and realistic image quality. The current surface-based system is very close to satisfying these two goals. However, after achieving realistic and real-time performance for navigation, the user subsequently required effective and realistic interaction with the anatomical model. Unfortunately, a surface representation is hollow, representing only surface topology. Thus, when operations such as digging, drilling, or cutting are desired, a very visible hole will expose surfaces behind those being worked upon, producing an unrealistic environment. Trying to reconstruct in real time a closed surface after such operations is difficult and time consuming.

However, these subtractive operations on a volumetric representation are easy to perform on a local per-voxel basis by changing the opacity of affected voxels to transparency. Note that our surface model was constructed from volume data. Our goal was to fill our empty surface model with the original volume data. When the user is not manipulating objects, the superior visual effects of the surface model are preserved. When cutting is performed on the surfaces, then appropriate surface elements will be removed and the volume representation will be visible. This effect requires a hybrid renderer that merges surface rendering and volume rendering.

The inherent difficulty in rendering both surface and volume data simultaneously is that a renderer must deal with two types of graphics primitives, i.e. polygons and voxels. Raycasting presents one solution because rays can be easily intersected with the surface and the volume can be adequately resampled. The problem is that this method is slow and graphics accelerators for this application are not currently available. On the other hand, graphics accelerator pipelines project primitives. In this case, when the surface model is opaque, then all that is required is to render the volume in front of the surface. In this way, the surface model may be rendered in a first pass and then the volume rendered in sorted order for correct compositing. For example, the surface may be rendered to a z-buffer, and the volume rendered in a front-to-back or back-to-front order, testing the z-buffer. When the

surface is not opaque then the difficulty stems from the fact that the view-dependent depth ordering of all graphic primitives must be preserved to correctly perform compositing. This problem has been addressed in the literature by several authors, including (Max 1990), (Guo 1995), (Albertelli 1997), and (Nielson 1997). An alternative solution we will mention for completeness is that the surface model may be voxelized and then only voxels need to be processed. As mentioned before, a purely volumetric solution is currently not feasible to satisfy the goals of real-time rendering with realistic image quality (Yagel et al., 1996), (Wiet et al., 1997, 1998), (Stredney et al., 1998).

Methods

Trying to achieve real-time volume rendering using current technology requires either extra hardware capabilities, such as parallel machines, or reduction in the dataset, resulting in a loss of image quality. We have explored a method that shows promise toward merging volumes and surfaces in real-time because it utilizes the same graphic accelerator hardware as used by our current surface-based system. This method uses the 3D texture hardware extension to perform volume rendering. The volume dataset is realized as a 3D texture and is visualized via slicing polygons that are colored via the well known solid texturing method. We now describe the pros and cons of this method in this project.

3D texturing technology is available on various platforms, including a number of workstations from Silicon Graphics, Inc. (SGI). Volume rendering of regular grids using this technology is straightforward. Polygons, placed parallel to the view plane and along the view direction, are sampling agents that are composited together in depth-sorted order to produce the final image. The volume is loaded into texture memory and the slicing polygons are spaced evenly along the eye direction, e.g. z-axis, and are mapped to 3D texture space where they are filled. For this filling process, 3D texture coordinates are interpolated and the texture volume is resampled using trilinear interpolation. The resolution of the fill, or rasterization, is determined by the resolution of the screen. We constructed a prototype in our project using the 3D texture-based volume rendering software package called VRP™ (Volume Rendering Primer) available from SGI. This code is written using OpenGL™ and renders surface geometry, called embedded geometry, defined in an Inventor™ scenegraph file format, simultaneously with volume data. The surface data can be shaded with specular highlights. VRP™ treats both data types as separate entities. When the embedded geometry is rendered with the volume, the embedded geometry is projected first. Then, the textured slicing planes are projected in sorted order with respect to each other. If the surface model is opaque then the part of the volume in front of the surface model will be rendered. Areas where the volume samples lie behind the surface will be discarded with the z-buffer test. If the slicing polygons (through the volume) are composited in back-to-front order, the correct result is achieved. If the surface model is semi-transparent, then this method is incorrect because the sorting order is wrong.

The advantage of this method is that the volume is realized as a set of polygons. Thus, it can easily be combined with the surface model, and graphics acceleration hardware is used to quickly render the volume. We are able to load and position, i.e. register, the surface and volume dataset into the scene using affine transformations such as translations, rotations, and scale. As mentioned above, the idea is that the surface model acts as a "shell" to enclose the volume data.

Results

Figure 1: Left: Surface Only. Right: Volume Middle Turbinate (in left field of view)

Volume Size -64^3 of 4 component data (RGBA)
No Iso-surface

Onyx2 IR w/ 64MB texture

(frames/second)		Samples (slices)		
Window Size	150	76	30	10
512x512	4.40	7.80	15.28	28.75
380x380	7.61	13.10	20.00	30.00

Window size - 380x380
Samples – 150 slices

Onyx IR w/ 64MB texture	Octane MXE w/ 4MB texture	Onyx2 IR w/ 64MB texture
7.22 Frames/ Second	7.27 Frames/ Second	7.61 Frames/ Second

Onyx2 IR w/ 64MB tex
Volume Size -128^3 of 4 component data (RGBA)
Window size - 380x380
Samples – 150 slices

Iso- surface – 29550 Triangles	No Iso-surface
4.73 Frames/ Second	5.59 Frames/ Second

Discussion

An initial evaluation of our current prototype shows that this approach is promising; however, we need to overcome some difficulties. The image quality is relatively high in relation to the rendering speed we are able to achieve. There are two levels of sampling for rendering. The first is realized with the screen-aligned slicing planes. Image quality improves as more planes are added, with the additional cost of speed, because more polygons need to be processed. The second level of sampling occurs when the texture is applied to the slicing polygons, where the sampling resolution is determined by screen resolution during rasterization, as mentioned before. There will certainly be a distraction if the difference in image quality between the surface and volume representation is large. Issues that will need to be addressed are the "fuzziness" of the volume representation in relation to the surface model and the striping that will occur at the transitions. The issue of achieving an acceptable frame rate is very much an open problem. A solution may include reducing the volume space of the volumetric objects using bounding boxes, thus reducing the size of the slicing planes and requiring less texturizing, i.e. rasterization. The polygon fill step

may introduce an unacceptable bottleneck as the viewer comes closer to the volume, because the projected slicing polygons will cover a larger portion of the screen. We will need to effectively evaluate the cost of this operation because our application will require such magnified views.

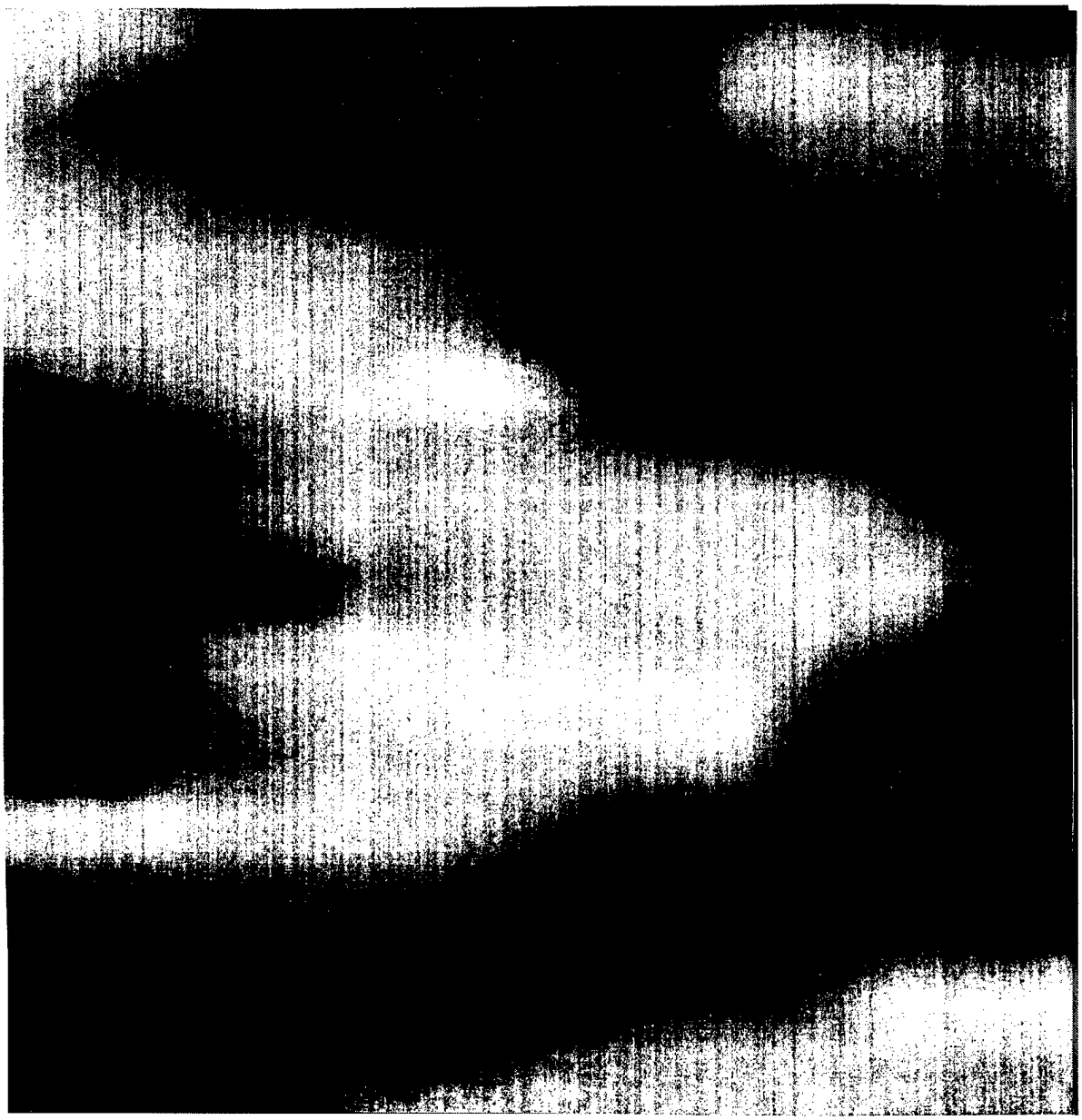


Figure 2. Medial View of Middle Turbinate. When Magnified, Notice Effect of "Fuzziness"

Another limitation is the relatively small size of the texture memory available on current machines when compared with the volume sizes we will need to handle. For example, the largest texture memory space in our lab is 64MB on an SGI ONYX2. So far, we have rendered small size datasets and only one volume at a time. This fact does not limit our

capability to visualize large datasets, but the cache overhead of uploading subvolumes for larger size volumes and multiple volumes still needs to be researched. In addition, technology promises to provide larger memories at lower cost, as has been the case with memory technology over the years.

References

Max N, Hanrahan P, and Crawfis R, "Area and Volume Coherence for Efficient Visualization of 3D Scalar Functions," *Computer Graphics*, Vol. 24, (November 1990) 27-33.

Guo B, "Interval Set: A Volume Rendering Technique Generalizing Isosurface Extraction," *IEEE Visualization '95* (Vis '95), Atlanta GA (Oct. 29 - Nov. 3, 1995) 3-10.

Albertelli G, and Crawfis R, "Efficient Subdivision of Finite-Element and CFD Datasets into Consistent Tetrahedra," *IEEE Visualization '98* (Vis '98), Phoenix AZ (Oct. 19 - Oct. 24, 1997) 213 - 219.

Nielson, GM, Sung J, "Interval Volume Tetrahedralization," *IEEE Visualization '98* (Vis '98), Phoenix, AZ, (Oct. 19 - Oct. 24, 1997) 221-228.

Stredney D, Wiet GJ, Yagel R, Sessanna D, Kurzion Y, Fontana M, Shareef N, Levin M, Martin K, and Okamura A, "A Comparative Analysis of Integrating Visual Representations with Haptic Displays," *Proceedings of Medicine Meets Virtual Reality*, KS Morgan et al. (Eds.) IOS Press, Amsterdam. 20-26.

Wiet GW, Stredney DL, Yagel R, and Sessanna D, "Using Advanced Simulation Technology for Cranial Base Tumor Evaluation," *Computers in Otolaryngology*, Vol. 31 (2) (April 1998) 341-356.

Wiet GJ, Yagel R, Stredney D, Schmalbrock P, Sessanna D, Kurzion Y, Rosenberg L, Levin M, and Martin K, "A Volumetric Approach to Virtual Simulation of Functional Endoscopic Sinus Surgery," *Proceedings of Medicine Meets Virtual Reality 5* (January 1997) Morgan et al., (Eds.), IOS Press, Amsterdam, 167-179.

Yagel R, Stredney D, Wiet GJ, Schmalbrock P, Rosenberg L, Sessanna D, and Kurzion Y, "Building a Virtual Environment for Endoscopic Sinus Surgery Simulation," *Comput. & Graphics*, Vol. 20 (December 1996) No. 6., 813-823.

APPENDIX D

FORMAL EVALUATION

Phase 2 Formal Evaluation of the Madigan Endoscopic Sinus Surgery Simulator

Prepared by:

Suzanne Weghorst
Christopher Airola
Human Interface Technology Laboratory
University of Washington, Seattle, WA

Charles V. Edmond Jr. MD FACS
Maj. Glen Mesaros M.D.
Madigan Army Medical Center, Ft. Lewis, WA

David Heskamp
Jeffrey Miller
Lockheed Martin Tactical Defense Systems, Akron, OH

Table of Contents

ABSTRACT

The Madigan Endoscopic Sinus Surgery (ESS) Simulator is a state-of-the-art procedural simulator developed by a multi-institution team led by researchers at Lockheed Martin. It includes force-feedback instrument and virtual endoscope interfaces developed by Immersion Corporation, and 3-D models of the paranasal sinus anatomy derived from the Visible Human dataset by researchers at the Ohio Supercomputer Center, supplemented by a variety of instructional aids, such as hoops for navigation paths and targets for injection sites. We describe here the results of the final phase of our formal evaluation of the simulator, led by researchers at the Human Interface Technology (HIT) Laboratory at The University of Washington. Formative evaluation efforts throughout the development phase of the project informed the system design and user interface requirements for the simulator. Summative evaluation of Version 1.2 of the system was conducted during the previous phase of the project, with an emphasis on its validity as an ESS simulator. The procedural validity of the simulator was supported by the strong correlation between performance on the ESS simulator and degree of prior ESS experience, and by the subjective evaluations by the experienced ESS surgeons. The primary focus of the current phase of the evaluation effort was on characterizing the degree and type of transfer of training from the simulator to actual operating room performance. Several approaches to assessing transfer of training were developed: real-time proctor ratings of OR performance; blind ratings of endoscopic videotape of initial OR procedures; and paper-and-pencil measures of relevant components of surgical competence. Results from these studies suggest that experience with the Madigan ESS simulator results in learning which transfers to OR performance and surgical competence.

1 Introduction

The use of computer simulation as a training tool for surgical procedures has been motivated by the success of computer simulation flight training as well as the high cost of medical education. In recent years, advances in interactive graphics and virtual reality technology have greatly enhanced our arsenal of instructional tools (Caird, 1996), moving these systems toward more general commercial graphics platforms.

A recent survey of prior research in surgical simulation suggests that each surgical procedure has a unique set of simulation requirements (Edmond et al., 1997). Research thus far includes abdominal laparoscopy (Cover et al., 1993; Hon, 1994; Merril, 1994; McGovern et al., 1994); limb surgery (Chen et al., 1992; Pieper et al., 1991); eye surgery (Peifer et al., 1994; Sagar et al., 1994); plastic surgery (Pieper, 1989); gastrointestinal endoscopy (Baillie et al., 1992; Bard, 1990; Gillies et al., 1992; Poon et al., 1988); anesthesiology (Good et al., 1993); epidural anesthesiology (Bostrom et al., 1993; Stredney et al., 1996); and interventional radiology procedures (Dawson et al., 1996). Much of this recent work in medical simulation has received impetus from the leadership of both DARPA (Satava, 1996) and the Army Medical Command.

Wickham (1994) summarizes the need for these novel and extensive training techniques for endoscopic surgery skills:

> Evaluation of new operative competence is urgently needed because of the rapidity of changes in interventional treatment. Training programmes must be established so that interventionists' training is similar to that of airline pilots. A surgeon or radiologist should not be allowed to treat patients with sophisticated and potentially dangerous instruments without the experience of simulated operations and closely supervised procedural training. Fully equipped training centers should be established with simulator laboratories where interventionists can develop the different hand-eye coordination required for the transition from open to endoscopic techniques . . . The need is urgent: the traditional methods of "see one, watch a video, do one" are completely inadequate preparation for minimally invasive techniques . . . A theoretical evaluation of competence by written or oral examination is totally insufficient to determine whether a clinician has gained the manual ability to carry out complex open or endoscopic surgery.

The motivation for medical training simulators is clear. It is important, however, to ask: Do they work? Is the simulator effective as a training device? What skills do students learn most effectively? Do the skills learned in the simulated training transfer to the operating room? How fast do students learn? How does an hour of simulator time compare to an hour of traditional training methods? Are these simulators currently cost effective? If not, when will they be cost effective for use in medical schools? And how can simulators best be integrated with medical school curriculum?

To answer these questions, formal evaluation of surgical simulation systems is essential. Without answers, one remains uncertain as to whether students are receiving surgical training or merely learning to be good simulator users.

The evaluation study described here has been guided by the work of many others in the field. As Hoffman et al. (1996) suggest, we have included the end users during the formative stages of the simulator design, a vital step in establishing educational goals and curriculum design. In developing our evaluation criteria we have taken into account both objective and subjective considerations (Robb, 1997), relied heavily on the basic surgical proficiency measures of time and accuracy (Johnson et al., 1996), and incorporated much of our evaluation protocol into the simulator itself (Hon, 1996; Kaufmann, 1997).

1.1 Simulator Overview

The Madigan Endoscopic Sinus Surgery (ESS) Simulator uses virtual reality technology, a force feedback (haptic) display, and 3D computer-based anatomy models as a tool to teach a variety of skills needed to perform such surgery.

The heart of the system is a 3D model of the human nasal sinus anatomy derived from the National Library of Medicine's Visible Human Database. Researchers at the Ohio Supercomputer Center (OSC) created segmented surface models of the sinus anatomy from photographic cryosections of the male dataset. The Lockheed Martin team then added stochastically generated surface textures to complete the anatomical model.

This model can be rendered in real time (30-60 frames per second, depending primarily on anatomical position) on a Silicon Graphics (SGI) Onyx System and viewed at NTSC resolution on a standard video monitor, thus simulating the view that a surgeon would see of a video endoscopic display.

To interact with the model and perform the simulated surgery, students in training use a pair of 6-degree-of-freedom input devices developed by Immersion Corporation. One such device represents the video endoscope and the other represents the surgical instrument, such as an injection needle or a forceps (Rosenberg and Stredney, 1996). An external view of an experienced ENT surgeon operating the system can be seen in Figure 1. Note that the position of the surgeon relative to the endoscopic monitor and simulated patient emulates a typical clinical configuration.

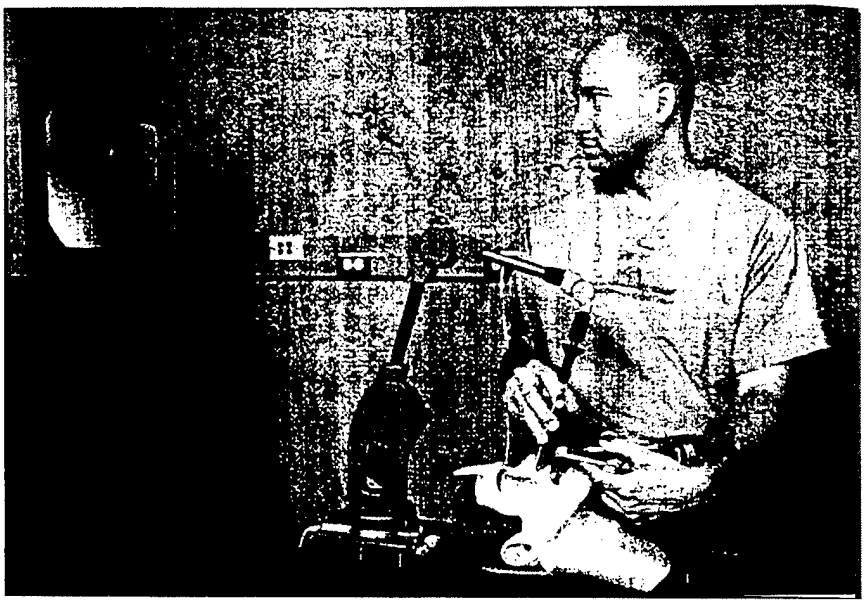


Figure 1. Operative Configuration of the Madigan Endoscopic Sinus Surgery Simulator

As the student manipulates the input devices, the simulator tracks the position and orientation of the devices, updates the positions and orientations of the virtual endoscope and virtual instrument, manipulates the virtual anatomical model accordingly, and displays the resulting virtual endoscopic view on the monitor. In addition, the system tracks the opening of the forceps handle of the instrument input device, as well as a foot switch for activation of the rotating virtual instruments (the microdebrider, round burr, router and bent router). All together, the system measures 14 degrees of freedom of the student's input.

The physical input devices are designed to resemble the feel of an endoscope and forceps and are assembled with a latex replica of a human head. The endoscope input device resides outside the head, while the instrument input device is inserted into the nostril of the latex head and attached to the position tracker inside the head. In addition to displaying the simulated endoscopic view, the system computes the forces that the sinus tissue would apply to the tip of the instrument during surgery and applies the computed force to the instrument input device, via mechanical coupling within the mannequin's head.

The proctor's console provides an interface to the run-time system parameters, along with optional radiographic views of the current anatomy and optional performance feedback summaries for the trainees.

Training Aids

Optional 3D graphical overlays are superimposed on the endoscopic view to provide performance aids for the student. These overlays include a path of circular hoops representing the desired endoscope trajectory, bullseye targets representing the desired injection sites, and text labels identifying anatomical feature landmarks.

The system provides voice audio feedback representing the current status of the surgical procedure, as well as a simulated heartbeat which responds to certain user actions.

Training Tasks

Students using the system are instructed to perform a simulated surgical procedure consisting of three subtasks: navigation, injection and dissection. In addition to archiving the frame-by-frame position of the devices, the system measures the time required to perform these tasks, as well as the accuracy with which they are performed, to generate an evaluation of the student's overall performance.

Three versions of the procedure were developed corresponding to three skill levels: novice, intermediate, and advanced. The primary differences among these procedures are the type of geometric model used and the presence of training aids.

The novice procedure uses a simplified abstract model of the anatomy consisting of injectable targets and a sequence of tool-specific dissectible objectives and corresponding obstacles or hazards. Intermediate and advanced procedures use the more complex surface sinus model generated by OSC. Novice and intermediate procedures use the training aids described above, whereas the advanced procedure is performed without benefit of these aids (to more accurately simulate the target procedure).

1.2 Evaluation Overview

Evaluation efforts for this project fall into two general categories:

- "Formative" evaluation, which attempts to provide design specification input to the development team during the development process
- "Summative" evaluation, which assesses the success of that effort by formally analyzing the effectiveness of the system

Details of our formative evaluation and our initial summative evaluation can be found in our earlier report (Heskamp et al, 1997; Weghorst et al, 1998). Specifically, we present evidence for the validity of the system as an ESS simulator. These findings are summarized in the first section below.

Our focus during the current phase was on the following three questions:

- **QUESTION 1** - What is the impact of experience with the simulator on actual OR performance?
- **QUESTION 2** - How does experience with the simulator affect individual components of surgical competence?
- **QUESTION 3** - What is the effectiveness of the various training aids and protocol methods incorporated into the simulator?

While our previous evaluation efforts had been focused on simulator *validation*, the current phase was focused primarily on *transfer of training* issues. Our methods and findings for these Phase 2 questions are detailed in subsequent sections.

2 Summary of Phase 1 Evaluation

Throughout the development of the ESS simulator, the Human Interface Technology Laboratory (HITL) team worked in close collaboration with Dr. Charles Edmond, staff otolaryngologist at Madigan Army Medical Center (MAMC). Our initial task was to define the system's performance requirements and to perform ongoing formative evaluation to provide design recommendations to the development teams at Lockheed Martin, OSC, and Immersion Corporation. Once the system was stabilized at Version 1.2 and shipped to the HIT Lab, we began a series of studies to assess the usability and efficacy of the system as a simulator for the ESS procedures of interest.

2.1 Phase 1 Formative Evaluation

Our methods of approach for the formative evaluation phase included the following:

- Endoscopic video analysis to determine simulator performance requirements
- Geometric complexity requirements analysis
- Prototype anatomical modeling
- Development of spatial awareness aids, interface features and rendering approaches
- Development of a prototype simulator with an integrated expert system assistant
- Development of a surgical training curriculum to be embedded in the simulator
- Survey of medical experts to determine feature and curriculum priorities

Dr. Edmond provided the core of domain expertise in sinus surgery. After a few weeks of study, the rest of the HITL team had gained a basic familiarity with sinus anatomy and surgical issues. During this time Dr. Edmond gained familiarity with the computer graphics tools available at HITL for use in prototyping simulator design and computer-assisted surgery applications.

In addition to design experiments based on prototype modeling, the HIT Lab served as a test site for successive versions of the anatomical model and the simulation system. Having the target hardware platform (SGI onyx) in-house made iterative evaluation a viable and useful approach. In particular, the proximity of the lab to MAMC made it relatively easy for the evaluation team to consult with Dr. Edmond regularly on simulator features and system performance.

Upon receiving each release of the ESS simulator, the HITL team would compile a list of feature enhancements, known bugs and other observations. This list would be prioritized by Dr. Edmond and returned to Lockheed Martin for inclusion in the subsequent release.

2.2 Evolution of the Training Framework

During the early phases of the project we examined several other surgical simulator systems to look for places to improve the state of the art. What we noticed was that although these

systems were making advances in anatomical modeling and user interaction, they were missing a structured educational component. In general these systems simulated a specific surgical domain or task only and were therefore "orphaned" experiences in the educational process.

We concluded that in order to be of significant educational value, one had to not only embed the simulator in the existing academic curriculum, but also develop curricula within the simulation itself. This would not only serve the educational process but would also facilitate our evaluation of the educational effectiveness of the simulator.

The prevailing paradigm in surgical education is usually summarized as: "See One, Do One, Teach One." Our goal was to improve on this paradigm by taking the "See One, Simulate Many, Do One" approach that had proven effective in the domain of flight training. To achieve this, we needed to develop a curriculum structure.

2.2.1 Task Analysis

Working with Dr. Edmond, we developed a taxonomy of ESS simulation objectives and simulator performance and interface features. Our objective, which proved far too demanding, was to elicit estimates from experienced ESS surgeons of the desirable system performance requirements for each training task. This effort did, however, provide us with a framework for approaching the issues of curriculum design in this domain.

2.2.2 ESS Domain Expert Survey

Twelve experienced ESS staff surgeons from several leading otolaryngology training programs were surveyed by Dr. Edmond to assess their judgments of the primary simulation requirements for physicians performing ESS procedures. The primary objective of this survey was to determine the curriculum needs and perceived importance of several of the candidate features of the system early in its development.

These domain experts were asked to rate the relative importance (from "critical need" to "not important") for 11 simulator characteristics: visual realism, spatial awareness training, haptic interaction, patient-specific modeling, psychomotor training, real-time interactivity, real-time (intra-operational) performance aids, standard surgical procedures, complications, advanced techniques, and pathophysiology. They were also asked to assess the value of these characteristics for target subjects at three levels of ENT experience: novice (i.e., junior residents), intermediate (senior residents), and expert (experienced ENT surgeons).

Results of this survey suggested that "spatial awareness" is the most crucial training need (of those presented in the survey) for all levels of subject experience, although "advanced techniques" were seen as equally critical for the experienced ESS target group. With the exception of "standard procedures" and perhaps "pathophysiology," all requirements were seen as equally important for all target levels or more important with increasing ESS experience.

It is interesting to note that almost all of these characteristics were rated on the "important" end of the scale. The characteristics deemed least important for simulation were "advanced

techniques" and "patient-specific modeling" for the novice group, and "standard procedures" for the advanced group. In addition, "real-time interactivity," "haptic interaction" and training in "complications" were seen as of only moderate importance for ESS novices.

2.2.3 Simulator Integration

A considerable amount of effort was spent during the development phase on establishing a "curriculum wrapper" for the simulator. We initially envisioned the simulator as embedded within a multimedia training system which would provide a meaningful context and set the protocol for the trainee.

The development team elected instead to integrate relevant aspects of the emerging ENT curriculum into the simulator itself and to postpone further efforts to develop a total curriculum "package." The reasons for this included:

- Development of a full-blown ESS curriculum was beyond the scope of this phase of the project.
- It became apparent that a staged protocol approach was necessary to make the simulator effective as a training context.
- Useful techniques emerged (such as the use of navigation hoops and injection targets) which could be integrated relatively easily.

The final product of the formative stage of the project was, in fact, more thoughtful with respect to an integrated ENT curriculum than had originally been anticipated and appears to be more useful as a training tool than other virtual reality medical simulators which have emerged in recent years.

2.3 Phase 1 Summative Evaluation

The primary objective of Phase 1 of our evaluation effort was to assess the procedural and task validity of the system with respect to the target ESS domain. Several studies were conducted using Version 1.2 of the simulator to determine this, along with an assessment of its perceived utility by experienced ENT surgeons.

2.3.1 Phase 1 Methods

Subjects: Run-time and survey data were collected for three groups of subjects on a common protocol progressing through the three basic ESS subtasks: navigation, injection, and dissection. Twelve non-MD subjects performed the tasks on a simplified abstract virtual model with instructional aids (hoops for the navigation path, injection targets, dissection spheres, auditory feedback about task completion, and simulated patient heart rhythm). Eight non-ENT MDs progressed from this "novice" model to a simulated anterior ethmoidectomy on an "intermediate" model with the aids embedded in the reconstructed and segmented paranasal anatomy. Finally, twelve MAMC otolaryngologists ranging from second-year ENT residents through senior staff progressed through both the abstract and

intermediate models, and then performed the simulated surgical procedure on an "advanced" model, consisting of the anatomy with no instructional aids.

We focused on these three groups to establish a baseline and asymptote for simulator performance, and to determine its validity for ESS procedures. We would expect that real world ESS experience would be positively correlated with performance on the simulated procedure. The first two groups also allowed us to "shake down" the system and the research protocols without sacrificing valuable ENT resident subjects, and provided baseline scores for untrained subjects. Finally, the non-ENT MDs provided us with valuable input about the extensibility of the simulator to non-ENT applications.

Methods: All subjects proceeded through a common protocol for one or more sessions, involving: (1) general orientation and consent form; (2) background questionnaire; (3) specific orientation to the simulator and tasks; (4) one or more simulation trials using one of the three training models; and (5) post-session debriefing and questionnaire. Performance measures were acquired at run time; in addition, all trials were videotaped for later analysis and "think-aloud" comments. Proctoring for each group was adjusted to their prior experience with the task. One proctor introduced the subject to the simulator and answered questions during the trial; a second proctor managed the records and the trial parameters for each subject.

Performance on each trial was automatically recorded by the system. Subtask scores were calculated as: *accuracy*optimal-time/completed-time.* To normalize across models optimal times were derived from the performance by two experienced ESS consultants on each subtask for each model. Navigation accuracy was based on the percentage of hoops negotiated (with the hoops in the advanced model rendered invisibly). Injection accuracy scores were based on the ratio of the percentage of each target injected to a criterion percentage for that object: 100% for the graphical targets, 25% for the middle turbinate, and 10% for the nasal wall. Similarly, dissection accuracy was based on a criterion dissection percentage for each dissectable object (markers, polyps, uncinate, ethmoid cells, ethmoid bulla, maxillary ostium, and bone fragments).

2.3.2 Phase 1 Conclusions

The validity of the Madigan simulator for the ESS domain was suggested by a number of findings:

- ENT subjects performed better than non-MD subjects on both the novice (abstract) and intermediate (anatomical with aids) models.
- Initial performance on the novice model was correlated with ENT residency level and degree of prior ESS experience.

- Analysis of instrument deviation from dissection targets provided convergent support for the validity of the run-time scoring algorithm.
- ENT subject ratings of the realism of the virtual anatomical model were consistently high on the post-session questionnaire and in open-ended comments.
- Post-training questionnaire responses confirmed that the simulator was generally perceived as valid and useful for ESS training by the ENT subjects themselves.

These findings provide convergent evidence that the Madigan ESS simulator represents a valid and useful implementation of the target ESS tasks. Furthermore, the thoughtful integration of an organized curriculum perspective appeared to enhance the pedagogical value of the system and allowed the developers to take better advantage of the training potential of virtual simulation.

The Phase 1 evaluation study also suggests a useful framework for incorporating systematic evaluation into the process of developing procedural training simulators in the medical domain. Incorporating both formative and summative aspects of evaluation greatly enhanced the development process and assures the continuing evolution of a usable and effective system. Ideally, subject performance on the simulator is reliably predictive of OR performance; the issue of transfer of training from the simulation system to the OR was thus of primary interest in Phase 2 of this evaluation effort.

3 Phase 2 Evaluation

3.1 Overview

Phase 2 evaluation studies focused on the training effectiveness of the Madigan ESS simulator. For these studies we used the beta release of Version 2.0 of the system (note that the final Version 2.0 release is now in place at MAMC). Version 2.0 (beta and final) differed from Version 1.2 by a number of significant requested features which are evaluated in a later section of this report:

- Hardware upgrade to SGI Infinite Reality graphics
- Reduced latency down to ~100 milliseconds from ~200 milliseconds
- Enhanced anatomical segmentation, deformability, and dissectability
- Addition of hazards in all models, including tool-appropriate abstract obstacles in the novice model and appropriate anatomical hazards (nerves, arteries and lamina) in the intermediate and advanced models
- Additional virtual instruments
- Optional voice control interface
- Embedded "Virtual Instructor"

- Improved haptic interface

To evaluate the effectiveness of the Madigan ESS simulator as a training tool, we systematically measured several indicators of surgical skill:

- Real-time performance in the OR as assessed by an attending staff ENT proctor
- Blind analysis of endoscopic video acquired during initial OR procedures for residents with and without simulation training
- Independent paper-and-pencil tests of several components of surgical competency

Subjects for this phase were 10 junior and senior ENT residents at MAMC, some of whom had had prior training with the simulator and some of whom had not. Since the number of available residents was limited and not all of them were advanced enough to perform real OR procedures, we were constrained to performing quasi-experimental studies rather than full factorial design experiments. As such, these should be considered "pilot" studies which provide some potentially useful methods for evaluating surgical simulators, as well as some suggestive findings for further exploration.

In addition to these measures of transfer of training, we also continued our assessment of the usefulness of various components of the simulator. Because of the number of features of interest, systematic controlled variable studies of utility were not possible within the scope of this project. Instead we provide subjective assessments by our experienced staff consultants and by the resident trainees themselves, as well as observations by the technical proctor. These assessments are useful for informing the development team about further system development and may provide research hypotheses for future studies.

3.1.1 Simulation Testing Protocol

Each subject was initially introduced to the overall configuration of the simulator and informed that we were evaluating it as a trainer for residents in Otolaryngology. They were introduced to the voice recognition software and were encouraged to familiarize themselves with the interface and commands by running through some of the available menus. They were informed that any voice commands they could not remember could be described to the proctor, who would then execute the command at the system console.

The instrumented endoscope was then introduced, and subjects were informed of the availability of a 30, 45 and 70 degree scope which could be swapped for the zero degree scope they would initially be given. They were also informed of their ability to rotate the virtual lens axially (while using the angled scopes) by rotating the shaft of the instrumented endoscope, and to rotate the virtual camera by speaking to the voice recognition software, which would rotate the scope axially in 5-degree increments.

Subjects were then introduced to the instrumented forceps and informed that they would be used to simulate all virtual dissection tools and the virtual bent needle for injection. The mechanics of the instrumented forceps (open and closed) were described as simulating the plunger for the syringe during Injection and opening and closing the jaws of the dissection tools, and the foot switch operation of the rotating tools was demonstrated.

Subjects were then given a brief verbal description of the three subtasks (Navigation, Injection and Dissection) and what would be required of them during the trial. Each subject was then shown a video of the virtual endoscopic view of the procedure performed by Dr. Mesaros (MAMC staff otolaryngologist and co-investigator). During the videotape, subjects were allowed to step up to the mannequin to become familiar with the instrumentation, and the proctor described the subtasks in more detail, including what was required for completion of the trial. The "blood effects scope" was shown in the video and reasons for it were described by the proctor, along with how to remedy the problem by wiping the instrumented scope on the foam pad located on the mannequin's forehead. Subjects were encouraged to ask any questions and to "think aloud" during the procedure.

After completing Model 1 (the "Novice Model" described below), the subjects proceeded to Model 2 ("Intermediate Model"). In this model, subjects were required to perform a total ethmoidectomy, which included performing the three subtasks of Model 1, with the training aids embedded in the virtual paranasal anatomy. Again, a video of the tasks was shown while the proctor explained, in detail, the subtasks and what was required for completion of the trial. During the video, the subjects were familiarized with the various training aids, which would be guiding them through the procedure. In addition, they were shown the text labels overlaying the anatomical structures, which would aid them in learning the paranasal anatomy through an endoscopic perspective, and which would serve as orientation cues throughout the trial. Each subject's familiarity with the procedure determined the need for further instruction by the proctor during the trial.

Subjects then proceeded to Model 3 ("Advanced Model"). This model required the subjects to perform a total ethmoidectomy and remove three polyps lateral to the middle turbinate without the help of the training aids. During this trial, the subjects were expected to remember the Navigation paths taken, as well as where to inject and what dissect to complete the procedure. Each subject's baseline familiarity with the ESS procedure determined the need for further instruction by the proctor during the trial.

3.1.2 Simulation Procedure

Novice Model: The Novice anatomical model consisted of only the surface of the face and the entrance to the nasal cavity. An abstract 3D grid pattern replaced the sinus anatomy, providing a less complex environment in which to learn the three basic procedural subtasks: Navigation, Injection and Dissection. Training aids were used to guide the subjects through

the subtasks. Navigation training aids consisted of virtual hoops and Injection training aids consisted of virtual targets in space.

During Navigation the subjects maneuvered the instrumented endoscope through four sets of virtual hoops. The paths of the hoops represented three passes (sets two and three combined are one pass) commonly taken before ESS surgery begins to gain familiarity with the patient's anatomy and to allow cleaning of the areas of interest.

Injection consisted of coordinating both the endoscope and instrumented forceps within the environment to inject five targets oriented obliquely in space. The instrumented forceps controlled a virtual needle for this subtask. The placement of the targets in space reflected the common areas of injection of a vasoconstrictor during a maxillary antrostomy.

During Dissection subjects were also required to use both the endoscope and the instrumented forceps. The subtask consisted of a tool-specific target object and a dissectable hazard surrounding the object. Subjects were to dissect the target objects with pre-selected virtual tools and to avoid the hazards. The instrumented forceps represented each of the tools most commonly used in the procedure: left side-biter, up-biter, microdebrider ("hummer"), sickle knife, router, left antrum punch, round burr, straight-biter, left scissors, bent router, and circular antrum punch.

Navigation through the four sets of hoops, injection of the five targets, and dissection of each of the obstacles was required for a complete score. If the hazards surrounding the Dissection targets were hit, the subject's score was decreased by a percentage proportionate to the amount of the hazard dissected. Digitized audio cues were given for each hoop negotiated in Navigation, for each target Injected, for percentages completed of each object during Dissection, for interaction with hazards, and for completion of each subtask.

Intermediate Model: The Intermediate model was composed of the Navigation hoops and Injection targets from the Novice model embedded within a virtual anatomical model of the paranasal sinus. Injection and dissection followed the protocol for a total ethmoidectomy. Text labels were overlaid on all anatomical structures with which the subject interacted.

Navigation through the four sets of hoops, injection of the five targets, dissection of the five anatomical structures, removal of two bone fragments placed in the uncinate process and removal of three bone fragments placed in the bulla ethmoidalis were required for a complete score. The hazards in this model were: anterior ethmoid artery, carotid artery, lamina papyracea, opthalmic artery, optic nerve, periorbita and periorbital fat. Audio cues were given for each hoop negotiated in Navigation, for each target hit in Injection, for each bone fragment removed, for interaction with hazards, for percentages completed for each anatomical structure in Dissection, and for completion of each subtask.

Advanced Model: The Advanced model was composed of the anatomical model only. Subjects were expected to perform the three subtasks without the training aids and to follow the protocol for a total ethmoidectomy. Three polyps were added superior/anterior to the bulla ethmoidalis.

During Navigation subjects were required to perform the three passes in the same order as in Models 1 and 2. First, the inferior pass along the floor of the nasal passage to the nasopharynx was performed. This was followed by an intermediate pass medial to the middle turbinate towards the upper aspect of the nasopharynx and sphenoid ostium, and then retracting the scope and approaching the ostial meatal complex from anterior to the middle turbinate. Finally, the superior pass was performed medial to the root of the middle turbinate towards the sphenoethmoidal recess.

During Injection, as during actual OR procedures, the subject was cued only by the amount of blanching (whitening) of the virtual tissue as to whether more vasoconstrictor was needed. Dissection followed the same procedure as for Model 2.

Total score on the Advanced model was based on: navigation through the three passes, injection of the areas of interest, dissection of the five anatomical structures, dissection of the three polyps, removal of two bone fragments placed in the uncinate process, and removal of three bone fragments placed in the bulla ethmoidalis. Hazards were the same as in Model 2. Audio cues were given only at the end of each of the Navigation passes, for removal of bone fragments and for percentages completed in Injection and Dissection.

3.2 Transfer of Training to OR Performance

Residents from the MAMC ENT program, having a range of prior experience with the ESS simulator, were evaluated for their operating room performance using two approaches: (1) real-time ratings by attending staff proctors, and (2) independent blind ratings of captured endoscopic videotapes by senior staff surgeons.

3.2.1 OR Performance Ratings by Staff Proctors

Methods: Operating room proctoring was performed by one of three senior staff physicians who had been working together on ENT training at MAMC and who had similar approaches to teaching ESS procedures. Informal comparisons indicated high inter-rater reliability, but this should be more formally assessed in future studies.

Five ENT residents were proctored while performing a routine ESS procedure. Proctor evaluations consisted of 10-point scale ratings (1 = "inadequate", 10 = "perfect") for each the following aspects of the procedure:

- Patient Preparation
    a. Turn the bed
    b. Patient positioning
    c. Intranasal preparation (placement of pledgets with topical vasoconstrictor and local anesthesia)
    d. Facial skin prep
    e. Surgical draping of patient
    f. Monitor positioning
    g. Set-up of endoscopic equipment (suction, endoscopes, microdebrider, Mayo stand, etc)
    h. CT scans on view box
- Navigation (Left and Right)
    a. Inferior pass
    b. Intermediate pass
    c. Superior pass
- Injection (Left and Right)
    a. Uncinate
    b. Root of middle turbinate
    c. Lateral aspect of middle turbinate
    d. Sphenopalatine artery
- Medialization of middle turbinate (Left and Right)
- Uncinectomy (Left and Right)
- Maxillary antrostomy (Left and Right)
- Anterior ethmoidectomy (Left and Right)
- Posterior ethmoidectomy (Left and Right)
- Sphenoidotomy (Left and Right)
- Overall score

Results: Although an attempt was made to select equivalent cases, the normal variability across patients is such that some of these individual components were not relevant to all cases, resulting in a somewhat sparse data matrix. In addition, with the junior residents, in particular, the attending staff proctor would sometimes perform the procedure on one side or would sometimes take over the instruments during particularly troublesome tasks. Consequently, we normalized for these variations by averaging the proctor's scores across all items which were relevant for that particular surgical case, resulting in a "Mean OR" proctor rating for each resident.

When we examine the bivariate correlations between these Mean OR ratings and the major measures of each resident's performance on Version 2.0 of the simulator (Novice Model Score, Intermediate Model Score, and Advanced Model Score), we find that, while Novice and Advanced Model scores are positively correlated with these OR proctor ratings ($r = .520$ and .874, respectively), the only statistically significant predictor of Mean OR Rating is their Intermediate Trial Score ($r = .959$, $p < .05$). This relationship may reflect the actual level of

skill acquired by these subjects (i.e., no longer novices but not yet advanced enough to perform the procedure independently).

3.2.2 Ratings of Videotaped Procedures by Experienced ENT Staff

Methods: In addition to the real-time ratings of resident OR performance during Phase 2, endoscopic recordings of first-time OR procedures had been archived for four junior residents, two with considerable experience on Version 1.2 of the simulator prior to their first OR procedure and two with no prior simulation training. A panel of four experienced ESS surgeons conducted independent evaluations of these videotapes, along with a comparison endoscopic recording of a comparable procedure performed by an experienced surgeon. Raters were blind to the identity of the surgeon in each videotape, including the one of the experienced surgeon. Each videotape was rated on a 10-point scale (1 = "inadequate", 10 = "perfect") for each of the following performance criteria:

- Navigation
- Injection
- Uncinectomy
- Anterior ethmoidectomy
- Maxillary antrostomy
- Orientation of videoimage
- Image-task alignment
- Proper depth of image for task
- Tool manipulation
- Tool selection
- Tool-tool dexterity
- Tissue respect
- Surgical confidence
- Case difficulty

Results: When we look at the bivariate correlation matrix for these staff ratings of initial endoscopic procedure it is interesting to note that all of the above measures were partially intercorrelated except for Case Difficulty and Navigation. This finding for Case Difficulty is as expected, since the raters were instructed to normalize their ratings across cases (i.e., to not score the tapes differentially based on case difficulty, but to focus on the basic surgical skills of interest).

The finding of no correlation for Navigation is perhaps due to the extreme variability across cases in navigability of the anatomy, and a subsequent confusion by the raters as to what they should be looking for. Supportive of this conclusion is the finding that there was a significant between groups ANOVA (analysis of variance) effect by rater for Navigation only

($p < .009$), out of all the rating measures, indicating low inter-rater reliability for that measure, again perhaps due to some confusion about what was to be measured.
While mean rating across items was, as expected, correlated with all of the other components, the strongest predictors of mean videotape rating were Tool Manipulation ($r = .962$), Tool Selection ($r = .961$), Maxillary Antrostomy ($r = .950$), and Surgical Confidence ($r = .948$).

When we look at the correlations among all of the rating measures, an interesting trend emerges: two major sets of variables are highly intercorrelated and appear to represent separate factors:

Factor 1
image-task alignment
orientation of image
proper depth of image
tool-tool dexterity
tissue respect
injection (weaker correlation)

Factor 2
surgical confidence
anterior ethmoidectomy
tool manipulation
maxillary antrostomy
uncinectomy

The three remaining variables fall out as follows:
tool selection (correlated with both of the above factors)
case difficulty (not correlated with any item)
navigation (negatively correlated with everything, but no significant correlations)

Factor 1 appears to coalesce around "scope and instrument control," while Factor 2 may be characterized as "dissection skill." It is interesting to note that Surgical Confidence is most highly correlated with this second factor, indicating perhaps that it is based upon experience in successful dissection.

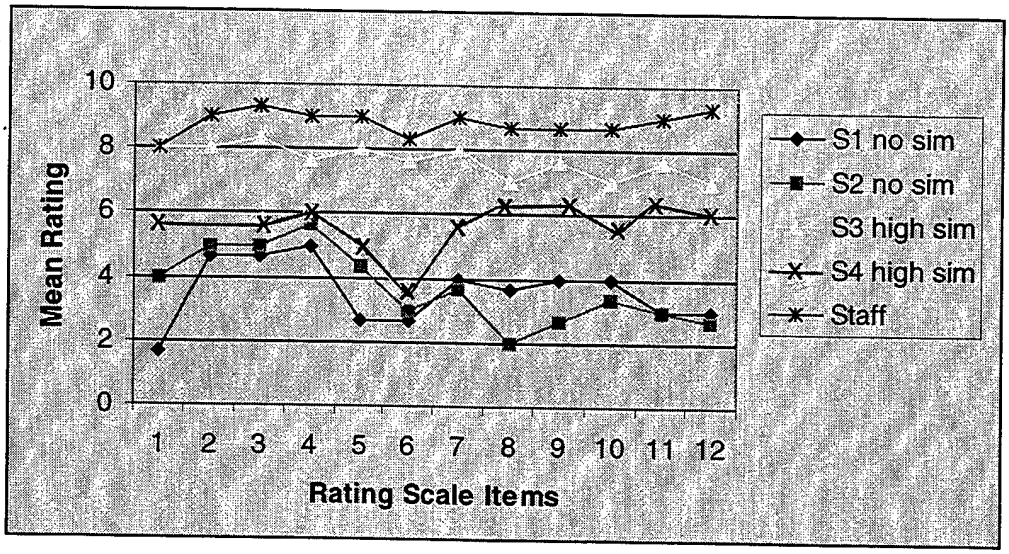


Figure 2. Mean ratings (across raters) for each of the videotape measures, organized by factor, with "scope and instrument control" items on the left and "dissection skill" items on the right. Note that Item 7 was correlated with both factors.

Figure 2 illustrates the mean ratings (across raters) for each of these measures organized by subject (i.e., although a somewhat unconventional use of a line graph, each line in the chart represents the performance of each subject across these measures). To help visualize the two major item groupings, rating scale items are arranged in this figure as follows:

1. Injection
2. Orientation of videoimage
3. Image-task alignment
4. Proper depth of image for task
5. Tool-tool dexterity
6. Tissue respect
7. Tool selection
8. Uncinectomy
9. Anterior ethmoidectomy
10. Maxillary antrostomy
11. Tool manipulation
12. Surgical confidence

It can be seen that simulation training appears to have positively affected these initial OR performance ratings, although Subject 4 appears to have benefited more on the second factor ("dissection skill") than on the "scope and instrument control" factor. While the two simulation-trained residents were rated consistently better than the other two residents across all measures (See Table 1 below), these differences were statistically significant for only two items (most likely due to the small number of subjects): Tool Manipulation (t =

6.00, p = .027) and Surgical Confidence (t = 6.96, p = .020). These results are nevertheless very promising and suggest that simulation training with this system does indeed have a positive transfer of training effect.

Table 1. Mean Ratings for Residents With and Without Prior Simulation Training Across Videotape Rating Criteria for Their First OR Procedure

Group Statistics

	Simulation Training	N	Mean	Std. Deviation	Std. Error Mean
Navigation	No Sim Training	2	7.6667	.0000	.0000
	Sim Training	2	6.5000	1.1785	.8333
Injection	No Sim Training	2	2.8333	1.6499	1.1667
	Sim Training	2	6.8333	1.6499	1.1667
Uncinectomy	No Sim Training	2	2.8333	1.1785	.8333
	Sim Training	2	6.6667	.4714	.3333
Anterior Ethmoidectomy	No Sim Training	2	3.3333	.9428	.6667
	Sim Training	2	7.0000	.9428	.6667
Maxillary Antrostomy	No Sim Training	2	3.6667	.4714	.3333
	Sim Training	2	6.3333	.9428	.6667
Orientation of Image	No Sim Training	2	4.8333	.2357	.1667
	Sim Training	2	6.8333	1.6499	1.1667
Image-Task Alignment	No Sim Training	2	4.8333	.2357	.1667
	Sim Training	2	7.0000	1.8856	1.3333
Proper Depth of Image	No Sim Training	2	5.3333	.4714	.3333
	Sim Training	2	6.8333	1.1785	.8333
Tool Manipulation	No Sim Training	2	3.0000	.0000	.0000
	Sim Training	2	7.0000	.9428	.6667
Tool Selection	No Sim Training	2	3.8333	.2357	.1667
	Sim Training	2	6.8333	1.6499	1.1667
Tool-Tool Dexterity	No Sim Training	2	3.5000	1.1785	.8333
	Sim Training	2	6.5000	2.1213	1.5000
Tissue Respect	No Sim Training	2	2.8333	.2357	.1667
	Sim Training	2	5.5000	3.0641	2.1667
Surgical Confidence	No Sim Training	2	2.8333	.2357	.1667
	Sim Training	2	6.5000	.7071	.5000
Case Difficulty	No Sim Training	2	5.0000	.0000	.0000
	Sim Training	2	7.1667	.7071	.5000
Overall Mean Rating	No Sim Training	2	4.0238	6.73E-02	4.8E-02
	Sim Training	2	6.6786	1.0943	.7738

ANOVAS for videotape ratings by the ENT docs revealed significant between groups effects by subject for all but Proper Depth of Image, Case Difficulty, Navigation, and Orientation of Videoimage. In other words, these four measures did not significantly discriminate among subjects, but all of the other measures did.

3.2.3 Ratings of Videotaped Procedures by Technical Staff

In an attempt to determine if reliable measures of surgical performance could be derived from endoscopic video by surgically inexperienced observers, we also performed a parallel (but independent) blind analysis by the two primary technical team members at the HIT Lab. While these ratings were not used to evaluate the transfer effects of simulator training, we include them here to suggest a possible methodology for future evaluation studies.

Methods: Each of 21 videotaped procedures were rated on a 7-point scale (1 = "extremely poor", 7 = "extremely proficient") independently by the observers for each of the following performance criteria:

- Scope Steadiness
- Scope Target Acquisition
- Ease of Scope Navigation
- Tool Steadiness
- Tool Target Acquisition
- Ease of Tool Navigation
- Injection Efficiency
- Dissection Efficiency
- Degree of Caution Exercised
- Tool/Scope Coordination
- Overall Assessment

Although these observers were familiar with the basic ESS procedures, they lacked experience with the variety of anatomy and case difficulty seen over time in the OR. Thus, their focus was on these more objectively observable characteristics of surgical performance.

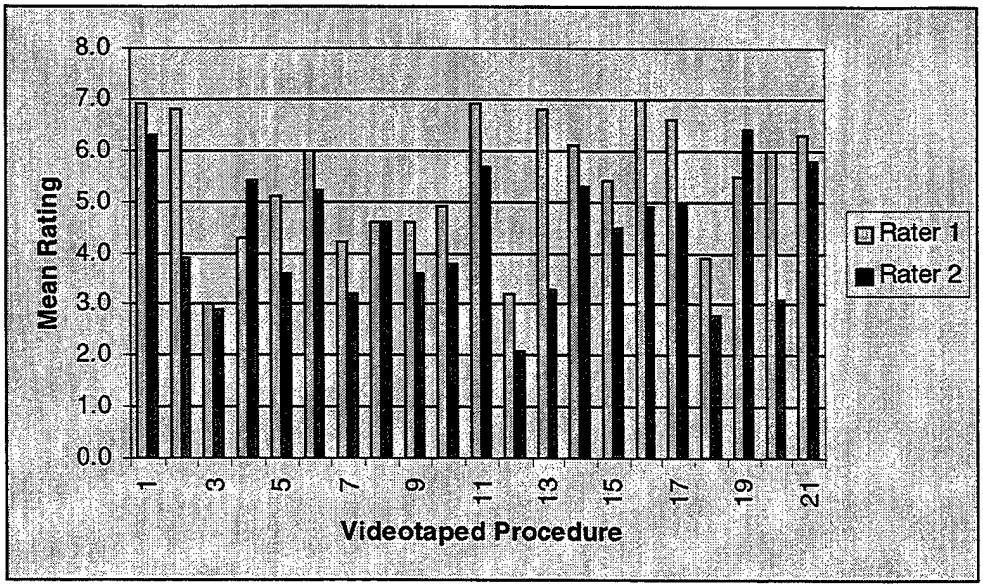


Figure 3. Mean ratings by both non-MD raters for each of 21 videotaped procedures

Results: Examining the non-MD ratings across the 21 videotaped procedures, we find a consistent mean difference between the two raters, with Rater 1 consistently rating the items significantly higher than Rater 2 (with the exception of Procedural Caution and Scope Steadiness, which did, however, approach significance). Despite this between-rater scaling difference, we do find a significant rank order correlation between the two raters for mean rating across all scale items ($r = .589$, $p < .005$), as can be seen by examining Figure 3.

When we look at inter-rater reliability for individual items on the non-MD rating scale we find positive correlations between raters for all measures. However, we find a fair amount of variability in the *strength* of those correlations: scores by the two raters were significantly correlated ($p < .05$) for Overall Rating, Scope Steadiness, Scope Target Acquisition, Ease of Scope Navigation, Injection Efficiency, and Scope-Tool Coordination. Ratings were *not* significantly correlated for Tool Steadiness, Tool Target Acquisition, Ease of Tool Navigation, Dissection Efficiency, and Procedural Caution.

Since the mean rating showed a strong inter-rater reliability we feel confident using it as an independent measure of surgical performance (as judged from the endoscopic videotapes) in future studies. In the current study, however, we were unable to validate these non-MD ratings against ratings by experienced ESS surgeons. This was due to differences between the two rating scale items and the non-equivalent videotape segments (i.e., the ESS surgeons rated selected representative segments of resident OR performance, while the non-MD raters viewed complete procedures, portions of which conducted by the attending physicians). In follow-on studies we will attempt to determine the validity of the non-MD rating approach by having both groups rate the same stimulus materials. It seems likely, however, that these non-MD ratings will be at best an adjunct to ratings by experienced surgeons, who are able to address more sophisticated surgical competency skills.

3.3 Transfer of Training to Components of Surgical Competency

The primary goal of this phase of the evaluation was to explore how experience with the simulator may differentially affect individual components of surgical competence. In order to tease apart these effects, the evaluation team, led by Dr. Edmond, measured subject performance on independent measures of:

1) Paranasal sinus anatomical knowledge (basic, complex, and anomalous).
2) Endoscopic view identification using different scope angles, including recognition that a particular view is ambiguous.
3) Procedural knowledge (both general and detailed for all simulated procedures).
4) Instrument knowledge (identification, function, and appropriate selection).
5) Visual spatial orientation (endoscopic).

Since no standards exist in the field for evaluating the components of surgical competency, these measures were developed in consultation with experienced ESS surgeons and were judged to have captured essential skills needed to perform ESS procedures.

Subjects

A total of 9 Otolaryngology residents from Madigan Army Medical Center completed all phases of the individual component evaluation. Residents from 5 residency year groups were enrolled: 2 fifth-year residents; 1 fourth-year resident; 2 third-year residents, 2-second-year residents and 2 first-year residents. All subjects were tested in the same setting using a videotape recording of a simulated procedure generated from the Madigan ENT surgical simulator. The simulator was used to generate recordings which were transferred to videotape for playback and testing of the five dependent measures of surgical competency. Six of the 9 subjects in the study group had experience with the simulator prior to this testing.

3.3.1 Paranasal Sinus Anatomical Knowledge (Basic, Complex and Anomalous)

Knowledge of the paranasal sinus anatomy is critical to diagnosing normal and abnormal states as well as differentiating aberrations that are associated with certain pathologic conditions. Knowledge of the surgical anatomy is a critical step towards surgical competence.

Methods: A videotaped recording of a simulated navigation of the Intermediate model (with labels turned off) was performed. At various stops in the navigation, the subjects were asked to identify the anatomical structure in view. A list of possibilities where supplied on the subjects' answer sheets.

Basic anatomy consisted of 11 structures and included the following: ethmoid bulla, middle turbinate, uncinate, maxillary ostium, sphenoid ostium, non-obstructing septum, inferior turbinate, maxillary antrum, sphenoid sinus, nasopharynx, eustachian tube orifice. *Complex* anatomy consisted of four structures, and included the following: frontal recess, lamina papyracea, skull base and posterior ethmoid cavity. *Anomalous* anatomy consisted of three structures, and included the following: concha bullosa, paradoxical middle turbinate and deviated septum.

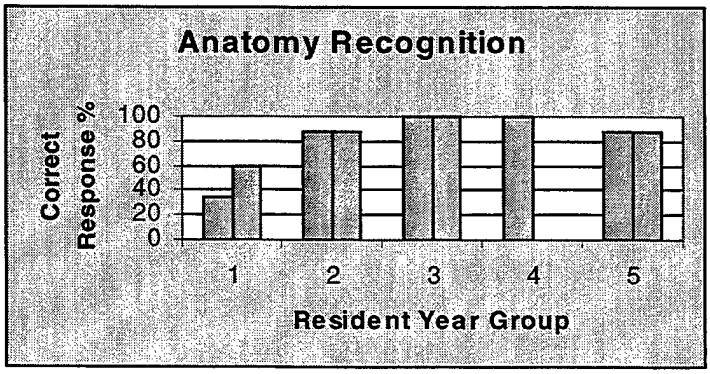


Figure 4. Histogram of Anatomy Recognition Test Performance by Year of Residency

Results: As a group, the residents missed 26 of 169 responses. The structure that was uniformly missed was the skull base. Five of 6 residents misidentified this structure as either the frontal recess or the posterior ethmoid cavity. In general, the residents had more difficulty with the complex and anomalous anatomy. Interestingly, the intermediate residents fared better than the R5s or R2s. This probably reflects their greater simulation time over the course of the evaluation of the simulator during both Phase 1 and Phase 2.

The fifth year residents had difficulty with the complex anatomy that involved the skull base, frontal recess and lamina. This particular area is quite complex in both cadaver and actual dissection, and confusing even to experienced physicians. The residents' responses were regionally correct, however, substituting frontal recess for skull base and lamina for posterior ethmoid cavity.

The fourth and third year residents overall performed the best, recognizing all of the virtual structures (17 of 17 structures). Again, the high scores for this particular group may reflect their increased overall simulation time with respect to the other residents.

The second year residents missed on average 2 of 17structures each. The junior resident encountered more difficulty with the basic anatomy and the anomalous anatomy, confusing the paradoxical middle turbinate for the concha bullosa, and maxillary ostium with maxillary sinus.

The first year residents had the greatest difficulty with the virtual anatomy. On average, they missed 9 of the 17 structures, encountering difficulty with basic, complex and anomalous anatomy. This group had no experience with the simulator, unlike all but one other resident (in the second year group).

3.3.2 Endoscopic View Identification Using Different Scope Angles, Including Recognition that a Particular View is Ambiguous

The goal of this component of the exam was to ascertain whether the residents could perceive differences in scope angle, given the maxillary ostium as the reference target.

Awareness of anatomy from different perspectives is an important surgical attribute that gives the surgeon a better sense of the three-dimensionality of the paranasal structures. In addition, familiarity with these different perspectives is often required to perform certain critical tasks during a sinus surgery. Inability to negotiate these perspectives limits the surgeon's effectiveness.

Methods: The residents were presented with 6 different target fields from the perspectives of a zero-degree, 30-degree and 70-degree scope angle. The view was presented as both a static image that was interchanged between target fields, as well as a dynamic view, consisting of navigation to the maxillary ostium with the selected scope.

Results: The fifth-year residents had difficulty differentiating between the zero- and 30-degree scopes, and between the 30-degree and the 70-degree scopes. On average, they were able to recognize 3 of the 6 views that were presented. The difficulty with this particular task could relate to the paucity of visual landmarks and the absence of psychomotor participation in the task, which may normally facilitate image and spatial perception.

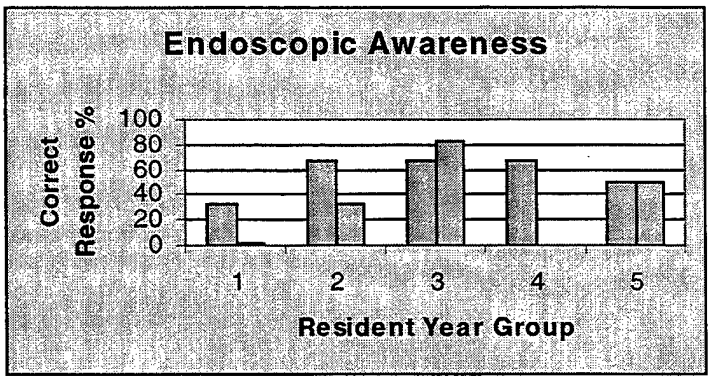


Figure 5. Histogram of Endoscopic Awareness Test Performance by Year of Residency

The fourth- and third-year residents again performed the best, with the fourth-year resident correctly identifying 4 of 6 views, and the third-year residents correctly identifying 5 of 6 and 4 of 6 views, respectively.

Difficulties were noted with all three scope angles, and as noted with the fifth-year residents, the difficulty with this particular task could relate to the absence of psychomotor participation in the task which normally facilitates image and spatial perception.

The second-year residents had mixed results, having more difficulty than the third- and fourth-year residents, and missing on average 3 of the 6 images. Surprisingly, the one resident who correctly identified 4 of 6 images had not used the simulator.

The first-year residents scored the worst overall, correctly identifying only 1 of 6 views on average. This is as expected, since they had not performed any endoscopic sinus surgery, nor had they had any experience with the simulator. They encountered difficulty with both static and dynamic images at all angles.

3.3.3 Procedural Knowledge

Endoscopic sinus surgery is an operative procedure that is performed in a sequential manner. The standard surgery is performed in an anterior to posterior format (front to back). The sequence of events initially consists of navigation, injection, and uncinectomy. After these initial steps are complete, the surgeon can proceed with maxillary antrostomy, anterior ethmoidectomy or dissection of the frontal recess. The surgeon's initial decision dictates the subsequent sequence of surgical steps necessary to complete the operative procedure.

Methods: The resident's knowledge of the operative sequence and their ability to decide alternatives to the procedure based on operative obstacles was assessed. The initial assessment was a multiple-choice questionnaire consisting of four questions that were posed after viewing a videotape of a phase of the virtual surgery. The residents were presented with four choices, and were instructed to choose the best response.

Results: The residents as a group did well, with 86% correct responses. Year of residency affected procedural knowledge as follows:

- 5th year residents achieved 88% correct responses (5a 75% and 5b 100%).
- 4th and 3rd year residents again performed the best, achieving 100% correct responses.
- 2nd year residents achieved 88% correct responses. (2a 100% and 2b 75%).
- 1st year residents achieved 63% correct responses (1a 100% and 1b 25%).

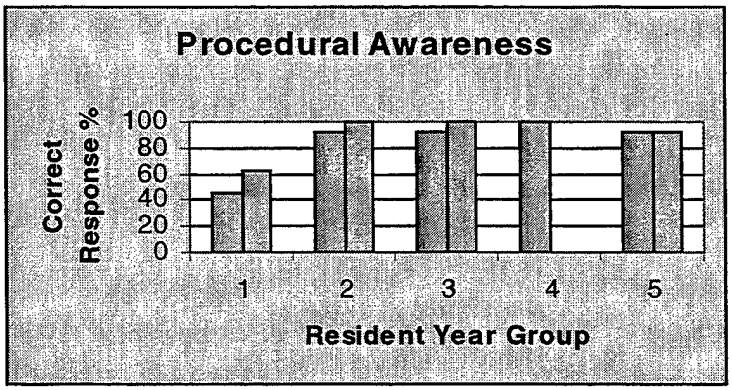


Figure 6. Histogram of Procedural Awareness Test Performance by Year of Residency

The second phase of the procedural knowledge assessment consisted of completing a wire diagram of the sequential steps in a standard endoscopic sinus procedure. Again, the residents were provided with a list of possible responses.

The residents as a group did well, achieving 86% correct responses. Overall, the fifth-, fourth-, third-, and second-year residents did extremely well, missing only two responses among them (5b, 4b). The percent correct for the more advanced residents was 99%. The first-year residents, in contrast, had significant difficulty with the procedural steps. The two first-year residents achieved only 50% correct responses. (1a 44% and 1b 55%).

Analysis of the overall scores on two tests revealed the following:

- 5th year residents achieved 92% correct responses.
- 4th and 3rd year residents achieved 96% correct responses.
- 2nd year residents achieved 96% correct responses.
- 1st year residents achieved 54 % correct responses.

3.3.4 Instrument Knowledge

Knowledge of the tools necessary to perform endoscopic sinus surgery is critical to a successful outcome. As important as knowledge of the surgical anatomy is a clear understanding of the surgical tools necessary to complete the procedure in a safe and effective manner. The instruments used in endoscopic sinus surgery have specific functions, and it is essential that the surgeon have a full understanding of the capabilities and uses of each individual surgical instrument.

Methods: Residents were tested on their knowledge of the tools used in endoscopic sinus surgery, as well as on their function and the region in which they are most commonly deployed. They were told that the function and region of use questions could have more than one response. Because of the relatively poor performance of the R1s on the tool *identification* portion of the test, only residents in year groups 2-5 were asked to complete the tool *function* portion, identifying the primary function for each individual tool. For the region of use test the residents were asked to indicate the location of most common use for each individual tool, and were allowed to submit more than one answer.

The tools were classified as basic and advanced. The *basic* tools consisted of the following: straight needle, bent needle, freer, seeker, left side-biter, right side-biter, microdebrider, bent router, scissors-right, scissors-left, straight-biting forceps, suction and sickle knife. The *advanced* tools consisted of the following: sphenoid punch, left maxillary antrum punch, right maxillary antrum punch, burr and aggressive router.

Results: The group as a whole had variable results. In general, the more senior residents had the better outcome. Specific aspects of instrument knowledge are detailed below.

Tool Identification

The fifth-year residents missed on average 2 of 19 responses (5a 3/19 and 5b 1/19). Resident 5a confused the bent router with the aggressive router and the orientation of the left and right side biters. Resident 5b confused the burr with the aggressive router. In contrast to the fifth-year residents, the fourth and third year residents were knowledgeable of all tools.

The second year residents missed 10 of 19 responses (2a 8/19 and 2b 2/19). Interestingly, the resident with no simulator time scored better on tool identification than the second year resident with simulator experience. Most of the difficulty encountered was with tool orientation and in confusing the advanced tools (i.e., burr for aggressive router, etc).

The first year residents had the most difficulty, missing on average 8 of 19 responses (1a 5/19 and 1b 11/19). Resident 1a confused the aggressive router with the sphenoid punch, the up-biter with the maxillary antrum punch, the side-biter with the up-biter and the microdebrider with the suction tool. Resident 1b encountered the most difficulty, identifying only the most basic tools.

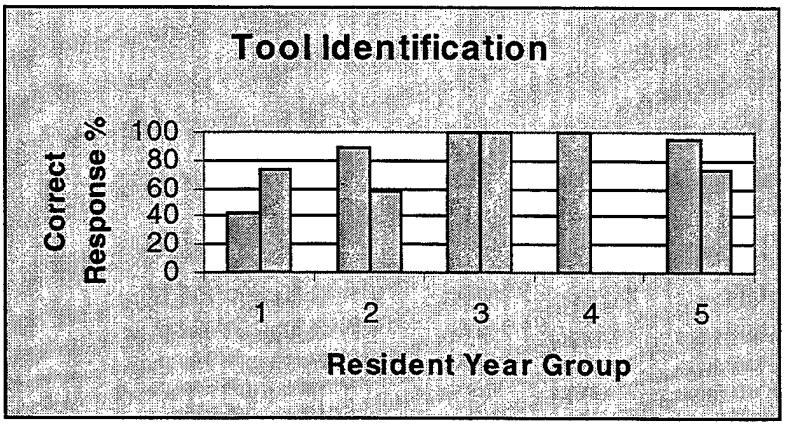


Figure 7. Histogram of Tool Identification Test Performance by Year of Residency

Tool Function

The list of possible options for tool function consisted of the following: injection, dissection, palpation and suction. Although more than one answer was possible, residents received full credit for listing any correct response.

The group as a whole achieved 100% correct responses. The residents had instructions to indicate additional responses for those tools that had more than one function. As year of residency increased from second to fifth year, the number of correct additional responses for each individual tool increased. The additional responses by the more senior residents suggest a better understanding of each tool's function and utility. For example, using a tool that was designed principally for dissection to palpate the anatomy is an advanced application for that tool. The more junior residents' responses were single responses. Applying additional functions to tools that were not implicitly designed for that function is a skill that appears to be learned through observation of more experienced surgeons and live surgery.

Region of Use

The instruments used in endoscopic sinus surgery are designed principally for certain procedural tasks during the course of surgery. Certain tools are designed for specific tasks within certain regions of the paranasal cavity, while others have universal application. The regions of use were defined as the following: anterior ethmoidectomy, posterior ethmoidectomy, maxillary antrostomy, sphenoid ostium, middle turbinate, uncinate process, frontal recess and through-out the paranasal sinuses.

The responses by groups were essentially identical. The residents for each year group missed on average 7 of 19 responses. Because of the large variance in responses, the reliability for this portion of the test is suspect; therefore, no specific conclusions are drawn.

3.3.5 Spatial Awareness

Spatial awareness is a critical skill necessary for safe and effective sinus surgery. The paranasal sinuses are quite complex, with a considerable amount of inter-patient and intra-patient variability in the paranasal sinus anatomy. In addition, the proximity of the brain and orbits to the paranasal sinuses magnifies the complexity and potential risks of these procedures. It is essential that the operative surgeon have complete confidence as to the position of his instruments relative to the critical landmarks.

The purpose of this particular test was to assess the resident's ability to identify the location of the endoscope within the virtual anatomy.

Methods: The residents were first presented static views of four anatomic structures consisting of the nasopharynx, inferior meatus, maxillary sinus and the sphenoid sinus. The static views were intentionally made difficult, with few landmarks to gauge the position of the anatomic structure within the virtual anatomy. This particular task is not normally performed in actual surgery; however, simulated tasks such as these allow us to evaluate the importance of navigation and visualization of peripheral landmarks to spatial awareness. In addition to the four static anatomic structures, the residents were presented the same structures in a dynamic fashion. The endoscope was navigated to the landmark, and the residents were then asked to identify the structure in view.

Results: The fifth-year residents correctly identified 50% of both the static and dynamic spatial images. The structures that were missed in common were the inferior meatus and the nasopharynx. Interestingly, the sphenoid ostium was correctly identified in both the static and dynamic setting for both fifth-year residents.

The fourth- and third-year residents performed similarly to the fifth-year residents, correctly identifying 50% (4th year 63% and 3rd year 44%) of the spatial images. The senior residents had more difficulty with the static images, uniformly missing the sphenoid ostium. Navigation within the virtual anatomy allowed the senior residents to identify the majority of the spatial anatomic landmarks.

The second-year residents correctly identified 38% of the static and dynamic spatial images (2a 25% and 2b 50%). Both residents had most of their difficulty with the static images, missing as a group7 of 8 landmarks. Of the dynamic images the area that gave these residents the most difficulty was the inferior meatus.

The first-year residents correctly identified 25% of the static and dynamic spatial images. (1a 50% and 1b 0%). Resident 1a had most of his difficulty with the static spatial images, missing three out of four, representing 75% of his errors. The only dynamic spatial image he had difficulty with was the inferior meatus, with which 5 of 8 residents had similar difficulty; the inferior meatus was the most commonly missed dynamic spatial image. Resident 1b had difficulty with both static and dynamic images.

3.3.6 Overall Competency Scores and Simulation Experience

Each resident's component scores were summed to provide the evaluation team with a cumulative competency measure. The pattern of the total correct percentage scores mirrored, not surprisingly, the performance patterns across year of residency seen above for the individual component scores.

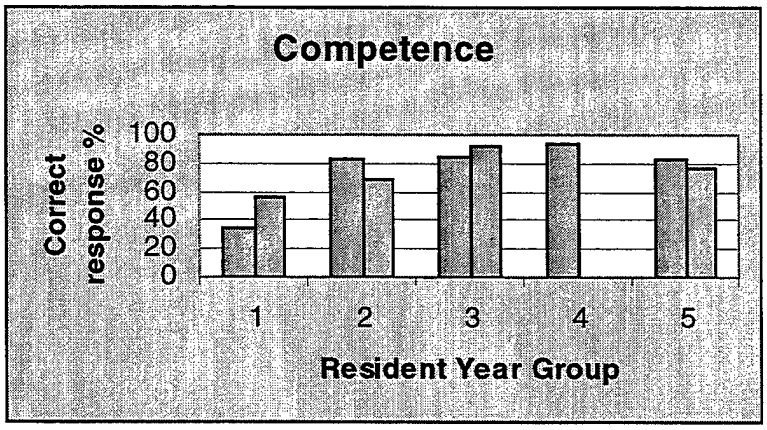


Figure 8. Histogram of Total Correct Response Percentage by Year of Residency

Although not statistically significant, the R3s and R4s performed better than the residents in the other year groups. Of particular interest to this evaluation study is the finding that the residents with the three highest scores on the competency evaluation also had 3 of the 4 highest cumulative simulation times for this group of 9 residents.

Indeed, when we look at the bivariate correlation between total score on the competency test and prior cumulative simulation time at the time of the test, we find a positive but non-significant relationship ($r = .540$). When we look at the relationship between prior simulation time and competency *component* test scores, we find a positive correlation with all components, but a statistically significant correlation with only Procedural Awareness ($r = .673$, $p < .05$). This finding is intriguing, suggesting perhaps that the simulator reinforces a standard procedure, or perhaps that procedural learning is more effective when the learner is "driving". These results suggest that further research may be warranted to validate these measures and to tease apart the transfer effects.

Finally, when we look at the interrelations of subscores on the Surgical Competency Test, we find the following:

- All of the subscores are significantly correlated with each other, except for Spatial Awareness, which is not correlated with any other subscore.
- The strongest subscore correlations are between Anatomical Recognition and Procedural Awareness ($r = .968$, $p < .01$) and between Anatomical Recognition and Tool Identification ($r = .808$, $p < .01$).

3.4 Relationships Among Surgical Performance Scores

When we examine the bivariate correlations among these surgical competency measures and the subjects' *scores* on the simulator, we find that overall competency test scores are significantly correlated with overall Novice trial scores ($r = .875$, $p < .01$), but *not* with Intermediate or Advanced model total scores ($r = .228$ and $.123$, respectively). In addition, several competency test subscores are significantly correlated with Novice trial score: Anatomical Recognition ($r = .915$, $p < .01$), Endoscopic Selection ($r = .769$, $p < .05$), and Procedural Awareness ($r = .956$, $p < .01$). Of the components of Novice trial score, novice Dissection scores were the strongest predictor of these surgical competency scores.

Tool Identification, while not significantly correlated with *overall* Novice trial score, was significantly correlated with some of its components: novice navigation accuracy ($r = .774$, $p < .05$), novice dissection time ($r = -.799$, $p < .05$), and novice dissection score($r = .805$, $p < .05$). Spatial Awareness was not significantly correlated with any of these simulator performance measures, but it is interesting to note that is was more strongly correlated with scores on the Intermediate and Advanced models than on the Novice model.

Looking at the intercorrelations among simulator scores for this group we find that Novice scores were predictive of Intermediate scores, but *not* of Advanced scores. Intermediate scores, on the other hand, were predictive of Advanced scores. However, the best Novice trial predictor of both Intermediate and Advanced scores was novice *hazard* score, even though novice hazard scores were not significantly correlated with any other novice score or with any competency score. This perhaps suggests that the most discriminating performance factor in the Novice model is the ability to avoid hazards, a skill that experienced surgeons would have acquired over time in the operating room.

Table 2 presents a matrix of bivariate correlations among the primary measures of interest for the 10 residents in the Phase 2 study: ESS Experience (as indicated by current year of residency), Prior Simulation Time (Version 1.2, relevant primarily for the Mean Video Ratings), Total Simulation Time (Version 1.2 plus Version 2.0), Scores on Version 2.0 (Novice, Intermediate, and Advanced), Mean OR Rating (by the staff proctor), Total Competency Test Score, and Mean Video Item Rating (for initial OR procedure). The reader should note that these correlations were generated using *pairwise* (rather than *listwise*) deletion, meaning that each correlation may represent only a subset of the total number of cases. The reader is encouraged to take into account the number of cases listed in the lower portion of the table for each pairwise comparison.

The correlations in Table 2 suggest the following:

- The strongest predictor of Mean OR Rating is Intermediate Model Score ($r = .959$, $p < .05$). The finding that simulation experience is a better predictor of these proctor ratings than is year of residency is also of interest, although year of residency is likely not the most robust measure of ESS experience.

Table 2. Correlations Among Major Simulation and Surgical Performance Measures

Correlations

		ESS Experience	Prior Simulation Time	Total Simulation Time	Novice trial score	Intermediate Model Score	Advanced Model Score	Mean OR Rating	Total Competency Test Score	Mean Video Item Rating
Pearson Correlation	ESS Experience	1.000	.372	.157	.698	-.188	-.264	.000	.655	.474
	Prior Simulation Time	.372	1.000	.877**	.706	.762	.662	.689	.540	.911
	Total Simulation Time	.157	.877**	1.000	.810*	.729	.624	.710	.561	.953*
	Novice trial score	.698	.706	.810*	1.000	.681	.592	.520	.875**	.779
	Intermediate Model Score	-.188	.762	.729	.681	1.000	.966**	.959*	.228	.748
	Advanced Model Score	-.264	.662	.624	.592	.966**	1.000	.874	.123	.546
	Mean OR Rating	.000	.689	.710	.520	.959*	.874	1.000	.579	.804
	Total Competency Test Score	.655	.540	.561	.875**	.228	.123	.579	1.000	.870
	Mean Video Item Rating	.474	.911	.953*	.779	.748	.546	.804	.870	1.000
Sig. (2-tailed)	ESS Experience	.	.289	.665	.054	.721	.613	1.000	.056	.526
	Prior Simulation Time	.289	.	.001	.050	.078	.152	.311	.133	.089
	Total Simulation Time	.665	.001	.	.015	.100	.186	.290	.116	.047
	Novice trial score	.054	.050	.015	.	.136	.215	.480	.010	.221
	Intermediate Model Score	.721	.078	.100	.136	.	.002	.041	.712	.252
	Advanced Model Score	.613	.152	.186	.215	.002	.	.126	.844	.454
	Mean OR Rating	1.000	.311	.290	.480	.041	.126	.	.421	.196
	Total Competency Test Score	.056	.133	.116	.010	.712	.844	.421	.	.130
	Mean Video Item Rating	.526	.089	.047	.221	.252	.454	.196	.130	.
N	ESS Experience	10	10	10	8	6	6	4	9	4
	Prior Simulation Time	10	10	10	8	6	6	4	9	4
	Total Simulation Time	10	10	10	8	6	6	4	9	4
	Novice trial score	8	8	8	8	6	6	4	7	4
	Intermediate Model Score	6	6	6	6	6	6	4	5	4
	Advanced Model Score	6	6	6	6	6	6	4	5	4
	Mean OR Rating	4	4	4	4	4	4	4	4	4
	Total Competency Test Score	9	9	9	7	5	5	4	9	4
	Mean Video Item Rating	4	4	4	4	4	4	4	4	4

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

- The strongest predictor of Total Competency Test Score is Novice Model Score ($r = .875$, $p < .01$).
- The strongest correlate of Mean Video Item Rating is Total Simulation Time ($r = .969$, $p < .01$), although simulator experience prior to the videotaped procedure approaches significance ($r = .911$).
- Total Simulation Time is highly correlated with Prior Simulation Time (as one would expect) and with Novice Model Score ($r = .810$, $p < .05$), perhaps suggesting improved simulation scores with practice.
- Intermediate Model Score is highly predictive of Advanced Model Score ($r = .966$, $p < .01$), perhaps indicating its value as a training scenario.

3.5 Effectiveness of Simulator Features

Seven of the 10 residents and 3 of the 5 staff ENT subjects completed a post-test questionnaire following their simulation training. Details of the questionnaire structure are provided in our earlier report, but its general objective was to assess the perceived usefulness of each of the components of the system, as well as the perceived training utility of the system as a whole. In addition, those subjects who had participated in the Phase 1 evaluation of Version 1.2 were asked to evaluate the changes to the system incorporated in Version 2.0. Results of this latter survey are presented in a subsequent section.

Although our relatively small pool of available ENT residents did not permit us to perform a full factorial design analysis of each component of the training system, we once again asked for subjective evaluations of their pedagogical utility. As with our findings in the previous phase, this cohort of residents indicated the following general assessments:

- Overall, the simulator was judged to be extremely useful for surgical training.
- The anatomical model was judged to be highly realistic, but several subjects suggested that the major anatomical landmarks and dissectable regions should be made more distinct for training purposes.
- Anatomy of lamina papyracea and ethmoid cells could be improved.
- Movement of the scope and surgical instruments was not as smooth as in the real procedure.
- Haptic feedback could be better and would also be useful on the endoscope itself, but was preferred on the surgical instrument, if a choice were required.
- Some subjects preferred the forces off during dissection of the anterior ethmoid cells, due to restricted access.
- The simulator was thought to be most useful for teaching anatomy, and for training hand-eye coordination, navigation skills, elementary dissection skills, and selection of appropriate instruments.
- Greater depth of field (i.e., more distant cut-off plane) was judged to be desirable.
- A seated operative posture was preferred to the current standing posture.

- The ability to slip through the virtual anatomy with the endoscope is disconcerting.
- Voice recognition was judged to be useful, but some subjects had difficulty with interpretation of their voice inputs.
- Some subjects experienced a slight delay in the update of the graphics, according to instrumented endoscope movement.
- Some subjects expressed the desire to have the ability to move the endoscope closer to the tool to get a more comfortable view of the anatomy.

These results were consistent with our earlier findings and are not discussed in further detail here.

3.6 Evaluation of Version 2.0 Enhancements

In Phase 1, several aspects of the ESS simulator were recommended for further development effort: patient model enhancement, hardware upgrades, software revisions, novice model changes, and haptic refinements. Working in close collaboration with Dr. Edmond, we compiled an expert assessment of these enhancements. These findings are summarized below.

3.6.1 Patient Model Enhancements

Recommendations were made for model enhancements that revolved around increasing the number of polygons to take advantage of the newer hardware , as well as segmentation of additional features. Throughout Phase 1, rendering lag was a major concern, and so the development team focused on limiting the number of polygons in the model. With the upgrade to a more powerful graphics engine they were able to increase the geometric complexity of the model to achieve greater realism.

Additional anatomic features were considered essential to performing a more complete dissection, and to increase the simulator's training effectiveness. Model enhancement consisting of both normal and variant anatomy added greatly to the overall depth of the trainer. Model enhancements within the following areas are assessed below: nasopharynx, sphenoid sinus, skull base, periorbital region and miscellaneous

Nasopharynx: Refinement within the virtual nasopharynx increased the model realism. The Eustachian tube was remodeled and additional depth was added at the inferior aspect of the nasopharynx, providing the illusion that the anatomy continued inferiorly. Furthermore, the nasopharynx was expanded laterally around the posterior choane, giving the proper perspective that one observes in live surgery.

Sphenoid sinus: Significant refinements were made within the sphenoid sinus to mimic the real anatomy. The addition of the carotid artery, optic nerve and pituitary indentations within the walls of the sphenoid sinus were critical spatial landmark additions that allowed the student to identify which sinus he/she had approached. Additionally, the presence of these structures provides the student with the capacity to entertain procedures outside the original scope of the simulator (e.g., Endoscopic hypophysectomy).

Skull base: The model's most significant advance was in the area of the skull base. With the addition of the agger nasi cell, nasofrontal duct, anterior ethmoid artery, and frontal sinus, a complete dissection of the sinuses became a reality. For training purposes, these additions were felt to be crucial toward creating a complete trainer.

Periorbital region: Orbital refinements were noteworthy, as well. Creation of a dissectable lamina, periorbita, periorbital fat, muscle and nerve were features that increased the complexity of the model and allowed the student to experience complications of sinus surgery that were not present within the Phase 1 model. It should be noted, however, that some of the experienced surgeons commented on the poor distinction between the lamina and the ethmoids.

The addition of a nasolacrimal duct introduces the potential for training students in a newly developed procedure, endoscopic dacryocystorhinostomy. However, additional effort toward making the lateral wall of the nasal cavity more dissectable will be necessary to make this procedure a viable addition to the list of simulated procedures.

Miscellaneous: Considerable effort was applied in the current phase toward increasing the total amount of anatomy that is dissectable and deformable. At the conclusion of Phase 1, it was surgically frustrating that anatomy traditionally dissectable within a normal surgery was not dissectable in the virtual sinus. This need has been accommodated in Version 2.0, and essentially all of the anatomy that is reasonably dissected in standard surgery is now dissectable within the virtual patient. Of additional concern was the number of fragments that remained after dissecting the ethmoid cavities. All of the free-floating fragments were found to be dissectable in the refined model. The uncinate process was made a movable object, and dissectable superiorly as well as anteriorily.

Pathologic variations were additional model enhancements that were recommended at the conclusion of Phase 1, and which became a reality in Version 2.0. The addition of a mucocele, dissectable nasal polyps, and an enhanced trauma model expanded the student's experience with additional abnormalities.

3.6.2 Hardware Upgrades

Phase 2 projections were to increase the polygon count and size of the virtual model. These anticipated increases in the complexity of the virtual model warranted upgrades to the

hardware. Computational limitations noted in Phase 1 resulted in less visual realism and perceptable latency when operating in the anatomically dense areas of surgical dissection (e.g., the ethmoid cavity). Several hardware modifications significantly increased real-time performance without sacrificing realism or model size and complexity. The addition of the Silicon Graphics Inc (SGI) Onyx 2 with Infinite Reality Graphics made an impact that was clearly noticeable to the clinical evaluation team. The new hardware increased the potential model complexity by 300%, more than adequately accommodating the new model. This feature allowed for less decimation and more polygons, with an overall effect of increased visual realism while maintaining real-time interaction rates.

3.6.3 Software Revisions

In Phase 1 the clinical evaluation team indicated a desire for greater interactivity and a closer resemblance to the operative experience. The addition of voice recognition and a "Virtual Instructor" accommodated this request.

Voice recognition software implemented with Dragon Dictate on a PC was deployed, allowing control of the simulator in response to commands spoken by the student. This particular enhancement was designed to mimic the communication that the surgeon has with the surgical technician when requesting surgical instrumentation. It allows for hands-free training when an assistant or proctor is not available. While voice recognition was found to be a positive addition to the simulator, detection inaccuracies across subjects diminished its popularity. The occasional delayed response was disruptive of the expected flow one would experience during a normal operative sequence. It should be noted that placement of the microphone is somewhat sensitive and voice recognition can fail if the speaker becomes excited or does not pause between commands. With additional training, the voice control option should be a useful component of the system

The addition of the "Virtual Instructor" was considered unique. The clinical team found this component to be extremely helpful in providing the student with direction when encountering difficulty with either procedural steps or anatomic recognition. This addition is analogous to the proctor in the operating room providing surgical advice during a procedure. In practice this component was not used as much as anticipated, however, primarily because a proctor was always present in the scenario in which the residents and staff were evaluated. We expect that it will be used more extensively in fully deployed training systems.

3.6.4 Novice Model Improvements

Improvements to the novice model were significant. Additional tools were introduced to the user, and provided the subject with a complete array of instruments used in endoscopic sinus surgery. A number of the new instruments were tools that are used for frontal recess

surgery, which complemented the extension of the intermediate and advanced models to include the frontal recess, lamina, skull base frontal duct, agger nasi cell and ethmoid arteries.

The tool specific dissectable objects were modified in the novice model to include adjacent obstacles (hazards) which challenged subjects to restrict their dissection to the confines of the dissectable object. The new hazards were designed to provide more realistic training in instrument control. Although time did not permit a detailed analysis of the instrument path data, we anticipate that it will show tighter control over dissection instruments than in Version1.2.

3.6.5 Haptic Refinements

One of the more challenging areas for evaluation was that of haptic system fidelity. Experienced and non-experienced surgeons' concerns were most vocal with regard to this feature during Phase 1. Considerable thought was initiated between phases in trying to decide if the force feedback should be on the endoscope as opposed to the tool. After a long debate and careful review of the experienced surgeons' comments, it was elected to continue with the forces on the tool. Ideally, forces would be delivered to both the endoscope and the forceps; however, this configuration was dismissed as too costly for this stage of development.

In addition, the placement of forces on two points along the shaft of the tool with interpolated forces between the points was considered desirable but not practical, given the constraints of the present configuration and budget. Throughout Phase 2, haptic fidelity continued to be an issue. In general the responses were mixed. For some tasks, the user preferred to have the forces on, but for some tasks, having the forces off was the desirable mode. The "feel" of surgery was found to be elusive. Improvements in force feedback were made in various areas, but significant gains overall were not witnessed.

3.7 Additional Observations by the Technical Proctor

Haptics

The technical (non-MD) proctor noted that, overall, there was significant improvement in the haptics in this version of the simulator. During the evaluation trials, subjects who had used the previous version of the simulator noted a dramatic improvement during Injection and initial Dissection.

Collision between the endoscope and instrumented forceps was often confused for haptic feedback from the system. Subjects would often mistake this interaction to be haptic feedback hindering them from entering an open area with the forceps, therefore falsely

indicating to them that the haptics were failing during the latter part of dissection in the anterior and posterior ethmoid cells. Because of this interaction, many of the residents who were also subjects in Phase 1 did not report any improvement in the haptic system, and one of the staff even considered the haptics to be more of a hindrance than a mechanism to better simulate a real procedure.

While it was unanimously noted that the ethmoid cells were better represented graphically in this version of the system, some subjects preferred to have the haptics inactivated during dissection of the ethmoid cells, allowing them greater freedom of movement with the forceps and endoscope. When the haptics were deactivated, subjects were allowed to penetrate the virtual anatomy with the tip of the instrumented forceps. This allowed them clearance from the encoders on the instrumented forceps, which was needed to obtain a normal view of the virtual anatomy with the endoscope. It should be noted that the unrealistic forces in the ethmoid area were corrected in the final release of software, which is currently available at Madigan, but was not available at the time of this study.

Finally, it was suggested by several subjects that 6 degree-of-freedom haptics be placed on both the forceps and the endoscope, and that the algorithms for the haptics include greater distinctions between bone and cartilage.

Familiarity with the Virtual Anatomy

Familiarity with the virtual anatomy played a major roll in the subject's recognition of active areas for Dissection. In contrast, the Navigation and Injection subtasks did not appear to be as affected by the subject's familiarity with the anatomy. Active areas of dissection for the graphical model of the anatomy needed to be pointed out to most subjects during their trials on the simulator. The anatomical structures which were judged to be most difficult to distinguish were the lamina papyrecea and the ethmoid cells.

Although, overall, the subjects thought that the new anatomical models provided a more realistic representation of the actual anatomy, they expressed a need for greater distinction between active areas and inactive areas. Most of the subjects complained of too short a viewing frustrum for the endoscope. This may have contributed to their need for the proctor's help in identifying the active areas of dissection.

Ability to "Push Through" the Virtual Anatomy During Trials

Because of system cost limitations, the instrumented endoscope and the shaft of the forceps were not outfitted with 6 DOF haptics. Because of this, subjects were able to unrealistically pass the viewpoint of the endoscope, the shaft of the endoscope, and the shafts of the virtual tools through the virtual anatomy.

In the previous version of the simulator, this was seen by most staff subjects as beneficial, forcing the residents to exercise greater hand eye coordination in maneuvering the

endoscope and tools within the virtual anatomy. Perhaps because of the significant improvements made in the graphical model, haptics and tools in Version 2.0, the benefit of exercising greater hand-eye coordination was not weighted as heavily as the possibility of the simulator obtaining full operating room realism.

The majority of the repeat subjects' comments were concentrated on how each aspect of the simulator differed from the real operating room, rather than what limited operating room skills residents could gain from using the simulator. This may indicate that these subjects were more directly correlating the simulated procedures to the actual OR procedures.

4 Summary of Findings

Our Phase 2 evaluations consisted of two general areas of emphasis: an assessment of changes made to the simulator since Phase 1, and an assessment of the impact of experience with the Madigan ESS simulator on the overall goal of surgical training effectiveness.

Evaluation of Version 2.0 Modifications

Our evaluation of the simulator modifications was conducted on an iterative basis, as in Phase 1, and is summarized in this report. Evaluation of the Version 2.0 enhancements indicated that:

- Model enhancement consisting of both normal and variant anatomy added greatly to the overall depth of the trainer.
- Essentially all of the anatomy that is reasonably dissected in standard ESS surgery is now dissectable within the virtual patient in Version 2.0.
- Graphics hardware upgrades allowed more complex anatomy to be displayed at real-time rates.
- Surgeons would still prefer a longer yon plane, but not at the expense of anatomical realism.
- Voice recognition was judged to be a valuable feature but will take some further adjustment and training to be optimally useful.
- The Virtual Instructor was considered a valuable feature, but was not fully tested in this phase, since proctors were always available during data collection sessions.
- New tools and tool-specific hazards were considered valuable additions to the Novice model.
- Improvements in force feedback were made in several areas, but haptic fidelity continues to be an issue for further development.

As with our Phase 1 findings, the overall response from these ENT subjects was overwhelmingly positive. The simulator was thought to be a very useful tool and should be

considered an integral part of Otolaryngology training programs. Continued work on visual realism, force feedback and model enhancement were recommended.

Transfer of Training Evaluation

The primary goal of our Phase 2 evaluation was to determine the impact of ESS simulator training on surgical effectiveness. To address this issue we developed a set of assessment methods, including real-time attending proctor rating scales, endoscopic video assessment scales, and several novel methods of evaluating components of surgical competence.

In their attempt to assess the impact of simulation experience on OR performance, Drs. Edmond and Mesaros encountered moderate difficulty striving to generate the appropriate cases for evaluation with the junior residents, for whom the biggest impact would be expected. In addition, controlling for operative exposure and experience across residents proved to be an unexpected challenge. Nonetheless, a number of promising trends emerged from the data:

- Intermediate Model Score on the simulator is a strong predictor of OR performance as measured by our proctor rating methods.
- Two somewhat independent factors appear to emerge from the rating scale used to perform blind assessments of initial OR performance: "scope and instrument control," and "dissection skill." Surgical Confidence is most highly correlated with the second factor, indicating perhaps that it is based upon experience with successful dissection.
- Prior simulation training on Version 1.2 appears to have positively affected initial OR performance ratings, as judged by senior ENT surgeons viewing anonymous endoscopic videotapes of those procedures. Strongest effects were found for Tool Manipulation and Surgical Confidence.
- Mean ratings of endoscopic videotapes by non-MD observers showed strong inter-rater reliability, and may prove useful as an adjunct independent measure of surgical performance in future studies.
- Several new assessment techniques were developed to measure individual components of surgical competency. While still a potential contribution to the field of surgical competency evaluation, future studies should perhaps be modified to use test materials generated from real endoscopic footage rather than simulated footage, to more directly address transfer of training to the clinical scenario.
- Results from a group of nine residents followed a consistent trend which suggested transfer of training from the ESS simulator to these components of surgical competency; the component most strongly correlated with prior simulation experience was Procedural Awareness.
- Novice trial score on the simulator was a significant predictor of overall competency score. Of the components of Novice trial score, novice Dissection scores were the strongest predictor of these surgical competency scores.

Given the limited population of available residents, the intent here was principally to establish a framework for evaluating transfer of simulation training to real-time OR performance and to several aspects of surgical competency as it relates to endoscopic sinus surgery. Objective measures were generated for evaluation of transfer and competency, and protocols were developed to control for extraneous variables, where possible. In addition to the findings derived from the pilot studies detailed in this report, these methods may offer a contribution to the field of medical simulator evaluation.

References

Baillie J., et al, The Future of Endoscopy Simulation: A Duke Perspective, Endoscopy, 24, 1992.

Bard, C., Simulation Modeling of the Colon; First International Symposium on Endoscopy Simulation, World Congresses of Gastroenterology, Sydney, 1990.

Bartlett, Worley, Simpson, Aylward, Bailey, Fish. (1995) Report out to the Modeling and Simulation Working group, Modeling and Simulation Benefits Task Force, Defense Modeling and Simulation Office, Alexandria, VA.

Billinghurst, M., Savage, J., Edmond, C., Oppenheimer, P. The Expert Surgical Assistant: An Intelligent Virtual Environment with Multimodal Input, Medicine Meets Virtual Reality IV, 1996.

Bostrom, M., Singh, S.K., Wiley, C.W., Design of an Interactive Lumbar Puncture Simulator with Tactile Feedback, IEEE Annual Virtual Reality Symposium, 429-435, 1993.

Burnside, Billy, U.S. Army Research Institute. Assessing the Capabilities of Training Simulations: A Method and Simulation Networking (SIMNET) Application, 1990.

Caird, J. (1996) Persistent Issues in the Application of Virtual Environment Systems to Training, IEEE Computer Society: Human Interaction with Complex Systems, Vol. 3, 124-132

Chen, D., Rosen, J., Zeltzer, D., Surgical Simulation Models: From Body Parts to Artificial Person, Proc. Image VI Conference, July 1992.

Cohen, MacRae, Jamieson. (June 1996) Teaching Effectiveness of Surgeons. The American Journal of Surgery, Vol 171, 612-614

Cover, S. et al., Interactively Deformable Models for Surgery Simulation; IEEE CG&A, 13,6,68-75, 1993.

Dawson et al., An Interactive Training/Simulator for Interventional Radiology, J. Vascular and Interventional Radiology, 7(1, suppl 2): 132-136, 1996.

Edmond, C.V., Heskamp, D., Sluis, D., Stredney, D., Sessanna, D., Wiet, G., Yagel, R., Weghorst, S., Oppenheimer, P., Miller, J., Levin, M., and Rosenberg, L., ENT endoscopic surgical training simulator, *Proceedings of Medicine Meets Virtual Reality V*, San Diego, CA, January 1997.

Geis, Kim, McAfee, Kang, Brennan. (March 1996) Synergistic benefits of combined technologies in complex, minimally invasive surgical procedures. Surgical Endoscopy, Vol 10, 1025-1028

Gillies, D. et al., Computer Simulation for Teaching Endoscopic Procedures, Endoscopy, 24, 1992.

Good, M.L., Gravenstein, J.S., Training for Safety in an Anesthesia Simulator, Seminars in Anesthesia, 12, 4, 1993.

Gupta, S. et al., Introduction of New Technologies in the Medical Undergraduate Curriculum, Medicine Meets Virtual Reality 4, 1996.

Heskamp, D.,, C. Edmond, S. Weghorst et al., ENT Surgical Simulator, Final Report, Cooperative Agreement No. DAMD17-95-2-5023, September 1997.

Hoffman, H.M., Murray, M., Irwin, A., McCracken, T., Developing a Virtual Reality-Multimedia System for Anatomy Training, Medicine Meets Virtual Reality 4, 1996.

Hon, D., Ixion's Laparoscopic Surgical Skills Simulator, Medicine Meets Virtual Reality II, 1994.

Hon, D., Medical Reality and Virtual Reality, Medicine Meets Virtual Reality 4, 1996.

Johnson R., Bhoyrul, S., Way, L., Satava, R., McGovern, K., Fletcher, J.D., Rangel, S., Loftin, R.B., Assessing a Virtual Reality Surgical Skills Simulator, Medicine Meets Virtual Reality II, 1994.

Kaufmann, D.M., Bell, W. Teaching and Assessing Clinical Skills Using Virtual Reality Medicine Meets Virtual Reality: Global Grid, 1997.

McGovern, K.T. and McGovern, L.T., The Virtual Clinic, A Virtual Reality Surgical Simulator, Medicine Meets Virtual Reality II, 1994.

Merril, J. Virtual Reality in Surgery and Medical Education, Medicine Meets Virtual Reality II, 1994.

Peifer, J., et al., Virtual Environment for Eye Surgery Simulation, Medicine Meets Virtual Reality II, 1994.

Pieper, S., CAPS: Computer-Aided Plastic Surgery, PhD Thesis, MIT, 1989.

Pieper, S., et al., A virtual environment system for simulation of leg surgery, Stereoscopic Displays and Applications II, 1991.

Poon, A., Williams, C., Gillies, D., The Use of Three-Dimensional Dynamic and Kinematic Modeling in the Design of a Colonoscopy Simulator, New Trends in Computer Graphics, Springer Verlag, 1988.

Rice, D., and Schaefer, S. Endoscopic Paranasal Sinus Surgery, Raven Press, NY. 1988.

Robb, R.A.; Virtual Endoscopy: Evaluation Using the Visible Human Datasets and Comparison with Real Endoscopy in Patients, Medicine Meets Virtual Reality: Global Healthcare Grid, 1997.

Rohen, J., Yokochi, C., Color Atlas of Anatomy, Igaku-Shoin, New York, 1983.

Rosenberg, L., Stredney, D., A Haptic Interface for Virtual Reality Simulation of Endoscopic Surgery, Medicine Meets Virtual Reality IV, 1996.

Rosse, C. The American Journal of Surgery, Vol 171. 612-614

Rosser J., Rosser L., Savalgi, (Feb 1997) Skill Acquisition and Assessment for Laparoscopic Surgery, Archives of Surgery, Vol. 132., 200-204

Sagar, M.A., et al., Virtual Environment and Model of the Eye for Surgical Simulation, Computer Graphics Proceedings, 205-212, 1994.

Satava, R., Doorway to the Future, Medicine Meets Virtual Reality: Global Healthcare Grid, 1997.

Stredney, D. et al., An Virtual Simulation Environment for Learning Epidural Anesthesia, Medicine Meets Virtual Reality 4, 1996.

Weghorst, S., Airola, C., Oppenheimer, P., Edmond, C.V., Patience, T., Heskamp, D., and Miller, J., Validation of the Madigan ESS simulator, *Proceedings of Medicine Meets Virtual Reality VI*, San Diego, CA, January 1998.

Wickham, J.E.A., (Jan 1994), Minimally Invasive Surgery: Future Developments, The British Medical Journal, Vol 308(6922), 193-196.