

AWARD NUMBER: W81XWH-11-C-0011

TITLE: Automated Support for da Vinci Surgical System

PRINCIPAL INVESTIGATOR: Jeffrey Berkley, Ph.D.

CONTRACTING ORGANIZATION: Mimic Technologies, Inc.  
Seattle, WA 98104

REPORT DATE: May 2011

TYPE OF REPORT: Final, Phase I

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

DISTRIBUTION STATEMENT: Approved for Public Release;  
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

<b>REPORT DOCUMENTATION PAGE</b>			<i>Form Approved</i> <i>OMB No. 0704-0188</i>		
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. <b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b>					
<b>1. REPORT DATE</b> 1 May 2011		<b>2. REPORT TYPE</b> Final, Phase I		<b>3. DATES COVERED</b> 1 Nov 2010 – 30 Apr 2011	
<b>4. TITLE AND SUBTITLE</b>  Automated Support for da Vinci Surgical System			<b>5a. CONTRACT NUMBER</b> W81XWH-11-C-0011		
			<b>5b. GRANT NUMBER</b>		
			<b>5c. PROGRAM ELEMENT NUMBER</b>		
<b>6. AUTHOR(S)</b>  Jeffrey Berkley, Ph.D.  E-Mail: jeff@mimicsimulation.com			<b>5d. PROJECT NUMBER</b>		
			<b>5e. TASK NUMBER</b>		
			<b>5f. WORK UNIT NUMBER</b>		
<b>7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)</b> Mimic Technologies, Inc. Seattle, WA 98104			<b>8. PERFORMING ORGANIZATION REPORT NUMBER</b>		
<b>9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)</b> U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			<b>10. SPONSOR/MONITOR'S ACRONYM(S)</b>		
			<b>11. SPONSOR/MONITOR'S REPORT NUMBER(S)</b>		
<b>12. DISTRIBUTION / AVAILABILITY STATEMENT</b> Approved for Public Release; Distribution Unlimited					
<b>13. SUPPLEMENTARY NOTES</b>					
<b>14. ABSTRACT</b> While adoption of da Vinci systems has been rapid worldwide, there exists a wide variance in surgical procedure performance impacting care quality, cost and patient safety negatively, due in part to inefficient training practices and limited mechanisms for objectively assessing surgical performance. To address the demand for improved training, Mimic Technologies has developed da Vinci simulators, the dV-Trainer and da Vinci Skills Simulator, in collaboration with Intuitive Surgical, which collect and analyze diverse performance data, and then recommend steps for addressing identified weaknesses. Conversely, Johns Hopkins University's Surgical Assistant Workstation collects and analyzes data from da Vinci during training and surgery. In this Phase I feasibility study, we developed a proof-of-concept working prototype framework capable of providing continuous surgical skills assessment and decision support throughout initial training and thereafter during surgery with the ultimate goal of accelerating a surgeon's acquisition of da Vinci surgical skills. Preliminary analysis of training tasks (anastomosis and peg transfer) across simulation and phantom training laboratories performed by the developed prototype framework highlights the promise of cross-platform data collection and performance measurement across training curriculum, and the opportunity for designing dynamic, customized curricula to a trainee's needs. Developed prototype framework is also a first effort towards cross-platform assessment of robotic surgical performance. This work explored several new concepts including a distributed system for decision support, the role played by instrument and hand pose in planning surgical tasks and metrics assessing this role, distinction of man-machine skills from surgical and task skills, and refinement of simulation training environments to match the corresponding real world training. Preliminary results show that selected performance metrics previously validated in simulation environment continue to hold in training exercises with a real robotic system. The corresponding Phase II effort will develop a full featured and validated prototype of the decision support system including its initial installation, validation, and use in surgical training at selected training centers.					
<b>15. SUBJECT TERMS</b> da Vinci, dV-Trainer, SAW, robotic surgery, simulation, automated assessment, resident training, NHIN, anastomosis, needle driving, peg transfer					
<b>16. SECURITY CLASSIFICATION OF:</b>			<b>17. LIMITATION OF ABSTRACT</b>	<b>18. NUMBER OF PAGES</b>	<b>19a. NAME OF RESPONSIBLE PERSON</b> USAMRMC
<b>a. REPORT</b> U	<b>b. ABSTRACT</b> U	<b>c. THIS PAGE</b> U			<b>19b. TELEPHONE NUMBER</b> (include area code)
			UU	35	

## Table of Contents

INTRODUCTION .....	7
RESULTS OF PHASE I RESEARCH .....	8
Overall Framework for Collection and Analysis of Surgeon Performance Data (Task 1).....	9
Surgeon Performance Data Collection, Local Storage & Display (Tasks 1 & 2).....	11
Instrument Motion File .....	11
Mimic Instrument Motion File Generator & Synchronizer .....	12
JHU Decision Support Metrics Generator & dV-Trainer Metrics Generator.....	12
Exercise Report File.....	15
Local Database.....	15
Performance Evaluation & Display .....	15
Surgeon Performance Data Archival (Tasks 1 & 3).....	17
Database Server .....	17
Server/client Software for Data Transfer to/from the Server.....	18
Training Exercises (Task 4).....	19
End-to-End Anastomosis .....	19
Peg Transfer.....	20
Data Collection & Analysis (Task 5).....	22
Preliminary Experiments .....	22
Large Scale Data Collection .....	24
Work Plan for Phase II Development (Task 6).....	28
KEY RESEARCH ACCOMPLISHMENTS .....	29
Discoveries.....	29
Decision Support Analysis Can Differentiate Non-proficient Surgeons from Proficient Surgeons.....	29
Man-machine Interface Effects in the dV-Trainer Compared to the <i>da Vinci</i> Surgical System .....	29
Additional Deliverables .....	30
Proficiency-based Scoring and Curriculum Development .....	30
Reportable Outcomes.....	31

W81XWH-11-C-0011 Automated Support for *da Vinci* Surgical System - Final Report

CONCLUSION..... 32  
    Research Impact and Applications..... 33  
REFERENCES ..... 35  
APPENDICES ..... 36  
PUBLICATIONS MADE AS A RESULT OF THE RESEARCH EFFORT ..... 36  
    Towards validation of robotic surgery training assessment across training platforms..... 36  
    Abstract..... 37  
PERSONNEL RECEIVING SALARY FROM THE RESEARCH EFFORT ..... 37

## INTRODUCTION

Robotic-assisted surgery is becoming commonplace in medicine. In 2010, approximately 278,000 surgical procedures were performed using the *da Vinci*® Surgical System (a product of Intuitive Surgical, Inc.), which is up 35% from 2009 [1]. Over 1,840 *da Vinci* Surgical Systems systems are in use in over 1,500 academic and community hospitals worldwide with 1,344 robots utilized in the US alone. However, even though the use of surgical robots is growing extremely rapidly, there is no consistent, scientifically accepted format for evaluating robotic-assisted surgical skills. Not only is it important to unify methodologies that promote skills acquisition during training, but there is also a need for systematic assessment of operative performance. The objective of the research effort is to address such needs with an automated system that collects and analyzes diverse data from training and surgery; identifies variances in training and operative care; and, provides clinically relevant decision support for recommending follow-on training that would improve surgical performance and outcomes.

During our Phase I effort, we designed and implemented a proof-of-concept working prototype of an overall framework that collects and analyzes surgeon performance data from simulation (dV-Trainer) and phantom laboratory (*da Vinci*) exercises. Two selected tasks - End-to-end anastomosis and a peg board ring manipulation module - were implemented in simulation and real phantom laboratories for surgeon performance assessment across training platforms based on a defined set of cross-platform metrics. This allowed us to begin collecting preliminary surgeon performance data in order to test our data collection framework. Collected data from different platforms was uploaded to established Mimic Technologies servers and shared between Mimic and Johns Hopkins for further detailed analysis. Preliminary performance evaluation results compared the users' performance to that of proficient users or average user performance and identify the user's skills with deficient performance.

This work describes the first prototype for measurement and assessment of robotic surgery training data across real and simulated training platforms. We studied, for the first time, common performance metrics for the same training tasks across two platforms in experiments performed by users on the *da Vinci* system and the dV-Trainer. The preliminary computations and evaluation results show trends similar to larger studies on individual platforms, and that the proficient and non-proficient users are differentiable using the studied metrics. We also show that performance metrics of training exercises previously validated in simulation environments hold in training exercises with a real robotic system.

This work included significant research findings such as a first archive of cross-platform robotic surgery training data, preliminary evaluation of common metrics across two platforms, and detection of variability in the man-machine interfaces of the two platforms that must be accounted for in any further analysis. This work explored several new concepts including a distributed system for decision support, the role played by instrument and hand pose in planning

surgical tasks and metrics assessing this role, distinction of man-machine skills from surgical and task skills, and refinement of simulation training environments to match the corresponding real world training. Preliminary results show that selected performance metrics previously validated in simulation environment continue to hold in training exercises with a real robotic system.

## RESULTS OF PHASE I RESEARCH

This Phase I feasibility study focused on the development of methods for collecting, sharing and analyzing experimental data across minimally invasive robotic surgery and training platforms - namely the *da Vinci* Surgical System and the dV-Trainer training platform. We conducted a detailed design and feasibility study, which roadmaps the development of an automated surgical support prototype for the *da Vinci* Surgical System and the *da Vinci* simulation. We developed a proof-of-concept working prototype and demonstrated the realization of technical objectives and functionality. Current results support the hypothesis that the development and implementation of a fully-functional automated support system is highly feasible.

The Phase I prototype demonstrated accomplishment of the planned technical tasks, in particular:

1. Design of an overall framework for collection and analysis of surgeon performance data from the dV-Trainer and from *da Vinci* Surgeon Consoles (*da Vinci* Skills Simulator and *da Vinci* Surgical Systems with phantom laboratories)
2. Extension of the JHU SAW open-source data format for performance data representation, and development of tools for exporting proprietary format data from experimental platforms into the open-source format
3. Establishment of an initial server-based system for managing the data collected across training platforms at Mimic and JHU
4. Development of simulated training tasks (anastomosis and peg transfer), and corresponding real phantom experimental environment for performing cross-platform measurements
5. Preliminary assessment and validation of task performance metrics for anastomosis and peg transfer data collected from the dV-Trainer and real phantom environment
6. Design of a work plan for Phase II development

We designed an open source, standards compliant framework (Figure 1) for collection and analysis of task performance data across *da Vinci* systems and its simulators. A preliminary version of this framework has been prototyped to store and analyze task performance data acquired from the Johns Hopkins' *da Vinci* S Surgical System, and dV-Trainer systems at Mimic Technologies.

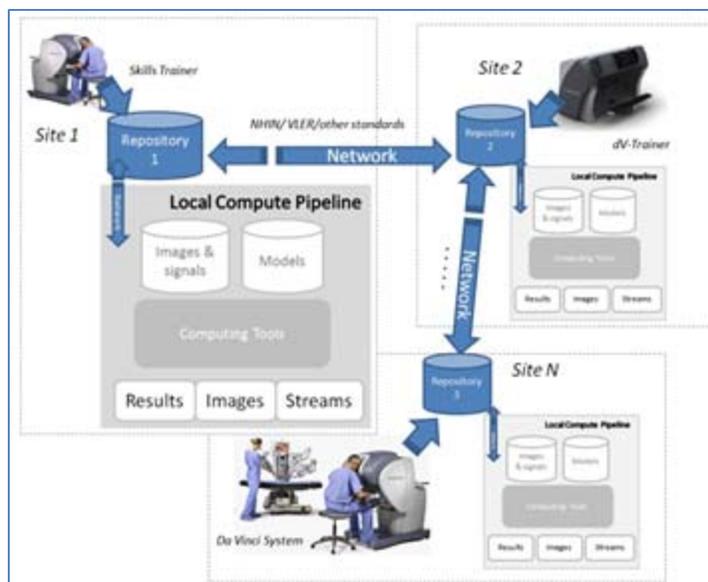


Figure 1 - Overview of the distributed decision support architecture for *da Vinci* systems and its simulators.

## Overall Framework for Collection and Analysis of Surgeon Performance Data (Task 1)

Prototype methods have been designed and implemented for collection and analysis of surgeon performance data in real phantom training exercises performed on *da Vinci* S system and in simulation exercises performed on dV-Trainer and *da Vinci* Skills Simulator.

Figure 2 outlines the architecture and task work flow for a real phantom laboratory exercise. A JHU archival workstation captures stereo-video and instrument motion data from the *da Vinci* in the SAW data format, and the archived data are post-processed by a set of custom-developed applications to compute cross-platform metrics using common analysis methods. Computed metrics can then be loaded in to the local database of the framework and performance evaluation is displayed via the developed graphical user interface (GUI).

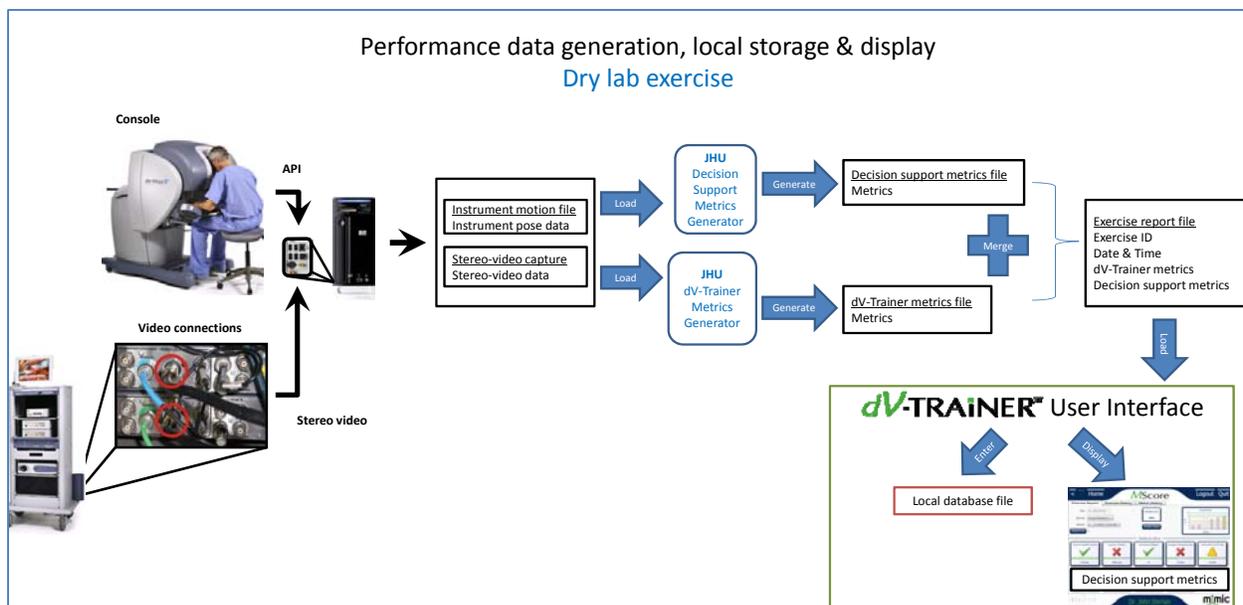


Figure 2 - Prototype for performance data generation, local storage and display in real phantom training laboratory exercise on *da Vinci*

Figure 3 outlines the corresponding architecture and task flow for a simulation exercise performed on a dV-Trainer or *da Vinci* Skills Simulator. A JHU archival workstation captures stereo-video and the dV-Trainer generates the simulation log file that includes a wide range of application programming interface (API) data collected during the performance of the exercise. The motion data in the log file is converted to instrument motion data. The log file, stereo-video and converted instrument motion data are then analyzed by the dV-Trainer and by a custom-developed application to compute cross-platform metrics using common analysis methods as in the previous work flow. Similarly, computed metrics can then be loaded in to the local database of the framework and performance evaluation is displayed via the developed graphical user interface (GUI).

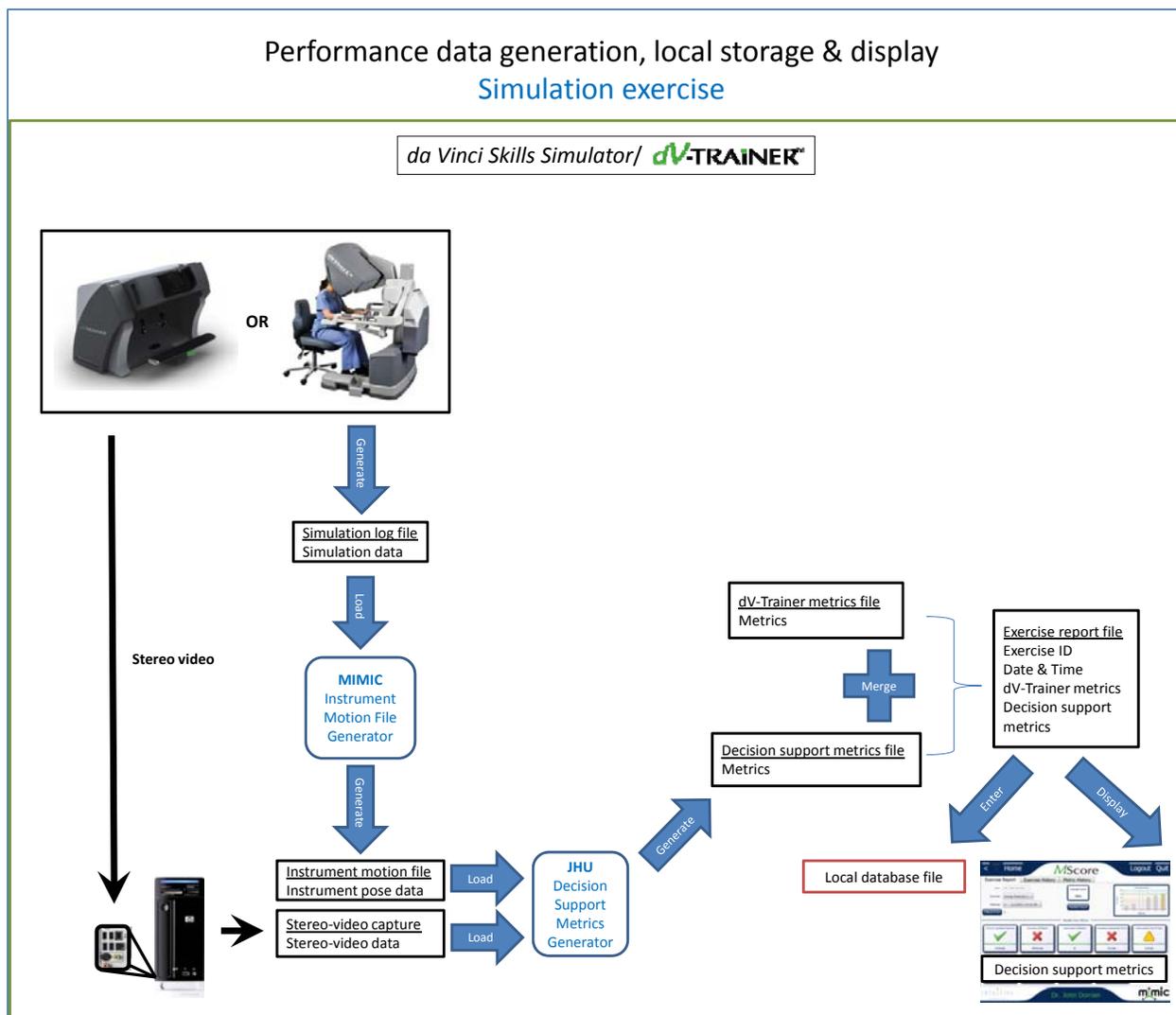


Figure 3 - Prototype for performance data generation, local storage and display in a simulation exercise on *da Vinci* Skills Simulator or *dV-Trainer*

## Surgeon Performance Data Collection, Local Storage & Display (Tasks 1 & 2)

While the surgeon is performing a training task, hand and instrument motions along with the captured stereo session video are processed to generate and store the surgeon's performance.

### Instrument Motion File

We created a non-proprietary data format for storing motion data that can be produced in training laboratories. The instrument motion file, task performance data acquired in the work flows presented above (Figures 2 and 3), is an extended and unencrypted version of the JHU SAW open source surgical performance data format that is already being used to store *da Vinci* motion and video data from several research centers around the country. This format reports sampled

pose information (*i.e.*, position, orientation, velocity and opening values) of hand motion and as well as instruments in any performed exercise at a desired sampling frequency in synchronization with the corresponding stereo-video data of the exercise.

### Mimic Instrument Motion File Generator & Synchronizer

We developed the Mimic Instrument Motion File Generator & Synchronizer, an online and offline post-processing tool, (Figures 2, 3, and 4) as a stand-alone application that converts the API data given in a simulation log file to the desired format represented in an instrument motion file. It also has a synchronization module that can synchronize an instrument motion file with its corresponding stereo-video.



Figure 4 - Mimic Instrument Motion File Generator & Synchronizer. Instrument motion file generation module (left) and synchronization module (right)

### JHU Decision Support Metrics Generator & dV-Trainer Metrics Generator

The advanced machine learning techniques that were developed for computation of the decision support metrics (Pose difference, Pose efficiency, etc.) are integrated in the stand-alone JHU Decision Support Metrics Generator application (Figures 2, 3 and 5). Another stand-alone application, dV-Trainer Metric Generator, is also developed to calculate the cross-platform dV-Trainer metrics (Economy of motion, Time to complete exercise, etc.). These applications are used for post-processing the collected surgeon performance data to derive performance metrics and can also be called within the dV-Trainer at the end of an exercise.

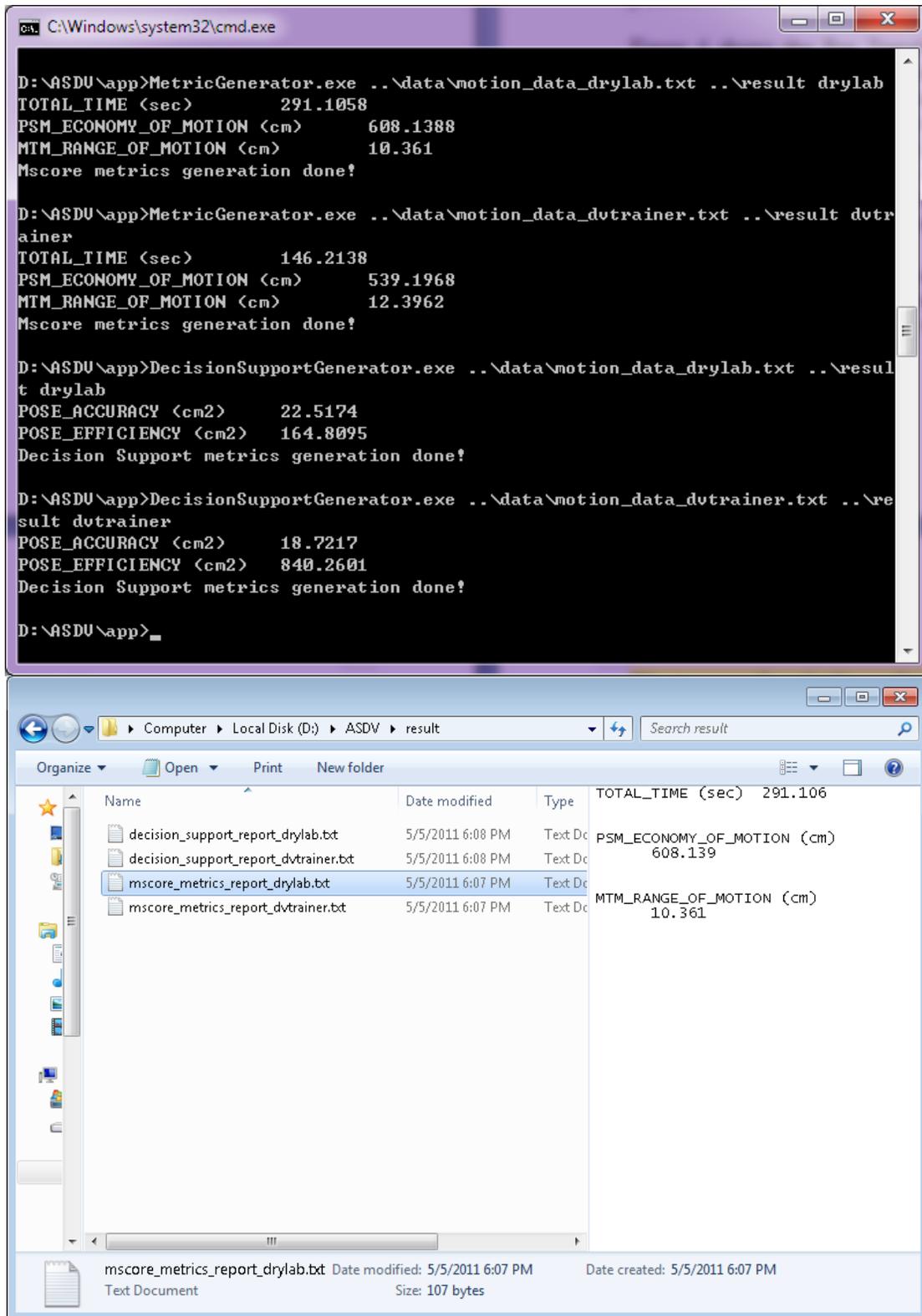


Figure 5 - JHU metric generator application screenshots (top) and generated result files (bottom).

A range of additional metrics have been developed for use in the preliminary study. These metrics extend the existing metrics currently available in the dV-Trainer, and have been shown to be better predictors of skill [2, 3]. These metrics are:

*Pose difference:* As the line distance traveled by an instrument does not capture the pose dexterity, this metric computes dexterity, this metric computes an area swept ( $\Sigma A_i$ ) by a unit length of the instrument over the period of the task (over the period of the task (

- Figure 6).
- *Pose efficiency:* The Pose efficiency uses the derivative of the pose difference, capturing large orientation velocities and large deviations and corrections by a user.
- *Pose accuracy:* A task metric integrates the notion of deviation from an ideal path. It computes a cumulative pose difference from an instrument traversing the “correct” path to the target. The “correct” path may be an analytical computation for a simple exercise or a model based on proficient executions for more complex paths.
- *Proficiency distance:* A proficiency distance is a measure of performance computing the distance along an established learning curve to the competency threshold.

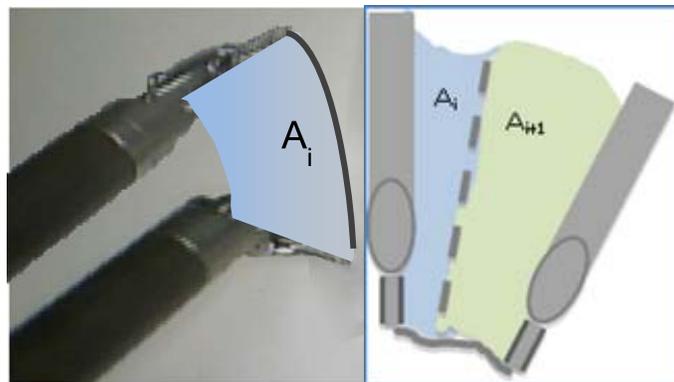


Figure 6 - A tip distance (the gray line) does not capture the pose variation ( $A_i$ ) which changes significantly between different instrument orientations with the same tip position trajectory.

Examples of the existing useful dV-Trainer metrics further investigated in this study include:

- *Master Workspace Range:* Larger of the two radii of motion of the user’s working volume on master handles
- *Instrument Collisions:* A count of two or more instruments coming into unintended contact
- *Economy of Motion:* Total distance travelled by all instruments
- *Instruments Out of View:* Total distance travelled by instruments outside a user’s field of view
- *Target Misses:* Number of missed needle targets.
- *Time to Complete Exercise:* Total elapsed time during the exercise from when the user enters following or camera control until the final target is completed.

## Exercise Report File

We defined a common exercise report file format (Figures 2 and 3) that contains computed Decision Support and dV-Trainer metrics along with some additional identification data such as ID number of the performed exercise and date and time of the session.

## Local Database

The dV-Trainer's local database contains performance data in terms of the existing dV-Trainer metrics. The database was modified to accommodate defined decision support metrics. The dV-Trainer user interface's features were expanded to support entering all session performance data given in an exercise report file to the local database at the end of an exercise.

## Performance Evaluation & Display

We designed and implemented a comprehensive performance evaluation tool in the dV-Trainer user interface, MScore, which provides objective assessment measuring robotic surgery skills across all computed metrics (Figure 7). In addition to viewing single exercise reports in detail and exporting them, users can keep track of their progress history over time. MScore's advanced features let administrators access individual exercise reports of users as well as search, sort and export the entire performance database for further review and statistical analysis.

MScore is automatically launched immediately after a simulation exercise is completed, and provides the user with an evaluation summary of his performance. Scores of individual metrics are calculated using customizable baseline values to present metrics with deficient performance to the user clearly using straightforward visualization tools (*e.g.*, check mark for high performance and cross mark for low performance). The default baseline values are determined by observation of the dV-Trainer owner performances over the years. User is provided with additional performance feedback on a specific metric to learn the ways of improving his skills on the metric. The user can also view his performance history over the metrics for all his previous attempts and is provided with a comparison to the average of all users' or proficient's performance.



Figure 7 - MScore provides the user with overall performance evaluation summary (left), and metric-specific detailed performance evaluation and feedback (right) for all exercises performed with the system. MScore displays also include learning curves for all computed metrics (bottom) comparing the user's performance history (blue line in charts) to the average of all users (grey line in charts) or proficient performance.

The administrator is provided with a comprehensive interface allowing detailed performance data searching, analysis and exporting capabilities (Figure 8). He can set a variety of searching criteria for viewing a subset of the performance database and/or can export the entire performance database in open source Comma Separated Value format (CSV) for further statistical analysis and review.



Figure 8 - MScore allows system administrators to search, analyze (left) and export (right) performance data

## Surgeon Performance Data Archival (Tasks 1 & 3)

Preliminary data repositories have been established at Mimic (in the form of a File Transfer Protocol server) and JHU. The data collected in feasibility experiments were archived in these repositories. Secure access to stored data is available in both repositories to the members of the project team.

### Database Server

We also designed a 3-tier system to provide flexible and scalable client-server architecture to share data collected from different platforms (Figure 9). The system consists of

- an HTTP web server with a secure web-based user interface and Simple Object Access Protocol (SOAP) end-point
- a Java based application server based on the business logic developed by Mimic for processing surgeon performance data collected from different training platforms
- an enterprise database server for storage and retrieval of collected surgeon performance data.

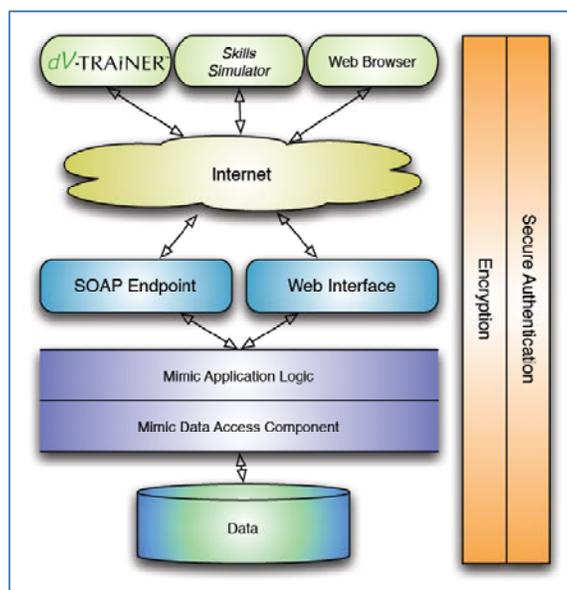


Figure 9 - 3-tier client-server design for secure sharing and transfer of surgeon performance data among different training platforms

The Mimic's MScore application will likely be the primary mechanism for analyzing, uploading and comparing performance data in the form of a dV-Trainer database. The web-server's interface will provide an alternative for those who do not have immediate access to an internet-ready dV-Trainer or *da Vinci*.

The second-tier implements an application server for managing the collected data and allows users to retrieve data stored on Mimic servers for future evaluation. The application server receives the data from the end-user and processes this information storing it in the third-tier database server. The application server can also retrieve the information from the database and present it to the user via the web-based interface.

We investigated the standards and requirements for transferring information over National Health Information Network (NHIN). NHIN has developed a specification of standards, protocols and governance to securely exchange health information between its nodes promoting a reference implementation, CONNECT open source gateway software solution, which is a JavaEE solution that relies on Web services using SOAP requests and responses to transfer information [4]. The SOAP protocol is based on sending XML formatted requests and responses between the clients and servers. CONNECT provides a server-based Primary Key Infrastructure (PKI) for authenticating network participants as well as client framework to customize the solution for private organizations.

We have implemented a preliminary working prototype of the designed 3-tier system based on CONNECT standards running on a Mimic server machine.

It should be noted that the Virtual Lifetime Electronic Record (VLER) standard was investigated. It is doubtful that surgeon performance data, especially when derived from simulation training and real phantom laboratories, should be applied to patient records. It is possible that it might be deemed appropriate to include a reference to NHIN transferred server data after surgery. This could easily be added as a de-identified text entry to a patient's VLER if needed, so the actual data would not be accessible by patients. It is highly unlikely that the medical community would accept the explicit inclusion of surgeon performance in a patient record. As patients could gain access to this data, there could be legal implications that would dissuade surgeons from wanting to participate in the assessment of their performance. Therefore, VLER was not addressed in Phase I feasibility study nor it will be addressed in Phase II development, but a performance reference could easily be exported to VLER at a later date.

### **Server/client Software for Data Transfer to/from the Server**

A prototype server/client application in the form of a web-based version of MScore was developed for uploading and downloading the local database of the dV-Trainer to and from the database server (Figure 10). The prototype application runs on the application server detailed in the previous section.

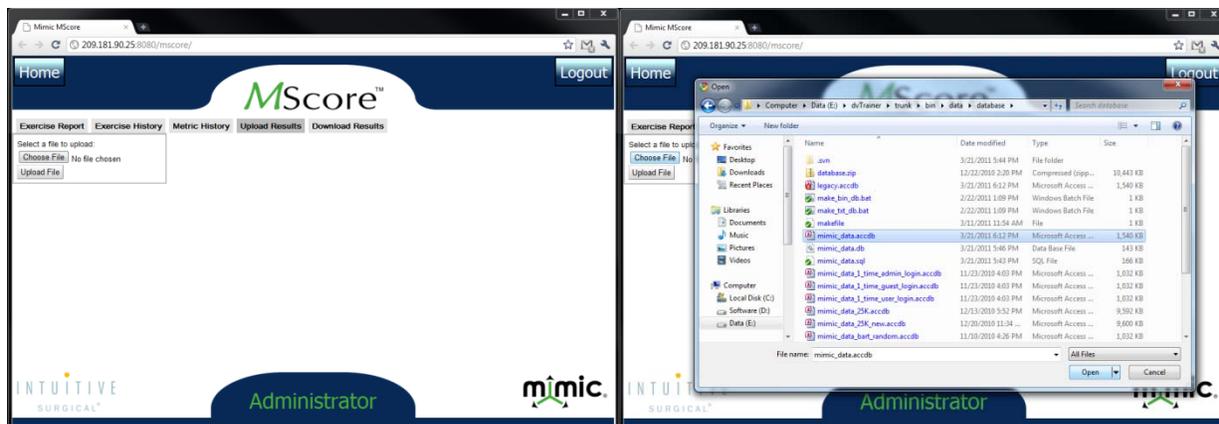


Figure 10 - A web-based prototype of MScore runs on the server (left) for uploading and downloading local database file of dV-Trainer to and from the server (right)

## Training Exercises (Task 4)

We developed common structured laboratory environments and corresponding experimental protocols for anastomosis and peg transfer that can be performed across all platforms for assessing cross-platform task performance.

### End-to-End Anastomosis

Figure 11 shows the needle states from a *da Vinci* S laboratory and the corresponding dV-Trainer task states for this task. This environment contains a simulated vessel to be anastomosed, together with markers and fiducials that allow automated computation of needle entry and exit points from captured video. We also capture the needle pose, which may improve the assessment of the intent of the user beyond the instrument pose captured by kinematic sensors. Initial layout of the simulated exercise was translated into robotic workspace coordinates, an initial setup of the anastomosis task was developed in real phantom laboratory platform, and experimental data was collected both in simulation and real phantom laboratories.



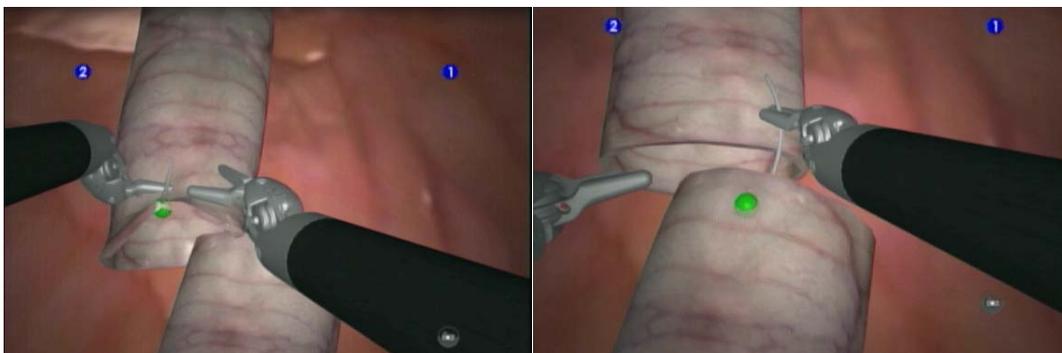


Figure 11 - End-to-end anastomosis task in real phantom (top), and simulation (bottom) environments

The task requires the repair of a simulated vessel by four needle throws on targets distributed 90 degrees apart along both edges of the simulated vessel. In the *da Vinci* version (Figure 11, top), this end-to-end anastomosis uses a 1" Penrose drain. The 3mm circle targets were placed 5mm away from the edge, and a 3-0 NSH-1 suturing needle was used to simulate the repair needle throws using two large needle drivers. A simulated vessel of similar size is placed in dV-Trainer scene and similarly oriented. The simulated vessel (Figure 11, bottom) also contains the same targets, and the needle is driven using two simulated large needle drivers as well. The exercise starts when the needle is picked up by either instrument from the vessel and it ends when the user throws the needle through the last target. The experimental protocol requires the users to complete the repair by moving clockwise through the targets in sequence, after appropriate familiarity with the respective setups.

### Peg Transfer

In addition to anastomosis, we used a second peg transfer structured laboratory environment that can be performed across all platforms for assessing task performance. Compared to the surgical skill in the previous task, this more elementary task relates more to system operation.

Figure 12 shows the Peg Transfer task from a *da Vinci* S laboratory and the corresponding simulation environment from a dV-Trainer exercise. Layout of the simulated exercise was translated into robotic workspace coordinates, and a setup of a real laboratory was developed using the same scale. This environment contains a peg board with a row of 6 pegs placed on the wall and a row of 2 pegs placed on the base of the board. Each peg's diameter is 2.4mm and height is 12.5mm. The pegs in the first row on the wall are placed 25mm away from the boundary and 20mm apart from each other at a height of 30mm. The pegs on the base are placed 50mm apart at a depth of 60mm.

Error! Reference source not found. Figure 12 – Peg Transfer task in real phantom (top), and simulation (bottom) environments

Six 8mm solid rings (with an inner diameter of 7mm) are initially placed on each of the six pegs on the first row on the wall. The experimental protocol requires the users to transfer each ring from its initial peg on the wall to the right peg on the base by

- first taking the ring off from the peg with the left instrument
- then transferring the ring from the left instrument to the right instrument
- and finally placing the ring on the right peg on the base with the right instrument

The task protocol requires each ring to be transferred using large needle drivers as explained above in order from the leftmost ring to the rightmost ring to practice hand-eye coordination and object manipulation. Experimental data was collected both in simulation and real phantom environments.

The training tasks described above are two-handed tasks; however the task goal is carried out by only one of the two hands holding an object (the needle in anastomosis or the ring in peg transfer) at any given time. Therefore, we segmented the instrument motion to include only those portions where instrument is holding an object.

## Data Collection & Analysis (Task 5)

### Preliminary Experiments

During the first phase of data collection, a single performance of a single task (anastomosis) trial was recorded from six different users – three each for the *da Vinci* and the dV-Trainer platforms. On each platform, we used a proficient user (practice time >100 hours on the respective platforms), and two users with varying but smaller amounts of previous training. Recordings started after sufficient practice time for warming up on the respective platform.

Figure 13 shows pose accuracy, the visualization of instrument motion of the proficient user performing the first needle throw in real phantom laboratory and the dV-Trainer motion data, respectively. Left and right instruments are drawn in blue-to-green and red-to-yellow tones respectively with brighter colors representing higher velocities.

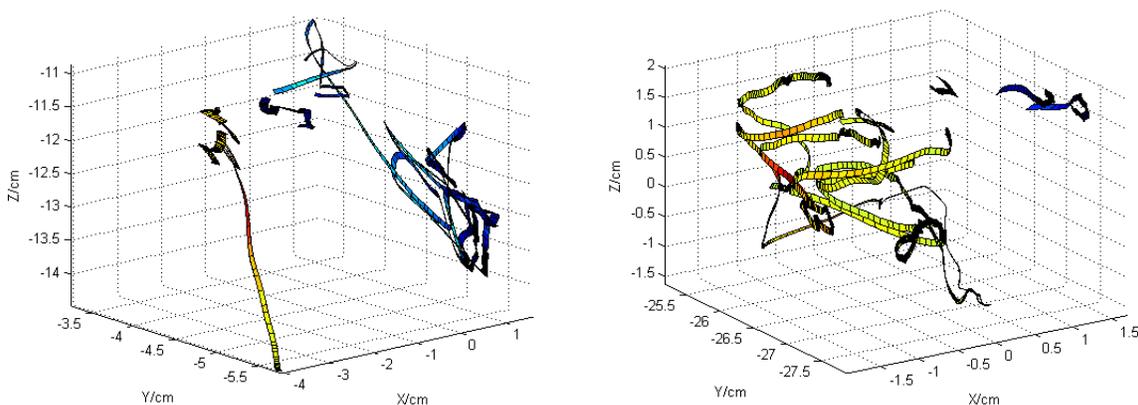


Figure 13 - The ribbon surfaces panned by the gripper during the first needle throw of a training task. Left: phantom laboratories. Right: simulation

The ranges of workspace used for performing the tasks are approximately [6cm, 4cm, 8cm] and [5cm, 10cm, 8cm] in real and simulated environments, and the instrument motions are subjectively similar except for the left and right-handed-ness of the subjects. This experimental setup therefore provides us with similar real and simulated environments to verify that the metrics described above are applicable in both training paradigms.

The proficient user provides the baseline, and the data from the other users is used to establish preliminary trends for the metrics. These metrics (normalized to fit on the same scale) with the *da Vinci* (User 1-3) and the dV-Trainer (User 4-6) are shown in Figure 14. User 1 and User 4 are the proficient users. All values are mapped linearly to the range of  $[0 + \delta_0; 1 - \delta_1]$  where  $\delta_0$  and  $\delta_1$  are small positive numbers allowing normalization to preserve the distribution of the values.

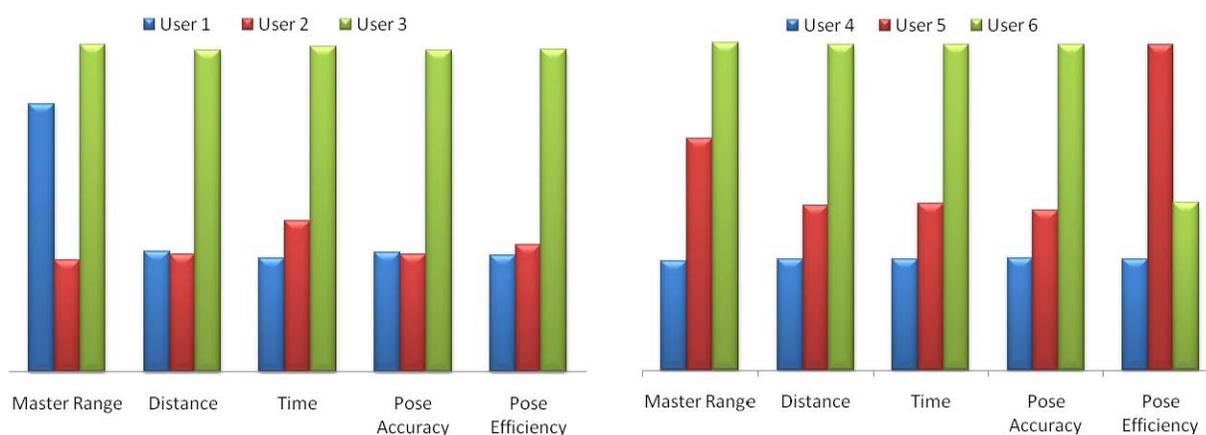


Figure 14 - Normalized metrics for anastomosis with the *da Vinci* (User 1 – 3) and the dV-Trainer (User 4 – 6)

The proficient users (1 and 4) generally outscore their novice counterparts in most of the computed metrics. Further in the simulation environment, these trends also follow the validation

studies for the metrics currently reported in those studies for other tasks, and the new metrics show separation similar to [2].

Table 1 below presents metric values of the performances of all users from two platforms.

metric	Real Phantom laboratories					
	U1(proficient)		U2		U3	
	holding	entire	holding	entire	holding	entire
<b>Master Workspace Range (cm)</b>	10.96085991	13.39659	9.229559	11.19484	11.62816	12.5909089
<b>Economy of Motion (cm)</b>	151.5282099	275.9289	149.0882	293.072	293.5951	563.5790679
<b>Time to Complete Exercise (sec)</b>	218.218		240.942		346.882	
<b>Pose Accuracy (area of the pose ribbon)(cm<sup>2</sup>)</b>	8.832729008	21.45309	8.728244	21.62153	18.81602	42.85939989
<b>Pose Efficiency (area of the velocity ribbon)(cm<sup>2</sup>/sec<sup>2</sup>)</b>	85.65410753	208.6871	88.93589	205.8313	150.8054	326.8043144
	Simulation laboratories					
	U1 (proficient)		U2		U3	
	holding	entire	holding	entire	holding	entire
<b>Master Workspace Range (cm)</b>	12.52970904	12.3962	16.1999	16.30311	21.93352	21.38037903
<b>Economy of Motion (cm)</b>	302.0009191	539.1968	305.8661	622.381	829.8543	1797.116919
<b>Time to Complete Exercise (sec)</b>	194.903		287.596		479.557	
<b>Pose Accuracy (area of the pose ribbon)(cm<sup>2</sup>)</b>	20.02292851	42.51423	21.96169	48.533	56.28769	138.2223654
<b>Pose Efficiency (area of the velocity ribbon)(cm<sup>2</sup>/sec<sup>2</sup>)</b>	863.3050854	1188.473	1186.035	1623.46	941.034	2572.947483

**Table 1 - Metric values of the performances of 3 users each across the two platforms for anastomosis.**

Figure 15 shows MScore metric evaluation summaries of the proficient (U1) and a non-proficient (U3) user based on preliminary data of the simulated anastomosis after the data is post-processed and loaded into the dV-Trainer. The current representation communicates specific skills needed, and the level of improvement required to user at the completion of each training task. As can clearly be seen from the figure, MScore can effectively show that a proficient user's performance is superior to that of a non-proficient user. For example, all raw metric values of the proficient user on the right are much better than that of the non-proficient user (the smaller the raw value, the better the performance). Also, individual metric scores of the proficient user calculated based on custom baseline values are higher than that of non-proficient user.



Figure 15 - MScore's metric evaluation summary of a non-proficient (left) and proficient (right) user from preliminary experiments in simulation. Different icons are used to quickly relay various thresholds of performance.

## Large Scale Data Collection

Upon early successful completion of the feasibility study, we undertook an extension of the project with acquisition of a much larger database, not included in the original scope of work.

This second phase of data collection aims to collect twelve performances of both tasks (anastomosis and peg transfer) trials from twelve different users – six each for the *da Vinci* as well as Skills Simulator and other six on the dV-Trainer as well as Skills Simulator platforms. On each platform, a proficient user (practice time >100 hours on the respective platforms), and five users with varying but smaller amounts of previous training will provide this data.

The dV-Trainer recordings started after sufficient practice time for warming up on the respective platform. This data collection also reduces the individual variability by using double the number of users and increasing the trials of performance by each user to twelve.

Six dV-Trainer users at MIMIC have already provided the experimental data and data collection at JHU will start once the IRB protocol is approved. Figure 16 shows the Pose Accuracy metric comparison between a proficient user and a non-proficient user across twelve trials. The distribution of the computed metrics indicates no learning in individual users across the twelve trials.

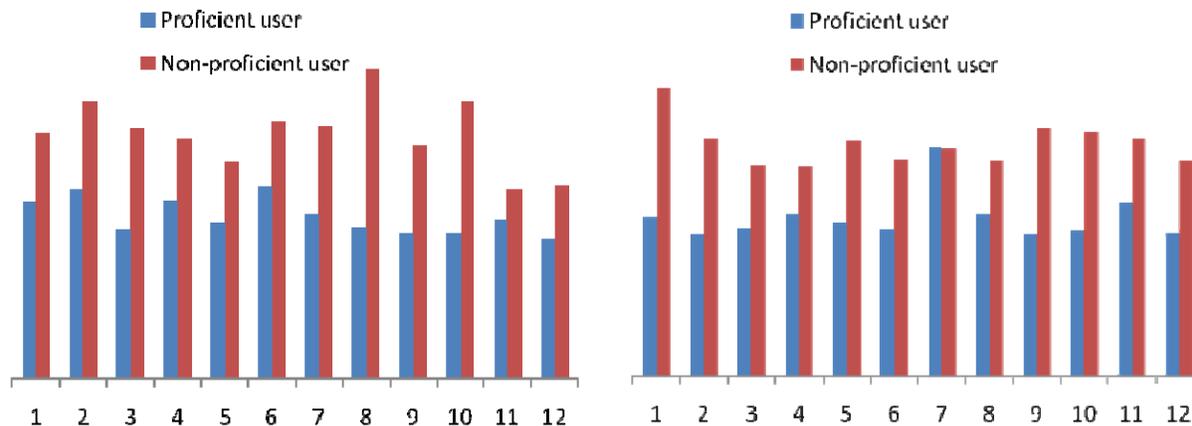


Figure 16 – Pose Accuracy metric comparison between a proficient user and a non-proficient user across twelve trials. Left: Anastomosis. Right: Peg transfer.

Figure 17 shows the normalized averaged (across the 12 trials) metrics of six users for the dV-Trainer, where User 1 is a proficient user. Here motions of both hands are included in computation. Pose Accuracy and Pose Efficiency provide the widest ranges of measurements across the subjects. Additional analysis awaits completion of data collection from the *da Vinci* system.

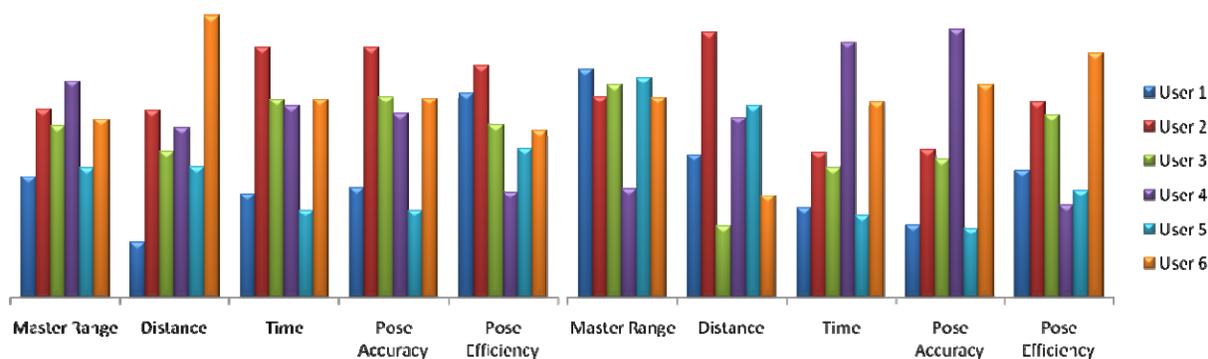


Figure 17 – Normalized averaged metrics with the dV-Trainer (Users 1 – 6, User 1 is proficient). Left: Anastomosis. Right: Peg transfer.

Figure 18 shows MScore metric evaluation summary of the best performance out of twelve trials of a proficient and a non-proficient user based on preliminary data of the simulated peg transfer and anastomosis. For both peg transfer (Figure 18 - top) and anastomosis (Figure 18 - bottom) exercises, it is shown that the proficient user's performance (Figure 18 - right) is superior to that of the non-proficient user (Figure 18 - left). For example, most raw metric values of the proficient user are much better if not the same than that of the non-proficient user (the

smaller the raw value, the better the performance). Also, individual metric scores of the proficient user calculated based on custom baselines are higher than or equal to that of non-proficient user. It is also represented in the Overall Score chart of the views that, generally, individual exercise scores of the proficient user (blue bar in the chart) are higher than the average scores of all six subjects (grey line in the chart), whereas that of the non-proficient user are lower than the average scores of all six subjects. Finally, it is worthwhile to add that the Overall Score of the non-proficient user shows improvement over the trials.

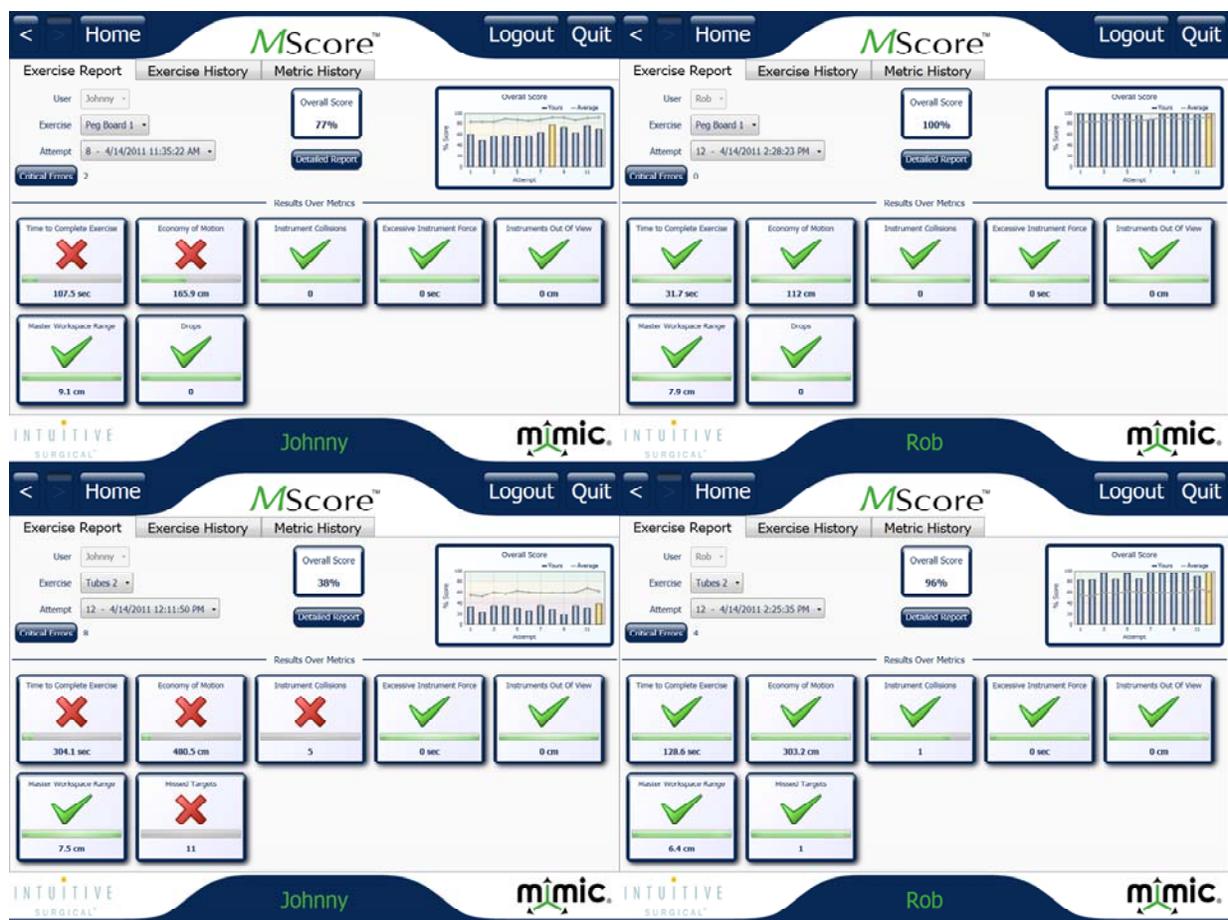


Figure 18 - MScore's metric evaluation summary of the best performances of a non-proficient (left) and proficient (right) user from preliminary experiments in simulated peg transfer (top) and anastomosis (bottom) exercises.

Figure 19 shows MScore's exercise progress history in terms of individual progress history of metrics collected during the exercise over twelve trials. It is shown that the proficient user's performance progress history per metric (Figure 19 - right) is in general superior to that of the non-proficient user (Figure 19 - left). It is also represented in individual metric progress history charts of the views that, generally, the proficient user (Figure 19 - right, blue line in the charts) has higher than average metric scores of all six subjects (Figure 19 - right, grey line in the charts), whereas the non-proficient user (Figure 19 - left, blue line in the charts) has lower than average metric scores of all six subjects (Figure 19 - left, grey line in the charts).



Figure 19 - MScore's exercise progress history view per metric of a non-proficient (left) and proficient (right) user from preliminary experiments in simulated peg transfer (top) and anastomosis exercises.

The new metrics reported here have not been computed previously in *da Vinci* experiments, although the simpler simulation metrics were computed for other similar phantom laboratory tasks [5]. These initial results show the promise of cross-platform computation of skill metrics, that if verified can allow for the proposed infrastructure to assess training metrics and parameters across all stages of training.

Table 2 below presents common performance metrics of all users from the dV-Trainer laboratories for the larger study. Additional analysis awaits IRB approval.

Metrics	Anastomosis					
	U1(Proficient)	U2	U3	U4	U5	U6
Master Workspace Range (cm)	128.6482	288.5554	250.3583	355.7322	150.1382	264.1687
Economy of Motion (cm)	11.39592	15.7034	14.37703	15.13084	13.87476	18.78911
Time to Complete Exercise (sec)	442.3579	813.6293	681.6936	665.1016	401.3126	682.996
Pose Accuracy (area of the pose ribbon)(cm <sup>2</sup> )	35.15971	63.31675	53.52775	50.07132	30.664	53.10483
Pose Efficiency (area of the velocity ribbon)(cm <sup>2</sup> /sec <sup>2</sup> )	1965.433	2269.855	1599.232	835.3991	1330.83	1537.931

Metrics	Peg Transfer					
	U1(Proficient)	U2	U3	U4	U5	U6
Master Workspace Range (cm)	27.4625	51.57192	41.33125	132.4828	34.687	52.35367
Economy of Motion (cm)	13.27937	18.56443	10.26047	14.9088	15.39405	11.52716
Time to Complete Exercise (sec)	128.1455	161.0158	151.616	227.1425	122.9271	190.765
Pose Accuracy (area of the pose ribbon)(cm <sup>2</sup> )	10.84063	13.5426	13.20471	17.86872	10.7118	15.88111
Pose Efficiency (area of the velocity ribbon)(cm <sup>2</sup> /sec <sup>2</sup> )	275.7386	377.2603	358.7542	223.6151	245.5768	452.1488
Master Workspace Range (cm)	27.4625	51.57192	41.33125	132.4828	34.687	52.35367

Table 2 - Averaged metric values of the performances of 6 users from the dV-Trainer from the larger study.

## Work Plan for Phase II Development (Task 6)

We submitted a proposal for Phase II development on March 17<sup>th</sup>, 2011. The following paragraphs summarize the work plan in Phase II development. More details can be found in the submitted proposal.

The current Phase I framework requires a number of manual steps to collect, assess, upload and compare data. Phase II will focus on automating this system for the purpose of commercialization. Phase II will result in a fully functional prototype that has commercial applicability. The prototype will be deployed at numerous test sites to verify its effectiveness. It is expected that the prototype will be refined for commercialization shortly after the completion of Phase II and that the framework will be quickly integrated with the dV-Trainer and *da Vinci* Skills Simulator products. Phase III funds might also be used to commercialize a version of the SAW system that could be utilized for assessment of phantom laboratories and *da Vinci* Surgery.

While Phase II development will certainly provide surgeons with additional assessment information related to their training and performance, the ultimate goal is to accelerate a surgeon's climb up the learning curve. Phase II development will be considered successful if it can be proven that the resulting assessment and feedback tools accelerate surgeon acquisition of *da Vinci* surgical skills. Our system must encourage surgeons to pursue additional training that they might not otherwise undertake without performance feedback. In addition, recommended training regimens should prove more efficient than training regimens that individuals might create for themselves without guidance. The recommended training regimens should also keep the trainee more engaged than an unguided training regimen.

The following is a list of technical objectives in Phase II development:

1. Adapt the current Mimic simulation software platform to enable internet access that will facilitate information exchange with Mimic servers.

2. Automate the current data collection and analysis framework for the dV-Trainer, *da Vinci* Skills Simulator and SAW system while conforming to the NHIN standards.
3. Develop additional post-training assessment tools that will help trainees understand their own performance deficiencies and recognize their need for follow-on training.
4. Identify and validate a set of baseline training exercises that can be conducted in simulation and phantom laboratories, which can be used to comprehensively identify surgical skills deficiencies.
5. Create a mechanism for automatically generating a customized training protocol derived from baseline exercise testing.
6. Validate the use of resulting customized training regimens intended to accelerate a surgeon's progress towards *da Vinci* proficiency.

## KEY RESEARCH ACCOMPLISHMENTS

*Bulleted list of key research accomplishments emanating from this research.*

### Discoveries

#### Decision Support Analysis Can Differentiate Non-proficient Surgeons from Proficient Surgeons

In preliminary results [3], the distinction between proficient and non-proficient users was maintained across the selected metrics. The current dataset is not sufficient to develop classification (novice, intermediate, proficient) methods and metric thresholds for graduation. We are developing these methods in related synergistic robotic surgery training studies, and will employ them upon collection of sufficient amount of data.

#### Man-machine Interface Effects in the dV-Trainer Compared to the *da Vinci* Surgical System

Comparison of metrics across the platforms requires that the simulation environment present a comparable man-machine interface and inertial properties for the simulated instruments, and camera. Any variability in these interactions would change the users' perception of the environment as well as their interactions with it significantly. Early results from limited data currently available suggest that further tuning of these parameters may be required to make the dV-Trainer substantially similar to a real *da Vinci* system. These findings, if validated in larger datasets, would be significant and new discoveries. While the development and tuning of man-machine interfaces is an ongoing process, the methods and findings discovered during the feasibility study add to the tools available to the research group and will expedite the development of a more efficient and realistic simulation training environment.

## Additional Deliverables

### Proficiency-based Scoring and Curriculum Development

We believe that the ultimate goals in robotic skills training are to accelerate a surgeon's climb up the learning curve and to make the surgeon keep his technical skills at top performance once proficiency is achieved. While the assessment and feedback tools developed in Phase I feasibility study certainly provides users with more advanced assessment information related to their performance and training, they are lacking the recommendation of a customized curriculum that is identified based on the training needs of individual users and that guides the users to improve their technical skills towards proficiency via proficiency-based measures. Such assessment and feedback tools should accelerate surgeon acquisition of *da Vinci* surgical skills by encouraging the surgeons to pursue additional training that they might not otherwise undertake without performance feedback. In addition, recommended training regimens should prove more efficient than training regimens that individuals might create for themselves without guidance. The recommended training regimens should also keep the trainee more engaged than an unguided training regimen.

One approach towards development of such assessment and feedback tools would be establishing proficiency by collecting expert data in a manner similar to that used when the Fundamentals of Laparoscopic Surgery (FLS) testing protocol was established [6]. Experts repeated an exercise five times and then all results (all trials and all subjects) were averaged and standard deviations were determined. Another option that was recommended to Mimic by ACS board members is to have experts repeat an exercise until they achieve two consecutive sessions of comparable metric data. Only the data from the consecutive performances would be used to determine averages. Regardless of the manner in which averages are determined, proficiency thresholds for each metric should be established by making use of the expert mean and standard deviation values.

We are planning to turn these ideas we came up with during our Phase I research effort into a new scoring system within the next generation of MScore, dV-Trainer's assessment tool. MScore currently uses a percentage-based scoring system to evaluate users. Expert data has been collected under a number of validation studies and this data has been used for deriving baseline values for individual metrics to calculate user performance score. We have started working on the implementation of a proficiency-based scoring and evaluation methodology to be integrated in MScore. The methodology is inspired by the current proficiency-based training curriculum in the FLS program [6]. The proposed methodology offers an evaluation that is simpler to understand and follow, and encourages a trainee to exceed expert proficiency and maintain that proficiency. Proficiency levels of metrics for each exercise are derived from the mean and standard deviation values from a defined number of performances of a number of expert surgeons. These proficiency values are then embedded in MScore to evaluate the user performance in comparison to the expert-derived values per metric. The trainee will pass a

metric only if he makes the proficiency level value of the metric, and he will pass a task only if he passes on all metrics of the task. Each task will be expected to be performed repetitively until the trainee passes the task. Furthermore, this level of performance will be expected to be achieved for a number of consecutive and nonconsecutive times for reinforcement and achieving proficiency on the task.

Figure 20 shows the mockups for an initial design of the proficiency-based scoring and visualization in MScore. We are planning to integrate MScore’s refined proficiency-based scoring and visualization in the next generation of Mimic’s dV-Trainer by the third-quarter of the year 2011 prior to the onset of Phase II efforts.



Figure 20 - Initial design of the proficiency-based scoring & visualization in MScore

## Reportable Outcomes

### Publications

1. Yixin Gao, Mert Sedef, Amod Jog, Greg Hager, Jeffrey Berkley, Rajesh Kumar, “Towards validation of robotic surgery training assessment across training platforms” IEEE/RSJ International Conference on Intelligent Robots and Systems, 2011 (in review)

Additional publications and intellectual property from this work are currently in preparation.

## CONCLUSION

During our Phase I effort, we designed and implemented a proof-of-concept working prototype of an overall framework that collects and analyzes surgeon performance data from simulation (the dV-Trainer and the *da Vinci* Skills Simulator) and phantom laboratory exercises. An anastomosis simulation module and a peg transfer simulation module were implemented in

simulation and phantom laboratories for surgeon performance assessment across different training platforms based on a defined set of cross-platform metrics. This allowed us to begin collecting preliminary surgeon performance data in order to test our data collection framework. Collected data from different platforms was uploaded to established Mimic Technologies servers and shared between Mimic and Johns Hopkins for further detailed analysis. Preliminary performance evaluation results compare a user's performance to that of proficient users or average user performance and identify the user's skills with deficient performance. The evaluation results are also capable of differentiating performance of a proficient user from that of a non-proficient user.

We successfully achieved all the technical objectives we proposed in the proof-of-concept working prototype we developed. We have tested our framework sufficiently to conclude that the development and implementation of a fully-functional automated support for the *da Vinci* Surgical System is highly feasible.

This work outlines the first prototype for measurement and assessment of robotic surgery training data across real and simulated training platforms. The system is designed to provide automated surgical skill and dexterity assessment on robotic and simulated *da Vinci* surgical training environments distributed at different sites.

We also report on customized metrics that capture efficiency of instrument manipulations in 6-DOF, extending our earlier work in [2]. In this research study, JHU presented the first work using motion data from the *da Vinci* Skills Simulator for classifying users of varying skills. Given the standardized environment of the *da Vinci* Skills Simulator, and the availability of the ground truth, skill measurements and feedback based on task motion hold the promise of effective automated objective assessment. Based on motion data of a simulated manipulation task from 17 users of varying skills, we demonstrate binary classification (proficient vs. trainee) of user skill with 87.5% accuracy. Alternate measures based on instrument pose, which are more relevant in the simulated environment, including a new measure of motion efficiency, were also presented and evaluated.

Lastly, we also study, for the first time, common performance metrics for the same training tasks across two platforms in experiments performed by users on the *da Vinci* system and the dV-Trainer. The preliminary computations show trends similar to larger studies on individual platforms, and that the proficient and non-proficient users are differentiable using the studied metrics. We also show that performance metrics of training exercises previously validated in simulation environments hold in training exercises with a real robotic system.

When developing surgical assessment protocols, an important step is to see how such mechanisms compare to gold standards for training and assessment. Unfortunately, it is debatable that such a gold standard exists for *da Vinci* assisted surgery, so there is no clear path

for development and validation of an automated support system. Both Mimic Technologies and JHU are actively involved with a variety of ongoing efforts to determine best training practices. Therefore, even in the absence of a gold standard for robotic-assisted surgery, the key personnel involved in this research effort drew upon their collective experience to propose a variety of assessment metrics that can be applied to various training modalities. The goal of this effort was not to develop a new gold standard for training, but rather to create tools that simplify data collection and offer new alternatives for performance assessment. Such tools might later help to determine a universally accepted gold standard, which could be based on simulation, inanimate models or animals.

One challenge of creating a robust data collection and assessment mechanism is to make it applicable to a variety of training modalities and surgery itself. Several studies in the literature reflect that skill assessment in simulation environments has so far only focused on evaluating the utility and validity of statistics such as task completion time, and instrument distance measured during a simulated task. Because of variable content, scenarios, and training objectives, it might not be practical to rely solely on traditional statistical analysis. Traditional analysis [7-14] does not distinguish between clinical task skills (*e.g.*, suturing) from the complex human-machine interactions needed to proficiently operate the robotic console (*e.g.*, master workspace adjustment, or camera control). Task time, structured assessment [11-13], and errors do not capture all the additional complexity inherent in robotic surgery. The developed prototype collects a wide range of data and tools are included to apply simple statistical analysis to this data. However, a big focus of this project was to apply advanced assessment algorithms based on machine learning [15-17], including further development of JHU statistical frameworks for segmenting, recognizing, and assessing surgeon task performances.

## Research Impact and Applications

The developed prototype would provide open-source encapsulation of Intuitive Surgical's proprietary Application Programming Interface (API) for collecting surgeon performance data at sufficient quality and granularity to provide meaningful evaluation and feedback in a standard format, and meeting National Health Information Network (NHIN) standards across a network of training devices. Such quantitative measurements would include tool, camera and master handle motion vectors including joint angles, velocity, and torque, Cartesian position and velocity, gripper angle, and synchronized stereo video data ("procedure data").

The integration of a customized, user-dependent and proficiency-based follow-on training and curriculum suggestion mechanism in the prototype would accelerate a surgeon's climb up the learning curve towards acquisition of *da Vinci* surgical skills more efficiently than training regimens with no guidance. Such an effective support system for the *da Vinci* would have significant commercial value, especially if used in conjunction with Mimic's dV-Trainer product line.

It is also expected that the resulting information system could prove valuable to institutions that have a stake in effective surgical training. Surgeon performance data would be tracked and stored on a secure remote server for authorized access of multiple research and training institutions using NHIN standards, which would make it possible for the users to compare themselves to the average trainee and proficient surgeon performance on a national basis. Governing medical bodies could use the system to establish training and procedure guidelines, such as when creating a Fundamentals of Robotic Surgery (FRS) program.

It should be noted that an extensive development project may be initiated as early as the end of 2011 to establish a Fundamentals of Robotic Surgery (FRS) testing program, which will be similar in nature to the FLS program. Assuming the project gets funded, Mimic will likely play a role by creating simulation exercises that replicate a series of phantom laboratories and to support validation studies of those exercises. Most likely, these phantom laboratories will be very similar to those found in FLS, but it is likely that brand new phantom laboratories will be created to support robot-specific skills. There are several ways a Phase II implementation of the Phase I feasibility study and an FRS project would complement each other. To our knowledge, the FRS project is not expected to cover the development of a system for automated data collection and upload to a central server. Considering all the review hours required to review FLS collected video, the automated processing resulting from this project would be a welcome addition to an FRS program. The goal of FRS is to create a testing protocol for credentialing surgeons for robotic surgery. However, there is nothing planned for optimizing training so that a surgeon can pass an FRS test. Therefore, the results of this project combined with the FRS results would mean a complete system for testing, training, data collection, performance comparison and credentialing. An FRS program would involve substantially more data collection to establish proficiency than what we plan for this project. However, the methods used in this project to establish a testing protocol and customized training regimen could be applied to the data collected from an FRS project. In the event that this project begins before an FRS project, we would certainly be in a good position to recommend exercises for inclusion in FRS.

Because of the rapidly expanding install base of Mimic's dV-Trainers and potential access to existing *da Vinci* Surgical Systems; it will eventually be possible to implement large-scale collection of longitudinal training data. This could lead to identification of common trends, learning curve behaviors and further research in robotic-assisted surgery training to improve performance and surgical outcomes while reducing case times.

As part of Phase I development and prior research at Johns Hopkins, dexterous skills have been being evaluated using JHU's algorithms for surgical gesture modeling and recognition work – what we refer to as the “Language of Surgery.” In this related work, JHU have developed techniques that allow the use of traditional speech modeling techniques (specifically, highly modified variations on Hidden Markov Models (HMMs) to model kinematic data acquired using the *da Vinci* API and through Mimic simulation software. This data, which comprises up to 200

variables, captures at rates of up to 100 times a second, is captured in synchrony with the associated stereo video data, forming a complete record of the performance of the recorded subject. JHU has shown that it is possible to train gesture-based models from proficient surgeons, and to synchronize surgical recordings of less skilled users to such proficient. In these synchronized recordings, it is possible to:

- Perform comparative statistical analysis, at both the task and gesture level, to identify level of skill.
- Develop methods for feedback from expert models to support training. This feedback can take both visual and haptic form.
- Detect variations in technique over time due to learning or fatigue.

## REFERENCES

1. Intuitive Surgical Inc., Investor Presentation Q1 2011, <http://phx.corporate-ir.net/External.File?item=UGFyZW50SUQ9OTE1MTR8Q2hpbGRJRD0tMXxUeXBIBT M=&t=1>
2. Jog, A, Itkowitz, B, Liu, M, DiMaio, S, Hager, G, Curet, M, Kumar, R, Towards integrating task information in skills Assessment for Dexterous Tasks in Surgery and Simulation, IEEE International Conference on Robotics and automation (ICRA), May 9-13, 2011
3. Gao, Y., Sedef, M., Jog, A., Hager, G., Berkley, J., Kumar, R., Towards validation of robotic surgery training assessment across training platforms, IEEE/RSJ International Conference on Intelligent Robots and Systems, 2011 (submitted)
4. CONNECT Community Portal. 2009. Nationwide Health Information Network. 05/02/2011. <http://www.connectopensource.org/>
5. T. Judkins, D. Oleynikov, and N. Stergiou, “Objective evaluation of expert and novice performance during robotic surgical training tasks,” *Surgical Endoscopy*, vol. 23, no. 3, pp. 590–597, 2009.
6. FUNDAMENTALS OF LAPAROSCOPIC SURGERY, Technical Skills Proficiency-Based Training Curriculum 05/04/2011. <http://www.flsprogram.org/wp-content/uploads/2010/08/Technical-Skills-Proficiency-Bases-Training-Curriculum.pdf>
7. Tendick, F., Downes, M., Goktekin, T., Cavusoglu, M., Feygin, D., Wu, X., Eyal, R., Hegarty, M., Way, L.: A virtual environment testbed for training laparoscopic surgical skills. *Presence: Teleoperators & Virtual Environments* 9(3) (2000) 236–255
8. Perrenot, C., Perez M., Tran N., Jehl JP., Nouri N., Abraham D., Felblinger J., Hubert J., Role of Simulation in Credentialing of Residents, 20<sup>th</sup> Worldwide Conference of Clinical Robotic Surgery Association (CRSA), Oct 1-2, 2010.
9. Grantcharov, T., Kristiansen, V., Bendix, J., Bardram, L., Rosenberg, J., Funch-Jensen, P.: Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *British Journal of Surgery* 91(2) (2004) 146–150
10. Seymour, N., Gallagher, A., Roman, S., OBrien, M., Bansal, V., Andersen, D., Satava, R.: Virtual reality training improves operating room performance: results of a randomized, double blinded study. *Annals of Surgery* 236(4) (2002) 458

11. Moorthy, K., Munz, Y., Sarker, S., Darzi, A.: Objective assessment of technical skills in surgery. *BMJ* 327(7422) (2003) 1032
12. Doyle, J., Webber, E., Sidhu, R.: A universal global rating scale for the evaluation of technical skills in the operating room. *The American Journal of Surgery* 193(5) (2007) 551–555
13. Reznick, R., Regehr, G., MacRae, H., Martin, J., McCulloch, W.: Testing technical skill via an innovative bench station examination. *The American Journal of Surgery* 173(3) (1997) 226–230
14. Kenney, P., Wszolek, M., Gould, J., Libertino, J., Moinzadeh, A., Face, Content, and Construct Validity of dV-Trainer, a Novel Virtual Reality Simulator for Robotic Surgery, *Journal of Urology*, 73(6) (2009)
15. Lin, H., Shafran, I., Yuh, D., Hager, G.: Towards automatic skill evaluation: Detection and segmentation of robotic-assisted surgical motions. *Computer Aided Surgery* 11(5) (2006) 220–230
16. Varadarajan, B., Reiley, C., Lin, H., Khudanpur, S., Hager, G.: Data-Derived Models for Segmentation with Application to Surgical Assessment and Training (2009)
17. Rosen, J., Solazzo, M., Hannaford, B., Sinanan, M.: Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden Markov model. *Journal of Image Guided Surgery* 7(1) (2002) 49–61

## APPENDICES

None.

## PUBLICATIONS MADE AS A RESULT OF THE RESEARCH EFFORT

### **Towards validation of robotic surgery training assessment across training platforms**

Yixin Gao, Mert Sedef, Amod Jog, Greg Hager, Jeffrey Berkley, Rajesh Kumar, “Towards validation of robotic surgery training assessment across training platforms” *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011 (submitted)

### **Abstract**

Robotic surgery is increasingly popular for a wide range of complex minimally invasive surgery procedures. To improve robotic surgery training, a skills trainer has recently been introduced, and a simulator is in advanced evaluation. These platforms report a range of time and motion based task metrics, and literature has investigated the validity of these metrics in training studies. However, the lack of a cross-platform data collection system has so far prevented a cross-platform investigation. Using a new architecture for collecting cross-platform motion data, we present the first study investigating whether metrics previously validated in simulation environments also hold in training exercises with a real robotic system. Our long term goal is to

assess both skills retention and skills transfer, and preliminary experiments for an anastomosis task in both simulated and real robotic environments towards this goal are presented.

## **PERSONNEL RECEIVING SALARY FROM THE RESEARCH EFFORT**

- Dr. Jeff Berkley (Mimic Technologies, Principal Investigator)
- Dr. Rajesh Kumar (Johns Hopkins University, Principal Investigator)
- Dr. Gregory Hager (Johns Hopkins University, Co-Principal Investigator)
- Yixin Gao (Johns Hopkins University, PhD Candidate)