Technical Correspondence

Expert Surgeons Can Smoothly Control Robotic Tools With a Discrete Control Interface

Marcia K. O'Malley, *Senior Member, IEEE*, Michael D. Byrne, Sean Estrada, Cassidy Duran, Daryl Schulz, and Jean Bismuth

***Abstract*—Objective assessment of surgical skill is gaining traction in a number of specialty fields. In robot-assisted surgery in particular, the availability of data from the operating console and patient-side robot offers the potential to derive objective metrics of performance based on tool movement kinematics. While these techniques are becoming established in the laparoscopic domain, current assessment techniques for robotic endovascular surgery are based primarily on observation, checklists, and grading scales. This work presents an objective and quantitative means of measuring technical competence based on analysis of the kinematics of endovascular tool tip motions controlled with a robotic interface. We designed an experiment that recorded catheter tip movement from 21 subjects performing fundamental endovascular robotic navigation tasks on a physical model. Motion-based measures of smoothness (spectral arc length and number of submovements) were computed and tested for correlation with subjective scores from a global rating scale assessment tool that has been validated for use when performing manual catheterization. Results show that the smoothness metrics that produced significant correlations with the global rating scale for manual catheterization show similar correlations for robotic catheterization. This finding is notable, since with the robotic interface, tool tip motion is commanded discretely via a control button interface, while in manual procedures the tools are controlled through continuous movements of the surgeon's hands. Logistic regression analysis using a single motion metric was capable of classifying subjects by expertise with better than 90% accuracy. These objective and quantitative metrics that capture movement quality could be incorporated into future training protocols to provide detailed feedback on trainee performance.**

***Index Terms*—Medical robotics, motion analysis, human-robot interaction.**

I. Introduction

Surgical robotic systems for endovascular procedures are gaining traction in the field of vascular surgery. Several robotic platforms (e.g., Hansen Magellan and Sensei, Stereotaxis Niobe, and Catheter Precision Amigo) are now commercially available for performing complex endovascular surgeries [1]. Even though the use of surgical robots for endovascular procedures is on the rise, little research has been done on the training necessary to gain proficiency when operating with these robotic devices. Traditionally, Halsted's "learning by doing" apprenticeship model has been used to assess technical competence for many domains in vascular surgery [2]. However, this "see one, do one, teach one" approach is time consuming, human resource intensive, and lacking in data-driven curricula and objective skill assessments [3], [4]. The Halsted model has also been criticized for being somewhat unstructured, because a resident's experience is based on the particular cases to which they are exposed. Vedula and Hager argue for the potential for surgical data science (SDS) to support skill evaluation and training in surgical domains, and note that surgical training has relied on assessment methods that are subjective and based on observation, a resource intensive approach [5]. In vascular surgery in particular, evaluation of skill has relied strongly on assessment of technical skills performance, with checklists and global rating scales the most commonly used [6].

Objective assessment of surgical skill in the laparoscopic domain is more developed than in the endovascular domain (see [7] for an extensive review). Previous laparoscopic studies have measured hand and instrument movements to assess the skill level of novice and expert surgeons operating the da Vinci robotic surgical device by analyzing qualitative [8] and quantitative [8]–[10] metrics and statistical models [11]–[13]. However, studies have shown that skills in one surgical paradigm (e.g., open surgery) do not usually correlate with skills in other surgical paradigms (e.g., minimally invasive surgery or endovascular surgery) [4]. Thus, there is a need to study skill performed using robotic platforms specifically in the context of endovascular procedures.

The quantitative metrics that have been investigated previously to assess performance on surgical robots can be classified as outcome-based (e.g., completion time), kinematic-based (e.g., peak speed), and motor control-inspired (e.g., jerk or number of submovements). A previous study showed the usefulness of different performance-based and smoothness-based metrics to quantify the performance of novice and expert surgeons when performing simple point-to-point movements using a Da Vinci Si robotic device [10]. Additional kinematic-based measures, such as path length, have also been used to differentiate expert versus novice behavior by studying movement of surgical instruments in robotic surgery [9], [14]. Recent studies have begun investigating the utility of time-based quantitative metrics to differentiate skill level between novice and expert surgeons performing endovascular procedures using robotic systems [15]. Fard *et al.* presented a machine-learning-based approach to classify expert and novice surgeons on a suturing task using six movement features and reported a classification accuracy of 85.7% [16]. Given the success of quantitative assessment of laparoscopic surgical skill using motion-based performance metrics,

Manuscript received January 21, 2018; revised October 2, 2018 and May 16, 2019; accepted May 18, 2019. Date of publication June 19, 2019; date of current version July 13, 2019. This work was supported in part by the National Science Foundation Grant IIS-1638073. *(Corresponding author: Marcia K. O'Malley.)*

M. K. O'Malley, and S. Estrada are with the Department of Mechanical Engineering, Rice University, Houston, TX 77005 USA (e-mail: omalleym@rice.edu).

M. D. Byrne is with the Department of Psychological Sciences, Rice University, Houston, TX 77005 USA (e-mail: byrne@rice.edu).

C. Duran, D. Schulz, and J. Bismuth are with the DeBakey Heart and Vascular Center, Houston Methodist, Houston, TX 77030 USA (e-mail: caduran@me.com; dgschulz@houstonmethodist.org; jbismuth@houstonmethodist.org).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/THMS.2019.2919744

(a) RCM (b) Remote Workstation (c) Surgical Tools

Fig. 1. (a) Hansen Magellan Remote Catheter Manipulator (RCM) (b) Hansen Magellan remote workstation featuring interactive motion controller, button interface, joystick and monitor (c) Steerable inner catheter (leader), guidewire, and outer catheter (sheath) used with the Magellan robotic system.

the current research explores the applicability of these objective measures for robotic surgery in the endovascular domain using the Hansen Magellan system (see Fig. 1).

In our previous work, we showed that metrics that capture movement smoothness produce statistically significant correlations with the observation-based assessment metrics in manual catheterization and can be used to differentiate skill among participants [17]. However, whether smoothness-based metrics could be used to assess performance in robotic catheterization is unclear, especially since the robot is controlled using a discrete button-type interface. In this work, we examine the potential to extend motion-based measures of surgical skill performance during tasks performed on physical models and surgical simulators to robotically executed endovascular tasks using the Hansen Magellan device. The experimental tasks that we use are designed to test different fundamental skills in endovascular surgery and are not procedure-specific. The tasks are representative of movements performed frequently during endovascular surgical cases.

Our primary goal is to understand whether motion-based smoothness metrics that have shown strong correlation with skill in manual endovascular surgery are also applicable to tasks performed with a robot outfitted with a user interface that does not mimic manual task performance. We investigated how a button-based control interface, where the motion is inherently discrete, affects surgical skill assessment using these smoothness-based objective performance metrics. An additional goal of this study is to investigate whether previous experience in manually performing endovascular procedures correlates with better performance when accomplishing the same procedures using the Hansen Magellan robotic device. In the laparoscopic domain, it has been shown that robotic surgical platforms provide novice robotic surgeons an early and persistent enabling effect that resulted in performance comparable to expert robotic surgeons [14]. In contrast, Nisky and Okamura showed that expert robotic surgeons were more effective at completing even simple movements with the robot compared to novices [10]. To understand the role of prior manual endovascular experience on robotic endovascular performance, we conducted logistic regressions for the motion based metrics computed during the performance of the fundamental tasks using the Magellan for two cases, once using experience groups based on previous experience performing robotic endovascular procedures, and once based on previous experience performing manual endovascular procedures. A tertiary goal is to investigate possible learning effects when performing tasks using the robotic device.

II. Methods

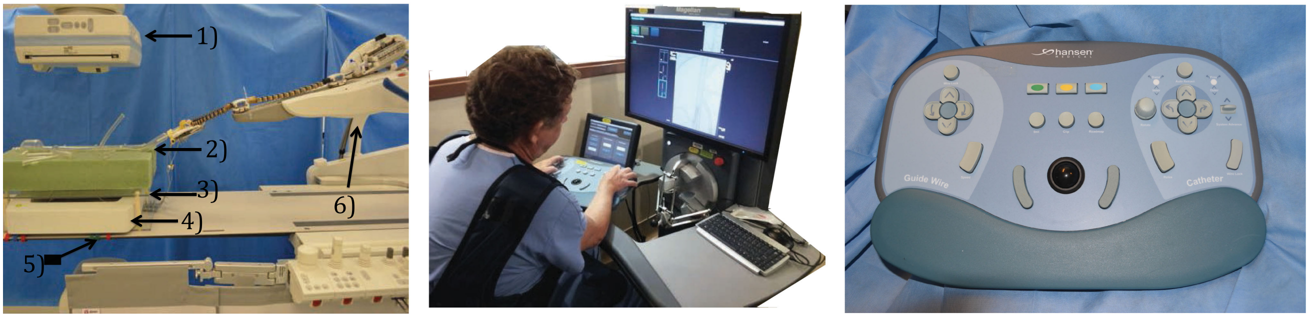
In this study, motion analysis, along with correlations and comparisons were used to identify quantitative metrics to differentiate the skill level of 21 subjects while they performed a set of fundamental endovascular surgical tasks using robotic catheterization with the Hansen Magellan device.

A. Hansen Magellan System

The Hansen Magellan robotic catheterization system (see Fig. 2) was designed to cannulate peripheral vessels and deliver simultaneous distal tip control of a catheter and a sheath from a centralized, remote workstation [18]. The Magellan system was engineered to meet the needs of vascular surgeons, interventional cardiologists and interventional radiologists [19]. Additionally, the system allows clinicians to perform complex surgical procedures, such as vessel navigation, selective angiogram generation, robotic guidewire control, and therapeutic device placement and delivery [19]. The Hansen Magellan system includes a "master" input device (either push-button or joystick) at a remote workstation that is used to provide catheter position commands to a remote catheter manipulator (RCM), which drives a steerable guiding catheter system that contains a flexible outer sheath, leading catheter, and a guidewire. A remote catheter control interface enables intuitive advancement and steering of the endovascular tools [19]. The RCM and catheter movements are controlled via motors and tension wires that ultimately determine the position of the outer sheath, catheter, and guidewire tips. The surgeon observes the task using live X-ray (fluoroscopic) images acquired with a mobile image intensifier. Fig. 1a–1c shows the RCM, surgical tools, and remote workstation for the Magellan system. The surgical tools can be advanced using a button interface or a 3-D joystick, both of which enable the surgeon to select the desired tool (guidewire, leader, or sheath), and then advance, retract, and rotate the tool. With each button press, the surgeon can advance the selected tool by a fixed increment. The tools can be moved continuously at a fixed rate by holding down the advance, retract, or rotate buttons. Preliminary experiments showed that novice users were more easily able to understand how to control the robotic device when the discrete button interface was used, compared to the 3-D joystick. Therefore, for this work, participants were instructed to only use the button interface to control the surgical tools.

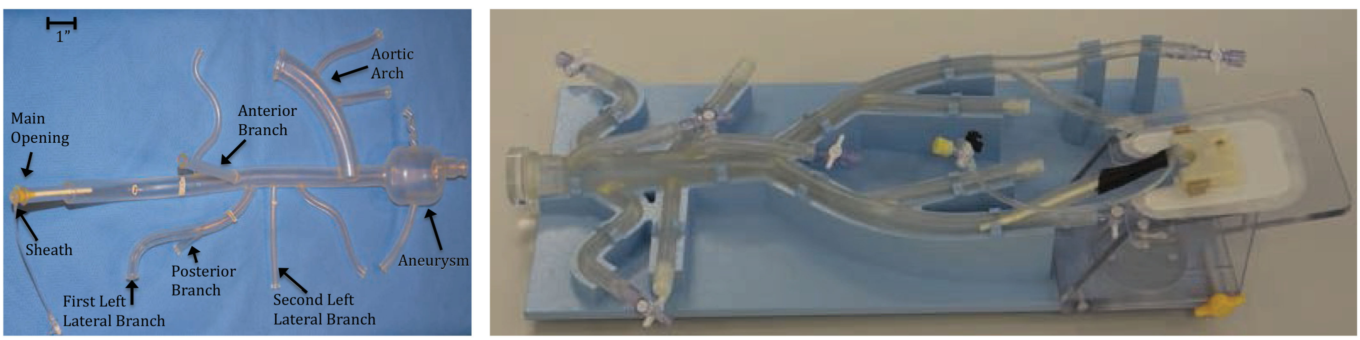
B. Fundamentals of Endovascular Skills Model

The Fundamentals of Endovascular Skills (FEVS) model was recently designed and constructed to use as a training tool for endovascular surgeons similar to the models currently used in the fundamentals of laparoscopic skill (FLS) curriculum [17]. The FEVS model is nonanatomical (though, anatomically inspired) and was designed to



(a) Robotic Catheterization Setup (b) Subject Manipulating Robot (c) Discrete Button Interface

Fig. 2. (a) Magellan setup, with 1) the face of the Siemen's C-arm to generate fluoroscopic images, 2) Fundamentals of Endovascular Skills (FEVS) model with styrafoam housing, 3) plexi-glass casing, 4) Window Field Generator, 5) velcro straps to secure setup to patient table, and 6) Remote Catheter Manipulation (RCM) device. (b) Subject manipulating the RCM and surgical tools from the remote workstation. (c) Button interface to control Magellan tools.



(a) FEVS Model (b) Training Model

Fig. 3. (a) FEVS model. (b) Model used for training on the Hansen Magellan robotic catheterization system.

provide a consistent environment in which to assess basic endovascular skills, as opposed to assessing specific procedures. The model was designed so that fundamental skills of endovascular surgery could be assessed, including basic catheter and guidewire skills, selective catheterization of all vascular beds, and arteriography of all vascular beds [20]. The FEVS model is shown in Fig. 3, with main features labeled.

C. Subjects

21 subjects (19 male, 2 female) participated in this study performed at the Houston Methodist Hospital. Seven subjects were either cardiology or vascular residents, six were cardiovascular fellows, six were attending physicians, one was a lab technician with significant expertise performing endovascular procedures in nonclinical scenarios, and one was an expert at driving the Magellan robotic device. The subjects ranged in experience from residents who were less than a year removed from medical school to attending surgeons with more than 20 years of experience in cardiovascular surgery. To investigate differences in skill, we grouped our participants based on their prior experience using the robotic surgery system. We defined novice subjects (16; 6 residents, 6 fellows, 3 attending, average age 35) as those having less than 30 h of prior experience using the robot, while expert subjects (5; 1 fellow, 3 attending, 1 lab technician, and 1 Magellan expert, average age 45) had more than 30 h experience using the robot. Because the robotic platform was not widely available technology, and because a limited number of procedures were approved to be performed with the robot at the time of our study, we opted for hours spent on the robot versus caseload to define expertise.

D. Robot Training

Prior to performing any tasks during the actual experiment, all subjects received one 90-min training session covering the basic operations required to drive the Hansen Magellan robotic catheterization system. This training was meant to give all subjects a fundamental understanding of the operation of the Hansen Magellan in an effort to ensure that data collected from the robot was representative of a subject's ability to perform endovascular tasks and not on a subject's familiarity with the basic operation of the robotic device. Fig. 3 shows the model that was used for training on the robot, which has been the standard model used for multiple years to train individuals on the operation of the Hansen Magellan device. The training model has no direct relationship to the FEVS model since each of these models was designed for a different purpose.

E. Tasks

The subjects performed four fundamental endovascular tasks: catheterization into the anterior, first left lateral, posterior, and second left lateral branches in the FEVS model. The first and second left lateral branches were cannulated while the portable imaging system was at 0° (or, Anterior/Posterior), the anterior branch was cannulated at 75° Left Anterior Oblique (LAO), and the posterior branch was cannulated at A/P until the catheter was sufficiently inside of the first left lateral branch, where the rest of the task was performed at 75° LAO. The sheath, catheter, and guidewire (shown in Fig. 1c) were controlled from the remote workstation and moved to navigate to the branch of interest. The goal was to move the catheter tip to an identified target point in the physical FEVS model (between 2–4 cm inside of the branch of interest) within five minutes. Participants completed each of the four tasks three times, once each in three separate sessions that occurred on different days within a one month period. In each session, the order of presentation of the tasks was counterbalanced. Subjects were not informed that skill learning would be evaluated.

F. Motion Analysis

The kinematic movement of the catheter tip was recorded using electromagnetic (EM) tracking technology to record the position and orientation of the catheter tip while using the robotic device to navigate into the physical model. A Northern Digital Incorporated Window Field Generator was used due to its ability to track flexible surgical tools while minimizing interference from live fluoroscopy to the generated electric field. The x, y, and z position and orientation about the x and y axes were read from a single 5-DOF sensor inserted into the catheter tip, with data collection occurring at a rate of 40 Hz.

G. Global Rating Scale for Endovascular Surgery

As described in [17], the Global Rating Scale for Endovascular Performance (GRSEP) is a structured grading tool that is used by a senior clinician to assess each subject based on their performance in endovascular skills. Even though the GRSEP has not been validated to assess performance using the robotic device, the grading tool was adapted and used during this experiment to compare structured grading results with assessment using motion analysis. One GRSEP was completed for each session (all subjects completed three sessions; therefore, three GRSEPs were conducted for each subject) as they completed the four endovascular tasks. The efficiency, wire/catheter manipulation, and device usage scores (each score measured on a scale of 1 to 5) were the portions of the global rating tool that were most applicable to identify trends with the computed motion metrics. A combined score of efficiency plus wire/catheter manipulation plus device usage (score from 0–15) was computed and compared to the quantitative metrics computed from data obtained from the EM sensors for catheterization tasks performed using the robot.

H. Quantitative Metrics

The notion of movement smoothness as an indicator of expertise is based on fundamental principles of human motor control, and movement smoothness is widely acknowledged as a characteristic of skill [21]. The quantitative metrics that characterize tool tip movement smoothness used in this study are computed from instrument kinematic data (catheter tip position and velocity). Each of the motor control inspired metrics uses a different method of computing movement smoothness (see [17] for full details about the metrics and equations necessary for calculating them). We used submovement extraction techniques [22] to derive four metrics (submovement duration and number of submovements, each computed with support-bounded lognormal (LGNB) curves [23] and minimum jerk profile curves [22] as the basis for the submovement extraction optimization procedure). We also used spectral arc length (SAL), a unitless frequency domain measure of movement smoothness, as a metric [24]. We chose these five metrics since they showed the strongest correlations with skill in manually performed endovascular tasks [17]. These metrics were computed using tangential speed data to assess skill level of surgical interventionalists while accomplishing tasks on the robot.

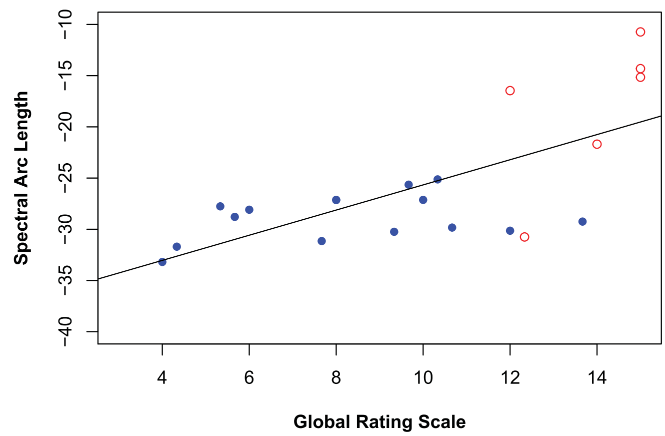


Fig. 4. Scatter plot showing relationship between spectral arc length and GRSEP. Novices are denoted with blue points and experts with red ○, with expertise based on prior exposure to robotic procedures. Lower values of SAL indicate smoother and more expert-like movements of the tools.

III. Results

A. Data Post-Processing

Two post-processing steps were conducted on data obtained from the electromagnetic sensors. In order to eliminate high-frequency background noise, the motion data were filtered with a Savitzky–Golay (S–G) filter using a third-degree polynomial with an 81-frame window. S–G filtering was applied to the measured motion data because of its ability to eliminate most of the noise while preserving the shape qualities of important peaks. For ANOVA-based analyses, observations more than three interquartile range (IQRs) from the cell mean were replaced with the subject's mean. A total of 9 points out of the 1260 points (21 subjects, 3 sessions, 4 tasks, 5 metrics) were replaced, which is less than 1% of the observations. Other approaches to outlier replacement are less conservative than our method of replacing with the subject mean, and a number of studies have been conducted to analyze in detail the effects of outlier replacement on statistical analysis of data (see, for example, [25], [26]).

B. Motion-Based Metrics and Global Rating Scale

Data were obtained both from the movement of the catheter tip and the global rating scale while subjects executed each of the experimental tasks using robotic catheterization. As described earlier, the GRSEP rating scale is a subjective measure of performance generated by an expert observer. Establishing relationships between these subjective measures and our objective motion-based measures is a first step toward validating these measures; it is not clear how meaningful motion metrics are when the motions are mediated through a button-controlled surgical. For each surgeon, one GRSEP measure was computed and compared with the average of each of the five motion metrics computed from data collected across all four tasks and all three sessions.

The metrics based on submovements and spectral arc length all produced significant correlations. Number of submovements produced $r(19) = 0.59$, $p = 0.005$ for the LGNB method and $r(19) = 0.69$, $p < 0.001$ for the minimum jerk method. Submovement duration yielded $r(19) = 0.73$, $p < 0.001$ for the LGNB method and $r(19) = 0.63$, $p = 0.002$ for the minimum jerk method. Spectral arc length produced similar results, $r(19) = 0.69$, $p < 0.001$. A scatter plot for spectral arc length appears in Fig. 4. Note that most of the highest ratings and best spectral arc lengths come from the expert surgeons.

C. Discriminating Novices From Experts

Another way to validate the motion metrics is to see the extent to which they can discriminate experts from novices. Binary classification using a continuous variable as the predictor can be done statistically with logistic regression. Statistical significance is, of course, important, but in the case where the regression is being used as a classifier, another measure is even more critical: classification accuracy. Essentially, if the distributions for the two groups on the predictor variable do not overlap, then group membership can be perfectly predicted. For example, if all the expert surgeons score higher on the metric than all the novice surgeons, then the metric can perfectly separate the groups. If the distributions overlap, however, then the classification will be imperfect. Classification accuracy thus reflects the degree of overlap of the two distributions and is an index of how discriminable the two groups are according to that predictor.

For instance, using GRSEP to predict expert versus novice produces a statistically significant result with $p = .036$. Classification accuracy is good at 86%, but this means it still misclassifies 3 of the 21 surgeons. Fortunately, the motion metrics generally do even better. The number of submovement metrics classify all but 2 correctly (90.5% correct classifications, $p = 0.03$) and the submovement duration metrics and the spectral arc length metrics classify all but one surgeon correctly (95% correct classifications; $p = 0.03$). Those regressions each used a single predictor. A logistic regression using two predictors, spectral arc length and submovement duration LGNB, produced 100% classification accuracy. Logistic regression is a procedure that generally requires large amounts of data to produce significant results. The fact that results were this strong with such a small sample (only 21 surgeons) is highly encouraging.

Note that these results held only if "expert" is defined as "expert with the robot" and not merely "expert at endoscopic surgery." To explore the importance of how expertise is defined, first we reclassified our participants based on their previous manual (not robotic) endovascular surgery experience. Here we defined novice subjects (11, 6 residents, 4 fellows, and 1 Magellan expert) as those having performed less than 50 previous cases and experts (10, 3 fellows, 6 attending, and 1 lab technician) as those having performed more than 50 previous cases. We had several surgeons who were experts with endoscopic surgery in general and novices with the robot, and we had one surgeon who was an expert with the robot and a novice in endoscopic surgery; thus, these are not identical groups. Overall, 10 subjects were novices on both platforms, 5 subjects were experts on both platforms, 5 subjects were experts in traditional surgery but novices on the robotic platform, and 1 subject was a novice in traditional surgery but an expert on the robotic platform. Motion metrics derived from data acquired during task completion on the Magellan robotic system generally failed to discriminate between individuals whose performance group (expert versus novice) was based on their prior manual surgery experience; the best metrics (submovement duration and spectral arc length) produced only 67% accurate classifications and were not statistically significant (best $p = 0.21$). Thus, the motion metrics show not who is an experienced surgeon overall, but who is an expert with the robot.

D. Learning

Another important feature of a measure is that it should show improvement as skill level increases. Surgeons who are already highly skilled should show little or no improvement over the sessions of our experiment, but less skilled surgeons should improve, and this should be reflected in the motion metrics as well. Thus, there should be an interaction between novice/expert and session, with the expert group having a shallow (or even zero) slope across sessions and the novice group showing more marked improvement.

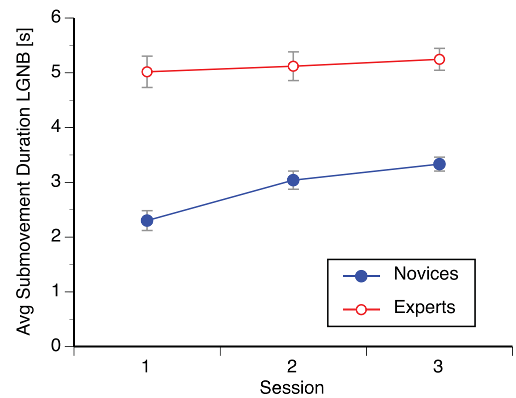


Fig. 5. Session by ability group interaction when performing catheterization on the physical model using the robot. Error bars represent the nonpooled standard error of the mean. Expertise is denoted based on prior experience with robotic procedures.

Submovement duration reflected exactly this pattern for both minimum jerk and LGNB methods; see Fig. 5. In both cases, the interaction was statistically significant –$F(2, 38) = 6.56$, $p = 0.004$ for LGNB and $F(2, 38) = 4.28$, $p = 0.021$ for minimum jerk– and simple effects analysis confirm no effect of session for the expert group and a significant effect of session for the novice group.

However, results were not strong for the other metrics. Number of submovements showed hints of the interaction under both methods (both $p = 0.09$), but there was no evidence whatsoever for spectral arc length ($p = 0.42$). In all cases, the main effect of session was significant (worst $p = 0.016$), indicating that the surgeons at least showed improvement in these measures, but there was insufficient evidence to conclude that this improvement was differential between novices and experts.

Overall, then, results here were somewhat mixed. Some metrics showed the expected pattern of more improvement for the less-skilled group relative to the more-skilled group, but others did so to a lesser extent or not at all.

IV. Discussion

In this work, we applied motion-based metrics that have been shown to reliably quantify movement smoothness in manually performed endovascular surgery tasks [17] to procedures performed with a robot. Both spectral arc length and submovement analysis metrics produced significant correlations with the GRSEP structured grading scale assessment of surgical skill. While at first this may seem unsurprising, it is important to highlight the unique differences between robotic and manual performance of endovascular surgery. With the Magellan system, the surgeon is controlling the advancement, retraction, and curvature of the catheters (leader and sheath) and guidewire with a pendant (button and knob) interface. To steer and manipulate the surgical tools, buttons are held or pushed repeatedly (depending on the preference of the surgeon) in a position-controlled fashion. Compared to the continuous control task nature of manual endovascular surgery, this is a discrete control task. Accordingly, we were surprised that the smoothness metrics that capture efficiency of tool tip movements still showed such significant correlation with the structured grading assessments. It was not obvious that the discrete control nature of the robotic surgery interface would produce smooth and coordinated movements of the surgical tools that could be quantified using metrics traditionally applied to continuous control tasks.

Logistic regression analysis showed that individual motion metrics can classify subjects into novices and experts with high accuracy. Furthermore, a linear combination of these metrics resulted in a perfect classification. It is unusual for logistic regression techniques to work well without extremely large data sets. Here, we had a relatively small data set for 21 subjects performing a navigation task to four different targets on three different days. Despite the small data set, the logistic regression based on the motion metrics achieved classification accuracies that exceeded those performed with the structured grading scale scores to predict expert versus novice. It is important to keep in mind that because the sample size was not large and was unevenly distributed, some over-fitting may have occurred; it is thus not guaranteed that this level of classification accuracy will generalize. Nonetheless, even if imperfect, such a finding suggests that these motion-based metrics could be used during task performance to evaluate skill of the trainee, and possibly even provide real-time feedback about their performance in terms of these metrics which are known to correlate to expertise. One challenge is that for this work, the computation of the SAL and submovement metrics was completed as a post-processing step. To provide feedback in real-time to trainees, the metrics would need to be computed in real-time as well, which presents additional challenges. Currently, the SAL and submovement metrics incorporate data from one complete trial, and thus could only be used as end-of-trial feedback. Further, the submovement algorithms are computationally expensive and don't lend themselves well to real-time implementation. Finally, it is unclear how to best present these metrics to subjects in a way that they would find intuitive and useful.

The logistic regression also showed that prior experience on the robot had an impact on performance of the fundamental tasks. Motion metrics derived from data acquired during task completion on the Magellan robotic system generally failed to discriminate between individuals whose performance group (expert versus novice) was based on their prior manual surgery experience. As has been seen in laparoscopic surgery, there can be a beneficial effect of prior robot experience on task performance [10]. The motion metrics that we propose correctly classify expert robotic endovascular surgeons, not necessarily expert manual endovascular surgeons. Individuals with expertise in manual procedures may require additional training if they wish to operate effectively with the robotic platform.

The significant main effect of session coupled with the session by ability group interactions implied possible improvement in motor skill execution for novice surgeons. This result was anticipated, since subjects in the novice ability group had limited to no experience using the robotic catheterization system. Although all subjects went through the same initial training regiment on the robot prior to performing the tasks in the experiment, the previous hours of robot experience of the expert subjects appeared to help them perform the tasks with more consistency and a higher degree of smoothness. One should be careful to extrapolate that the results that suggested learning occurred also implied that actual surgical skill was learned, as opposed to subjects simply becoming more comfortable with the simulator over time. In fact, prior work assessing performance in a Fitts' type targeting task completed with a da Vinci surgical robot showed that even for these very simple movements that don't mimic surgical tasks, there were differences in performance between those experienced on the robot and those not experienced on the robot [10]. Although our results are consistent with session-to-session learning, whether observed improvements correlate with improvements in actual surgical skill is still an open research question.

V. Conclusion

We seek consistent and standardized methods of evaluating surgical performance when using robotic devices for minimally invasive surgical procedures. In this research, we recorded tool tip movements during endovascular surgery navigation tasks performed with a Magellan robotic system and an inanimate vascular model. Simultaneously, assessments using standardized structured grading tools were conducted. Correlations between quantitative metrics of tool movement smoothness and assessment data were computed and demonstrated that spectral arc length and submovement analysis-based metrics are appropriate to apply to the discrete control task of robotic endovascular surgery. Further, we showed that logistic regression analysis using a single motion metric can classify subjects into robotic surgery novices and experts with high accuracy, and a linear combination of these metrics resulted in a perfect classification. Finally, a significant effect of session was observed, and with one metric we were able to identify a significant interaction between session and expertise group. Our findings suggest that these motion-based metrics could be used during task performance to evaluate skill of the trainee, and possibly even provide real-time feedback about their performance.

Acknowledgment

The authors greatly acknowledge the resources and facilities provided by the Houston Methodist Research Institute (HMRI) and technical assistance from Hansen Medical Incorporated and Northern Digital Incorporated (NDI). In addition, the authors appreciate the constructive feedback provided by Priyanshu Agarwal, William Jantscher, and Shivam Pandey.

References

[1] H. Rafii-Tari, C. J. Payne, and G.-Z. Yang, "Current and emerging robot-assisted endovascular catheterization technologies: A review," *Ann. Biomed. Eng.*, vol. 42, no. 4, pp. 697–715, 2014.

[2] B. N. Carter, "The fruition of halsted's concept of surgical training." *Surgery*, vol. 32, no. 3, pp. 518–527, 1952.

[3] R. K. Reznick and H. MacRae, "Teaching surgical skills—changes in the wind," *New England J. Medicine*, vol. 355, no. 25, pp. 2664–2669, 2006.

[4] C. E. Reiley, H. C. Lin, D. D. Yuh, and G. D. Hager, "Review of methods for objective surgical skill evaluation," *Surgical Endoscopy*, vol. 25, no. 2, pp. 356–366, 2011.

[5] S. S. Vedula and G. D. Hager, "Surgical data science: the new knowledge domain," *Innov. Surg Sci.*, vol. 2, no. 3, pp. 109–121, 2017.

[6] E. L. Mitchell *et al.*, "A systematic review of assessment of skill acquisition and operative competency in vascular surgical training," *J. Vascular Surgery*, vol. 59, no. 5, pp. 1440–1455, 2014.

[7] I. Oropesa *et al.*, "Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery," *J. Surgical Res.*, vol. 171, no. 1, pp. e81–e95, 2011.

[8] J. D. Hernandez *et al.*, "Qualitative and quantitative analysis of the learning curve of a simulated surgical task on the da vinci system," *Surgical Endoscopy Other Interventional Techn.*, vol. 18, no. 3, pp. 372–378, 2004.

[9] T. J. Tausch, T. M. Kowalewski, L. W. White, P. S. McDonough, T. C. Brand, and T. S. Lendvay, "Content and construct validation of a robotic surgery curriculum using an electromagnetic instrument tracker," *J. Urology*, vol. 188, no. 3, pp. 919–923, 2012.

[10] I. Nisky, A. M. Okamura, and M. H. Hsieh, "Effects of robotic manipulators on movements of novices and surgeons," *Surgical Endoscopy*, vol. 28, no. 7, pp. 2145–2158, 2014.

[11] R. Kumar *et al.*, "Assessing system operation skills in robotic surgery trainees," *Int. J. Med. Robot. Comput. Assisted Surgery*, vol. 8, no. 1, pp. 118–124, 2012.

[12] A. Jog *et al.*, "Towards integrating task information in skills assessment for dexterous tasks in surgery and simulation," in *Proc. Int. Conf. Robot. Autom.* IEEE, 2011, pp. 5273–5278.

[13] R. Kumar *et al.*, “Objective measures for longitudinal assessment of robotic surgery training,” *J. Thoracic Cardiovascular Surgery*, vol. 143, no. 3, pp. 528–534, 2012.

[14] V. Chandra *et al.*, “A comparison of laparoscopic and robotic assisted suturing performance by experts and novices,” *Surgery*, vol. 147, no. 6, pp. 830–839, 2010.

[15] C. V. Riga, N. J. Cheshire, M. S. Hamady, and C. D. Bicknell, “The role of robotic endovascular catheters in fenestrated stent grafting,” *J. Vascular Surgery*, vol. 51, no. 4, pp. 810–820, 2010.

[16] M. J. Fard, S. Ameri, R. B. Chinnam, A. K. Pandya, M. D. Klein, and R. D. Ellis, “Machine learning approach for skill evaluation in robotic-assisted surgery,” 2016, arXiv:1611.05136.

[17] S. Estrada, C. Duran, D. Schulz, J. Bismuth, M. D. Byrne, and M. K. O’Malley, “Smoothness of surgical tool tip motion correlates to skill in endovascular tasks,” *IEEE Trans. Human-Mach. Syst.*, vol. 46, no. 5, pp. 647–659, Oct. 2016.

[18] J. Bismuth, C. Duran, M. Stankovic, B. Gersak, and A. B. Lumsden, “A first-in-man study of the role of flexible robotics in overcoming navigation challenges in the iliofemoral arteries.,” *J. Vascular Surgery*, vol. 57, no. 2, pp. 14S–19S, 2013.

[19] F. Soleimani, F. Moll, D. Wallace, J. Bismuth, and B. Gersak, “Robots and medicine: Shaping and defining the future of surgery, endovascular surgery, electrophysiology and interventional radiology,” *Slovenian Med. J.*, vol. 80, no. 7, pp. 614–631, 2011.

[20] G. B. Zelenock, T. S. Huber, L. M. Messina, A. B. Lumsden, and G. L. Moneta, *Mastery of Vascular and Endovascular Surgery, Section I - Basic Considerations and Peri-Operative Care*. Philadelphia, PA, USA: Lippincott Williams & Wilkins, 2006.

[21] T. Flash and N. Hogan, “The coordination of arm movements: an experimentally confirmed mathematical model,” *J. Neuroscience*, vol. 5, no. 7, pp. 1688–1703, 1985.

[22] B. R. Rohrer, “Evolution of movement smoothness and submovement patterns in persons with stroke,” Ph.D. dissertation, Dept. of Mechanical Engineering. Massachusetts Inst. of Technology, Cambridge, MA, USA, 2002.

[23] R. Plamondon, A. M. Alimi, P. Yergeau, and F. Leclerc, “Modelling velocity profiles of rapid movements: a comparative study,” *Biol. Cybern.*, vol. 69, no. 2, pp. 119–128, 1993.

[24] S. Balasubramanian, A. Melendez-Calderon, and E. Burdet, “A robust and sensitive metric for quantifying movement smoothness,” *Biomed. Eng., IEEE Trans.*, vol. 59, no. 8, pp. 2126–2136, 2012.

[25] J. M. Orr, P. R. Sackett, and C. L. Dubois, “Outlier detection and treatment in i/o psychology: A survey of researcher beliefs and an empirical illustration,” *Personnel Psychol.*, vol. 44, no. 3, pp. 473–486, 1991.

[26] J. W. Osborne and A. Overbay, “The power of outliers (and why researchers should always check for them),” *Practical Assessment, Res. Eval.*, vol. 9, no. 6, pp. 1–12, 2004.