Towards Automated Performance Assessment using Velocity-based Motion Quality Metrics

Barathwaj Murali[1], Viony M. Belvroy[3,4], Shivam Pandey[2], Michael D. Byrne[2]
Jean Bismuth[3] & Marcia K. O'Malley[1]

***Abstract*— Acquiring proficiency in endovascular surgery requires a significant investment of time and resources, both from trainees who are developing skills and from experienced surgeons who must serve as evaluators. These experienced surgeons typically provide feedback to trainees using structured grading scales that offer a qualitative and subjective assessment of performance. To address these limitations, we previously established that spectral arc length (SPARC), a frequency-domain measure of movement smoothness, was a quantitative and objective indicator of surgical experience. Still, trainees have indicated that performance feedback based on SPARC is not intuitive or easily understandable. In this work, we evaluate the potential of alternative quantitative measures of endovascular tool navigation proficiency. One set of metrics is available from a commercial endovascular surgical simulator, and another set of metrics is derived from tool tip velocity profiles. Results indicate that average guidewire tip velocities and idle times (the amount of time the guidewire remains stationary) are significantly different across experience groups. In contrast, only one of the performance metrics currently implemented on the simulator shows significant differences across experience groups. Subsequent analysis showed that average velocity and idle time correlate strongly with SPARC for these tasks. These results support the potential of metrics based on tool tip velocity for real-time objective assessment of endovascular skill. Further, these metrics, which correlate strongly to movement smoothness, are likely to be easier for participants to interpret than feedback based on spectral arc length, which could positively effect training effectiveness.**

I. INTRODUCTION

Surgeons are increasingly adopting minimally invasive endovascular surgical techniques to perform a variety of diagnostic and therapeutic interventions, such as carotid artery stenting, abdominal aortic aneurysm repair, and aortic valve replacement [1], [2], [3]. Compared to open surgery, these techniques can result in shorter operation times and hospital stays, less blood loss, and lower rates of atrial fibrillation and systemic complications [2], [3]. Several studies observe a relationship between operator experience and postoperative complication rates [3], [4], with greater experience leading to higher chances of success, shorter procedure times and longer amounts of time before reoperation [5], [4].

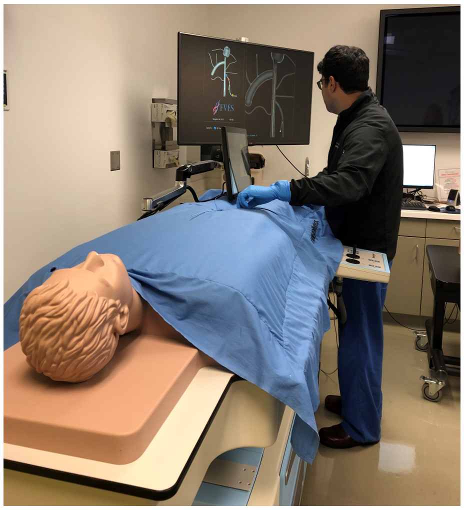

Fig. 1. Trainee performing endovascular navigation task with ANGIO Mentor Suite at Houston Methodist Institute for Technology, Innovation, and Education (MITIE)

Endovascular surgery requires acquisition of a significant amount of both procedural knowledge and manual dexterity. For a typical procedure, surgeons must be able to manipulate guidewires and catheters to provide and maintain access to target structures and for visualization and deployment of therapeutic devices such as endografts or valve replacements [6]. Depending on the procedure, surgeons may also exchange guidewires and catheters within a sheath to improve access to a vessel and reduce the risk of vascular injury [6]. Simulators, such as the one shown in Fig. 1, are frequently used during training to allow rehearsal of procedural steps and practice to improve manual dexterity [7].

Traditional methods of evaluating surgical performance consist of examiners, often senior-level attending physicians, completing Likert-style surveys such as the Objective Structured Assessment of Technical Skill (OSATS) and the Imperial College Evaluation of Procedural Skill (ICEPS) to measure general and procedural skills [8], [9], [10]. For endovascular surgery in particular, the Global Rating Assessment Device for Endovascular Skill (GRADES) rating scale is the recommended assessment tool, testing procedural efficiency and autonomy, fluoroscopic imaging and contrast use, device deployment, and tool manipulation [11]. Despite their widespread use, these qualitative assessments require the time and resources of experienced surgeons to administer the assessments, and the tools lack objectivity when measuring performance [12].

There is a growing body of research exploring the use of

---

*This work was supported by National Science Foundation grant IIS-1638073 and the National Science Foundation Graduate Research Fellowship (GRFP) Grant No. 1842494.

[1]Department of Mechanical Engineering, Rice University, Houston, TX, 77005 USA [2]Department of Psychological Sciences, Rice University, Houston, TX, 77005 USA [3]DeBakey Heart & Vascular Center, Houston Methodist, Houston, TX, 77005 USA [4]University Medical Center Utrecht, Utrecht, Netherlands

tool motion data from instrumented or simulated surgical tools as the basis for *quantitative and objective* performance assessment [13], [14], [15]. Kinematic and force data obtained directly from surgical procedures can provide a basis for more comprehensive assessment frameworks, with techniques ranging from validating global performance metrics to developing probabilistic models with motion data from experienced surgeons [16], [17], [18]. While there has been a considerable effort along these lines for laparoscopic and robotic surgery, application to the endovascular domain remains underdeveloped. We previously established that movement smoothness, a performance metric used in human motor control research that strongly indicates healthy and coordinated movement, provides a promising means for objective performance evaluation in endovascular procedures, given its strong correlation with experience level determined by global rating scales across manual, simulation, and robotic platforms [14], [19]. In this prior work, all data analysis was completed after data collection occurred, due to challenges in extracting tool tip motions from the various platforms and significant data processing efforts.

While it is important to have objective measures of surgical performance like movement smoothness, the real potential of such assessment is to provide feedback to trainees *as they practice*. Towards this end, we have explored methods for real-time feedback of movement smoothness during dexterous manual tasks similar in nature to endovascular procedures [20]. While we were able to demonstrate that movement smoothness feedback (provided via a vibrotactile cue) influenced task completion strategies in more beneficial ways than position feedback or no feedback, subjects reported that the feedback was difficult to understand and interpret [21]. Therefore, we seek measures of performance that are more easily computed, and more intuitive for users to understand and act upon.

In this paper, we evaluate the suitability of a suite of performance metrics, developed for a commercially available endovascular simulator, to assess manual skill for a set of navigation tasks. We also extend the comparison to include a frequency-domain measure of movement smoothness known as spectral arc length (or SPARC [22]) and other quantitative performance metrics derived from guidewire and catheter tool tip velocity profiles, including average velocity, idle time, and path length. Section II describes our experimental methods. Results, including both analysis of variance and correlations between metrics, are presented in Section III. We discuss our findings in Section IV.

II. METHODS

We asked participants to complete endovascular navigation tasks using the ANGIO Mentor system pictured in Fig.1. Participants were asked to manipulate the guidewire and catheter to reach targets in a preloaded module containing a virtualized anatomical training model, illustrated in Fig.2. We assessed their performance using a variety of motion-based metrics and compared these metrics based on participants' expertise level as defined by prior caseload.

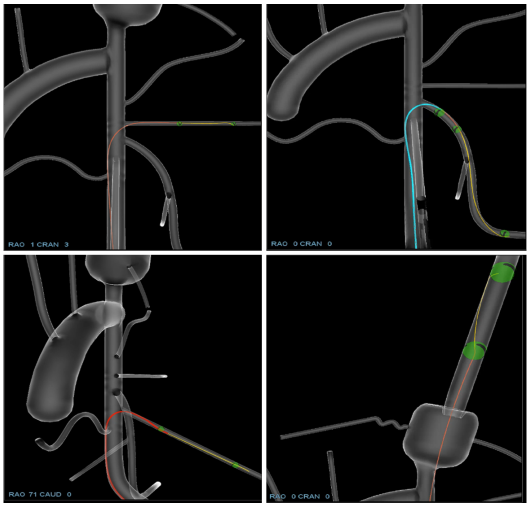

Fig. 2. Four navigation tasks with targets. a) Right angle bifurcation. b) Cannulation of branch vessel. c) Cannulation of aneurysmal branch vessel. d) Gate cannulation through aneurysmal segment.

A. Subjects

Subjects of all experience levels were recruited at various professional meetings of vascular and endovascular surgeons, as well as at the Houston Methodist Institute for Technology Innovation and Education (MITIE). A total of 52 individuals (40 male, 12 female, 30 novices, 11 intermediates, and 11 experts) participated in our study. The number of endovascular procedures performed with and without supervision determined the experience level of each subject, with novices defined as subjects who had performed less than 50 cases. Intermediates and experts were defined as subjects who had performed 50-500 cases and over 500 cases, respectively. This division of experience level by caseload is supported by evidence of a sharp change in procedural success rates after approximately 50-65 consecutive cases for abdominal aortic aneurysm repair, after which there was little appreciable change in success rates [23]. A similar result was observed in carotid artery stenting cases, in which a noticeable decrease in neurological complication and 30-day mortality rates occurred after the first 50 consecutive cases [1]. The novice group consisted of 18 students, 7 residents, 4 fellows, and 1 industry professional. The intermediate group consisted of 5 residents, 3 fellows, and 3 attendings. The expert group consisted of 3 residents, 7 attendings, and 1 physician assistant with experience in vascular surgery. Subjects represented a wide range of medical specialties, including anaesthesiology, general surgery, cardiology/cardiothoracic surgery, and vascular/endovascular surgery. All subjects provided informed consent for their participation and the study was approved by the Rice University Institutional Review Board.

B. Materials

We used the ANGIO Mentor Flex endovascular simulator (3D Systems, Littleton CO) at the professional meetings and an ANGIO Mentor Ultimate simulator (3D Systems, Littleton CO) at MITIE to collect motion data. Optical sensors on the simulator recorded the translation and rotation of practice guidewires, catheters, and sheaths inserted into the device, while a preloaded module containing a virtualized training model used by the Fundamentals of Vascular and Endovascular Surgery (FVES) platform in [11] simulated physical tool tip motions and streamed X, Y, and Z position data of each tool tip over a TCP network connection at varying sampling rates between 15-60 Hz. The module also streamed the differences between adjacent position values for computing velocity throughout each task.

C. Performance metrics

Performance metrics generated by the commercial simulator were recorded, along with a set of metrics computed from the tool tip positions and velocities provided by the data stream from the training module.

A set of on-board time and motion-based performance metrics computed by the simulation software were recorded. These metrics included maximum guidewire and catheter velocity, distance and number of occurrences of guidewire retraction (retrograde wire motion), the total distance travelled by the catheter without a leading wire, and the distance moved by the guidewire during cannulation while the catheter advanced into the vessel branch.

From the tool tip velocity data, we computed spectral arc length (SPARC), a frequency-domain measure of movement smoothness that was previously shown to be significantly correlated to experience level for endovascular procedures performed on manual, simulator, and robotic platforms [14], [19]. Its robustness to noise and sensitivity to small variations within the physiological range of healthy movement makes it a desirable metric for evaluating performance in the surgical domain [22]. Additionally, the relatively low computational burden of calculating SPARC shows promise for both online and offline performance evaluation and feedback [21].

Three additional metrics were computed from the velocity data, with a focus on identifying metrics that might be suitable for real-time performance feedback. Average tool tip velocity, idle time, and path length were identified as candidate metrics for their ease of interpretation and calculation from velocity data. Average tool tip velocity was calculated for each tool by using their tangential velocity profile. The average tool tip velocity presents another promising metric given its significant correlation to experience level in other catheter-based surgical domains such as transoesophageal echocardiography [24]. Idle time is defined by the total amount of time during a motion task in which the surgical tools remain stationary, and was shown to correlate to experience level in open surgery [25]. Similar to average velocity, idle times likely provide another measure of cognitive engagement, with higher values evident in individuals with less experience [25]. Idle time was calculated by determining the total amount of time in which the tools moved at tangential velocities below a threshold value, defined as 0.5 mm/s to account for motion artifacts such as deceleration of flexible tool tips. Since time-based metrics are task-dependent, we compute idle time as a fraction of the overall task completion time. Path length of each tool was calculated by integrating the tangential velocity profile across the entire task using trapezoidal numerical integration. The path length of tool tip motion correlates well with experience level in other domains [26], making it a metric of interest for endovascular performance evaluation. Lower path lengths may potentially correlate with experience level, similar to trends observed in video recordings of carotid artery stenting procedures [27].

D. Procedure

After consenting to participate in the study and prior to starting the first task, subjects completed a short survey that collected information on their level of medical training, specialty, familiarity with cardiovascular procedures and with using the commercial simulator, and the number of supervised and unsupervised endovascular cases performed. Participants recruited at the professional society meetings approached a booth containing a simulator, arriving in 15-20 minute rotations during which they completed between one and four target navigation tasks depending on the time available with the particular subject. Participants recruited from MITIE in Houston completed all four navigation tasks.

Each task consisted of navigating a guidewire, catheter, and if present, a sheath to the color-coded targets shown in Fig.2. The simulator computer screen displayed basic navigation guidelines before each task, and participants were given approximately 1-2 minutes to familiarize themselves with this information for navigating each tool to its respective target. After subjects were ready to proceed, they performed the task until either successfully reaching the tool targets or until the simulation timed out (at between 3 and 5 minutes, depending on task). As the simulation environment used an anatomically inspired model, the tasks corresponded to different navigation situations necessary for competency in endovascular surgery. Fluoroscopy was not simulated, but each task required the use of a virtual C-arm for facilitating target navigation. Almost all participants completed the right angle bifurcation task (see Fig.2), and then (time permitting) proceeded to complete additional tasks (cannulation of the branch vessel, cannulation of aneurysmal branch vessel, gate cannulation through aneurysmal segment) [11]. Most novice and intermediate subjects performed one or two tasks, while most expert subjects performed two to four tasks. Tasks were not repeated and each session did not exceed 15 minutes.

After completing their final navigation task, subjects completed an additional custom questionnaire that provided information on their perceived experience level and differences between experienced and inexperienced surgeons, as well as the amount of cognitive engagement necessary to correctly manipulate endovascular tools. The questionnaire also inquired as to the difficult aspects of endovascular navigation, and whether subjects preferred receiving feedback, either during or after each task.

E. Data analysis

A selection of on-board metrics that provided the most relevant indicators of motion quality calculated by the FVES module were recorded for each subject, along with the time series data of X, Y, Z position and differences between adjacent position values for each tool present in the task. The tool tip data provided the full trajectory of each tool from the beginning of the task until either each tool reached the target, or the simulation timed out.

A third order Savitzky-Golay filter with a window length of 21 samples was implemented to remove high frequency noise from the tool tip data while preserving the waveform shape of each signal [14]. The data were transformed to a constant sampling frequency of 60 Hz using linear interpolation for frequency analysis and calculation of SPARC before filtering. The tool tip velocity profiles were determined by taking the differences in tool tip position data provided by the simulator and dividing by the interpolated sampling time of 16.67 ms.

As in [28], any performance metric calculated from motion data containing critical failures, defined as instances in which the catheter advanced into the branch of interest before the guidewire during cannulation, was excluded as these can lead to severe complications in real-life procedures. Critical failures were detected by determining if the catheter tool tip crossed the opening of the vessel branch before the guidewire. Of the 115 individual motion trials across all subjects, 20 trials were excluded due to critical failures, resulting in data from 7 subjects (all novices) being removed from analysis. Outlier removal was not performed.

To identify the performance metrics that produced statistically significant differences between experience level, we performed a one-way independent measures ANOVA for each metric. Before performing each ANOVA, we verified that data for each metric and experience level followed a normal distribution and that the homogeneity of variance was satisfied using O'Brien's test [29]. Performance metrics for subjects who had completed more than one navigation task were averaged to provide a single value for each subject. Pairwise comparisons were performed using Tukey HSD for metrics that produced significant ANOVA results.

III. RESULTS

We compared various objective metrics, evaluated during the performance of simulated endovascular navigation tasks, across expertise groups. Then, we examined correlations between these metrics and a gold standard measure of tool movement smoothness that has been shown to correlate strongly with surgical skill. Finally, we evaluated participants responses to pre- and post- testing questionnaires.

A. Performance differences between experience levels

One metric provided on the commercial endovascular simulator, the distance travelled by the catheter without a leading wire, showed significant differences across our three expertise groups. For the remainder of the on-board metrics, there were no significant differences across expertise groups. Results are reported in TABLE I.

The velocity-based metrics calculated from the streamed guidewire tool tip data (see TABLE II), produced significant differences across expertise groups for SPARC $F(2,42) = 9.38, p < .001$, average velocity $F(2,42) = 10.66, p < .001$, and idle time $F(2,42) = 8.18, p = 0.001$. There was no significant difference in path length across our expertise groups. There were no significant differences in performance based on metrics computed from catheter tip motion.

Mean and standard error values for each velocity-based metric and experience level are shown in Fig. 3. Tukey HSD pairwise comparison tests indicated that group means of SPARC are significantly lower for experts ($M = -26.8, SD = 8.98$) than for novices ($M = -49.03, SD = 16.6$). The average guidewire velocity of experts ($M = 58.2, SD = 14.4$) was significantly higher than that of novices ($M = 34.7, SD = 12.3$) and intermediates ($M = 42.4, SD = 16.5$). Idle time indicated significant differences between experts ($M = 0.212, SD = 0.075$) and novices ($M = 0.358, SD = 0.109$).

Note that the effect size for group differences was quite large for all metrics other than path length, suggesting that these results are robust. Even the worst of these (idle time) was still better than the best metric from the FVES module.

B. Correlation of metrics with spectral arc length

Spectral arc length (SPARC) has been shown to strongly correlate with endovascular surgical skill [14], [19]. We computed the correlation between the on-board and velocity-based metrics with SPARC (see TABLE III) using robust linear regression to minimize the effects of potential outlier data. The on-board metrics from the FVES module on the ANGIO Mentor demonstrated weak to moderate correlations with SPARC, though three metrics were significantly correlated to SPARC (retrograde wire motion occurrences ($r(43) = 0.37; p = .013$), retrograde wire motion distance ($r(43) = 0.38; p = .010$), and the distance the catheter travelled without a leading wire ($r(23) = 0.48; p = .017$). In contrast, the velocity-based metrics of average velocity and idle time are moderately and significantly correlated to SPARC. Path length, also computed from tool tip velocity data, was not significantly correlated to SPARC. The correlations between velocity-based metrics and SPARC are illustrated graphically in Fig. 4. The linear regression of SPARC and average guidewire velocity produced the highest significant correlation ($r(43) = 0.72, p < .001$), followed by SPARC and idle time ($r(43) = 0.70, p < .001$).

C. Survey responses

Responses to surveys from all experience levels indicated that experts visibly possess motion fluidity and smoothness. Subjects with more experience suggested that gaining an intuition of catheter dynamics, wire control and proper visualization techniques is necessary for novices to gain competency, in addition to developing an ability to find alternative navigation techniques and avoiding damage to the vessel walls. Experts generally remarked that maneuvering the tool tip comes either as second nature or requires more planning only complex vasculatures, which likely suggests that they navigate tools faster and have less idle periods

TABLE I

ANOVA RESULTS AND EFFECT SIZES FOR ON-BOARD METRICS CALCULATED AT THE END OF EACH MOTION TRIAL BY THE FVES MODULE. STATISTICALLY SIGNIFICANT VALUES IN BOLD.

Metric	ANOVA Test Result	Effect Size (Cohen's *f*)
Max. Guidewire Velocity (in/s)	F(2, 41) = 0.12; p = .89	0.08
Max. Catheter Velocity (in/s)	F(2, 41) = 0.33; p = .72	0.13
Guidewire Movement during Cannulation (mm)	F(2, 41) = 1.07; p = .35	0.23
Retrograde Wire Motion Frequency	F(2, 41) = 0.32; p = .73	0.12
Retrograde Wire Motion Distance (mm)	F(2, 41) = 0.06; p = .94	0.05
Catheter Travel Without Leading Wire (mm)	**F(2, 21) = 3.51; p = .048**	**0.58**

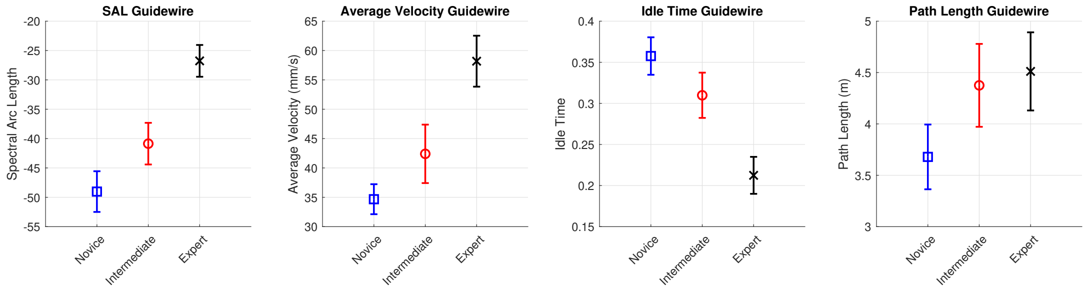

Fig. 3. Plots of mean and standard error of spectral arc length, average velocity, idle time, and path length of the guidewire for each experience level.

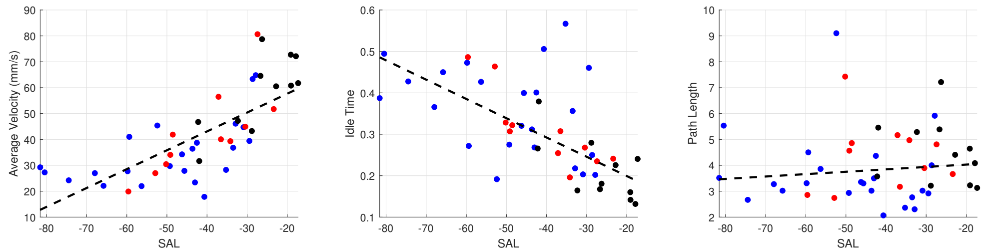

Fig. 4. Linear regression of velocity-based metrics (average velocity, idle time, and path length) and spectral arc length. Expert data is indicated with a black circle, intermediates with a red circle, and novices in blue.

TABLE II

ANOVA RESULTS AND EFFECT SIZES FOR PERFORMANCE METRICS CALCULATED FROM THE TANGENTIAL VELOCITY PROFILE OF GUIDEWIRE MOTION DATA. STATISTICALLY SIGNIFICANT VALUES IN BOLD.

Metric	ANOVA Test Result	Effect Size (Cohen's *f*)
SPARC	**F(2, 42) = 9.38; p <.001**	**0.67**
Avg. Velocity	**F(2, 42) = 10.66; p <.001**	**0.71**
Idle Time	**F(2, 42) = 8.18; p = .001**	**0.62**
Path Length	F(2, 42) = 2.67; p = .20	0.28

TABLE III

PEARSON R CORRELATION COEFFICIENTS AND ACCOMPANYING P VALUES FOR EACH METRIC COMPARED WITH SPARC.

Metric	Pearson r and p-value
Avg. Velocity	**r(42) = 0.72; p <.001**
Idle Time	**r(42) = 0.70; p <.001**
Path Length (m)	r(42) = 0.21; p = .16
Max Guidewire Velocity (in/s)	r(42) = 0.24; p = .11
Max Catheter Velocity (in/s)	r(42) = 0.03; p = .84
Wire Motion during Cannulation (mm)	r(42) = 0.21; p = .16
Retrograde Wire Motion Counts	**r(42) = 0.37; p = .013**
Retrograde Wire Motion Distance (mm)	**r(42) = 0.38; p = .01**
Catheter Travel Without Wire (mm)	**r(22) = 0.48; p = .017**

because of a greater knowledge of tool behaviors, along with more procedural knowledge.

IV. DISCUSSION

The value of objective assessment of surgical skill is well-recognized [12]. Our prior work has shown that tool tip movement smoothness is a robust metric for evaluating surgical skill in the endovascular domain [14], [19]; however, some movement smoothness metrics are computationally intensive, and real-time performance feedback based on movement smoothness has been reported to be difficult to understand and interpret [21]. In this paper, we explored the potential of a suite of metrics available on the ANGIO Mentor endovascular simulator to stand in for movement smoothness as an objective assessment of participants' skill levels. Then, we examined velocity-based performance metrics that are likely more intuitive for users to understand than movement smoothness. For all metrics, we examined correlations with our gold standard metric, spectral arc length, a frequency-domain measure of movement smoothness.

Of the on-board metrics provided on the commercial endovascular simulator, only the distance travelled by the catheter without a leading wire showed significant differences across our three expertise groups. One may question if a small sample size is the reason that other simulator metrics failed to show significant differences across expertise groups; however, since most of the effect sizes were small for these other metrics, it is unlikely that this would be the case. The only metric provided by the FVES module as an objective measure of manual skill that showed significant differences between expertise groups was catheter travel without leading wire. This metric was not available for most subjects since the data exported by the simulator contained missing values for several subjects and tasks. Further, this metric captures procedural knowledge differences between experienced and novice operators in tool manipulation strategies involving the relative motion between guidewires and catheters, rather than capturing manual skill. While this metric could be used as a basis for performance feedback that is more intuitive and easier to understand than movement smoothness, the measures computed from the guidewire tip velocity profile (e.g., average velocity and idle time) exhibited larger effect sizes and stronger correlations with SPARC.

Average velocity excelled as a reliable motion quality metric, showing the most pronounced differences between novices and experts as per its significance and strong effect size. Idle time also showed significant differences between novice and expert groups and had a strong effect size. Average velocity might also provide a measure of the cognitive engagement of the operator, as it is likely that more experienced operators with a greater amount of familiarity with endovascular navigation techniques would navigate the tools at higher average velocities than those with less experience.

Idle time provided further insight into other challenging aspects of the environment encountered during each navigation task, especially from a lack of familiarity with the visualization controls noticed across experience level. Idle tool motions likely depended not only on subjects orienting the view of the model to aid with navigation, but also on the amount of conscious motion planning effort required during each task [25]. The significantly lower idle times in experts despite their learning of the visualization controls implies that their motion planning ability is more developed than that of novices.

The non-significant differences between experience level for path length supports our previous observations that path length does not correlate with skill [28], although other studies have shown the opposite for similar disciplines [18], [27]. Path length as a candidate measure of skill is therefore likely to be procedure-dependent, and is not an indicator of expertise in our tasks due to more complex interactions between tools and vessel walls than in other domains.

Although SPARC is highly effective in determining experience level in this study as in our past work [14], [19], in a previous study, subjects considered it difficult to interpret the meaning of SPARC when provided as a performance feedback measure [20]. The high and significant linear correlations observed between SPARC, average velocity, and idle time (see Fig. 4) are encouraging results towards pinpointing alternative metrics that can convey performance feedback similar to SPARC, but in a more intuitive fashion.

The lack of significance of each motion metric derived from catheter tip motions provides evidence that proficiency of guidewire navigation contributes most to differences observed between experience levels.

Subjects occasionally preloaded the tools into the simulator and performed the navigation tasks in an environment prone to distractions. Despite these factors, the velocity-based metrics showed substantial differences between experience groups. Idle time, while effective, is more sensitive to external factors and dependent on the velocity threshold selected to define idle motion periods. Similar to [25], further refinement of selecting this threshold is necessary for maximizing its effectiveness as a performance metric.

Over 50% of novices indicated their preference for receiving online performance feedback while performing a motion task as a feature that future iterations of the system should include. Intermediate and expert subjects also showed a desire for receiving online feedback, although the number of survey responses for these groups were much smaller than that of the novice group. Overall, novices and intermediates would prefer some form of feedback over none, while experts would prefer either online feedback or no feedback.

V. CONCLUSION

We evaluated the suitability of a suite of performance metrics, developed for a commercially available endovascular simulator, to assess manual skill for a set of navigation tasks. Additionally, we computed a set of velocity-based objective metrics from data available from the simulator. The strong correlations and pronounced differences observed between different experience levels for guidewire tip velocity-based metrics further supports their validity as reliable and robust indicators of endovascular performance. Adopting such metrics has the potential to improve the utility of the FVES module and similar modules for objective assessment of endovascular surgical skill. In contrast, most of the motion-based metrics available from the simulator had small effect sizes and did not exhibit significant differences between experience levels, with the exception of the catheter travel metric, which had data integrity issues. The strong linear correlations between average velocity and idle time and spectral arc length, our gold standard for objective assessment of endovascular surgical skill, indicate that these metrics could serve as intuitive measures of task performance. Such measures have the potential to be easier to interpret when provided as performance feedback than spectral arc length. We believe that these findings are a promising step towards automating performance evaluation for endovascular navigation tasks using tool tip motion data.

ACKNOWLEDGMENT

The authors thank Wyatt Crider for his valuable support with data collection and analysis. We also thank Simbionix for their support with hardware and software during the professional meetings, and the staff at MITIE for their accommodations.

REFERENCES

[1] R. Ahmadi, A. Willfort, M. M. S. Wilfried Lang, E. Alt, M. E. Gschwandtner, M. Haumer, T. Maca, H. Ehringer, , and E. Minar, "Carotid artery stenting: Effect of learning curve and intermediate-term morphological outcome," *Journal of Endovascular Therapy*, vol. 8, no. 6, pp. 539–546, 2001.

[2] M. Prinssen, E. L. Verhoeven, J. Buth, P. W. Cuypers, M. R. van Sambeek, R. Balm, E. Buskens, D. E. Grobbee, and J. D. Blankensteijn, "A randomized trial comparing conventional and endovascular repair of abdominal aortic aneurysms," *The New England Journal of Medicine*, vol. 351, no. 16, pp. 1607–1618, 2004.

[3] C. R. Smith, M. B. Leon, M. J. Mack, D. C. Miller, J. W. Moses, L. G. Svensson, E. M. Tuzcu, J. G. Webb, G. P. Fontana, R. R. Makkar, M. Williams, T. Dewey, S. Kapadia, V. Babaliaros, V. H. Thourani, P. Corso, A. D. Pichard, J. E. Bavaria, H. C. Herrmann, J. J. Akin, W. N. Anderson, D. Wang, and S. J. Pocock, "Transcatheter versus surgical aortic-valve replacement in high-risk patients," *The New England Journal of Medicine*, vol. 364, no. 23, pp. 2187–2198, 2011.

[4] P. Lurz, L. Coats, S. Khambadkone, J. Nordmeyer, Y. Boudjemline, S. Schievano, V. Muthurangu, T. Y. Lee, G. Parenzan, G. Derrick, S. Cullen, F. Walker, V. Tsang, J. Deanfield, A. M. Taylor, and P. Bonhoeffer, "Percutaneous pulmonary valve implantation: Impact of evolving technology and learning curve on clinical outcome," *Circulation*, vol. 117, no. 15, pp. 1964–1972, 2008.

[5] C. S. Rihal, R. A. Nishimura, and D. R. Holmes, "Percutaneous balloon mitral valvuloplasty: The learning curve," *American Heart Journal*, vol. 122, no. 6, pp. 1750–1756, 1991.

[6] A. Khoynezhad and M. Kruse, "Guidewires, catheters, and sheaths used for thoracic endografting procedures," *Journal of cardiac surgery*, vol. 24, pp. 113–119, 2008.

[7] J. Bismuth, M. A. Donovan, M. K. O'Malley, H. F. El Sayed, J. J. Naoum, E. K. Peden, M. G. Davies, and A. B. Lumsden, "Incorporating simulation in vascular surgery education," *Journal of vascular surgery*, vol. 52, no. 4, pp. 1072–1080, 2010.

[8] R. Reznick, G. Regehr, H. MacRae, J. Martin, and W. McCulloch, "Testing technical skill via an innovative "bench station" examination," *The American Journal of Surgery*, vol. 173, no. 3, pp. 226–230, Mar. 1997. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0002961097895979

[9] M. J. Jackson, V. Pandey, and J. H. N. Wolfe, "Training for Infrainguinal Bypass Surgery," *European Journal of Vascular and Endovascular Surgery*, vol. 26, no. 5, pp. 457–466, Nov. 2003. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1078588403003174

[10] M. C. Vassiliou, L. S. Feldman, C. G. Andrew, S. Bergman, K. Leffondré, D. Stanbridge, and G. M. Fried, "A global assessment tool for evaluation of intraoperative laparoscopic skills," *The American Journal of Surgery*, vol. 190, no. 1, pp. 107–113, Jul. 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0002961005003478

[11] C. Duran, S. Estrada, M. O'Malley, M. Sheahan, M. Shames, J. T. Lee, and J. Bismuth, "The model for fundamentals of endovascular surgery (fevs) successfully defines the competent endovascular surgeon," *Journal of Vascular Surgery*, vol. 62, no. 6, pp. 1660–1666, 2015.

[12] K. Moorthy, Y. Munz, S. K. Sarker, and A. Darzi, "Objective assessment of technical skills in surgery," *BMJ*, vol. 327, no. 7422, pp. 1032–1037, Oct. 2003. [Online]. Available: https://www.bmj.com/content/327/7422/1032

[13] A. Schwein, B. Kramer, P. Chinnadurai, S. Walker, M. O'Malley, A. Lumsden, and J. Bismuth, "Flexible robotics with electromagnetic tracking improve safety and efficiency during in vitro endovascular navigation," *Journal of Vascular Surgery*, vol. 63, no. 1, pp. 285–286, 2016.

[14] S. Estrada, C. Duran, D. Schulz, J. Bismuth, M. Byrne, and M. O'Malley, "Smoothness of surgical tool tip motion correlates to skill in endovascular tasks," *IEEE Transactions on Human Machine Systems*, vol. 46, no. 5, pp. 647–659, 2016.

[15] 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," *The Journal of Urology*, vol. 188, no. 3, pp. 919–923, Sep. 2012.

[16] 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, Jul. 2014. [Online]. Available: https://doi.org/10.1007/s00464-014-3446-5

[17] S. Krishnan, A. Garg, S. Patil, C. Lea, G. Hager, P. Abbeel, and K. Goldberg, "Transition State Clustering: Unsupervised Surgical Trajectory Segmentation For Robot Learning," *International Journal of Robotics Research*, pp. 1–15, 2016.

[18] H. Rafii-Tari, C. J. Payne, J. Liu, C. Riga, C. Bicknell, and G. Yang, "Towards automated surgical skill evaluation of endovascular catheterization tasks based on force and motion signatures," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, May 2015, pp. 1789–1794.

[19] M. K. O'Malley, M. D. Byrne, S. Estrada, C. Duran, D. Schulz, and J. Bismuth, "Expert surgeons can smoothly control robotic tools with a discrete control interface," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 4, p. 388–394, 2019.

[20] W. H. Jantscher, "Using real-time smoothness metrics to deliver haptic performance cues for a dexterous task," Master's thesis, Rice University, 2018.

[21] *Toward improved surgical training: Delivering smoothness feedback using haptic cues*. San Francisco, CA: IEEE, 03/2018 2018.

[22] S. Balasubramanian, A. Melendez-Calderon, A. Roby-Brami, and E. Burdet, "On the analysis of movement smoothness," *Journal of NeuroEngineering and Rehabilitation*, vol. 12, no. 1, p. 112, Dec. 2015. [Online]. Available: https://doi.org/10.1186/s12984-015-0090-9

[23] A. C. Lobato, J. Rodríguez-López, and E. B. Diethrich, "Learning curve for endovascular abdominal aortic aneurysm repair: evaluation of a 277-patient single-center experience." *Journal of endovascular therapy : an official journal of the International Society of Endovascular Specialists*, vol. 9, no. 3, pp. 262–268, 2002.

[24] E. B. Mazomenos, F. Vasconcelos, J. Smelt, H. Prescott, M. Jahangiri, B. Martin, A. Smith, S. Wright, and D. Stoyanov, "Motion-Based Technical Skills Assessment in Transoesophageal Echocardiography," in *Medical Imaging and Augmented Reality*, G. Zheng, H. Liao, P. Jannin, P. Cattin, and S.-L. Lee, Eds. Cham: Springer International Publishing, 2016, vol. 9805, pp. 96–103.

[25] A.-L. D. D'Angelo, D. N. Rutherford, R. D. Ray, S. Laufer, C. Kwan, E. R. Cohen, A. Mason, and C. M. Pugh, "Idle time: an underdeveloped performance metric for assessing surgical skill," *The American Journal of Surgery*, vol. 209, no. 4, pp. 645–651, Apr. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0002961015000045

[26] R. Aggarwal, T. P. Grantcharov, J. R. Eriksen, D. Blirup, V. B. Kristiansen, P. Funch-Jensen, and A. Darzi, "An Evidence-Based Virtual Reality Training Program for Novice Laparoscopic Surgeons," *Annals of Surgery*, vol. 244, no. 2, pp. 310–314, Aug. 2006. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1602164/

[27] A. E. Rolls, C. V. Riga, C. D. Bicknell, D. V. Stoyanov, C. V. Shah, I. Van Herzeele, M. Hamady, and N. J. Cheshire, "A Pilot Study of Video-motion Analysis in Endovascular Surgery: Development of Real-time Discriminatory Skill Metrics," *European Journal of Vascular and Endovascular Surgery*, vol. 45, no. 5, pp. 509–515, May 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1078588413001214

[28] *On the development of objective metrics for surgical skills evaluation based on tool motion*. IEEE, 2014.

[29] R. G. O'Brien, "A simple test for variance effects in experimental designs." *Psychological Bulletin*, vol. 89, no. 3, p. 570, 1981. [Online]. Available: https://psycnet.apa.org/fulltext/1981-20271-001.pdf