Barathwaj Murali[1]
Department of Mechanical Engineering,
Rice University,
Houston, TX 77005
e-mail: bm43@rice.edu

Viony M. Belvroy
DeBakey Heart & Vascular Center,
Houston Methodist Hospital,
Houston, TX 77030
e-mail: vionybelvroy@hotmail.com

Shivam Pandey
Department of Psychological Sciences,
Rice University,
Houston, TX 77005
e-mail: sp45@rice.edu

Jean Bismuth
DeBakey Heart & Vascular Center,
Houston Methodist Hospital,
Houston, TX 77030
e-mail: jbismuth@houstonmethodist.org

Michael D. Byrne
Department of Psychological Sciences,
Rice University,
Houston, TX 77005
e-mail: byrne@rice.edu

Marcia K. O'Malley
Department of Mechanical Engineering,
Rice University,
Houston, TX 77005
e-mail: omalleym@rice.edu

Velocity-Domain Motion Quality Measures for Surgical Performance Evaluation and Feedback

Endovascular navigation proficiency requires a significant amount of manual dexterity from surgeons. Objective performance measures derived from endovascular tool tip kinematics have been shown to correlate with expertise; however, such metrics have not yet been used during training as a basis for real-time performance feedback. This paper evaluates a set of velocity-based performance measures derived from guidewire motion to determine their suitability for online performance evaluation and feedback. We evaluated the endovascular navigation skill of 75 participants using three metrics (spectral arc length, average velocity, and idle time) as they steered tools to anatomical targets using a virtual reality simulator. First, we examined the effect of navigation task and experience level on performance and found that novice performance was significantly different from intermediate and expert performance. Then we computed correlations between measures calculated online and spectral arc length, our "gold standard" metric, calculated offline (at the end of the trial, using data from the entire trial). Our results suggest that average velocity and idle time calculated online are strongly and consistently correlated with spectral arc length computed offline, which was not the case when comparing spectral arc length computed online and offline. Average velocity and idle time, both time-domain based performance measures, are therefore more suitable measures than spectral arc length, a frequency-domain based metric, to use as the basis of online performance feedback. Future work is needed to determine how to best provide real-time performance feedback to endovascular surgery trainees based on these metrics.
[DOI: 10.1115/1.4049310]

1 Introduction

Minimally invasive endovascular procedures are increasingly becoming the intervention of choice, given their numerous postoperative and functional advantages over open surgery. These procedures are especially preferable for individuals who are either ineligible for open surgery or who face high risk due to comorbidities or advanced age [1–3]. Endovascular procedures cover a wide range of diagnostic and therapeutic interventions, such as aortic valve placement, carotid artery stenting, and aortic aneurysm repair [2,4–7], and can result in significantly shorter operation times and hospital stays, lower complication rates, less blood loss, and lower rates of postoperative mechanical ventilation and atrial fibrillation than the equivalent open procedures [3,8].

Surgeons perform endovascular procedures by navigating flexible tools inserted into the body at a small incision. These procedures generally involve guidewires to provide and maintain access to anatomical structures, such as the site of an aneurysm. Surgeons can then introduce catheters to visualize anatomical structures using radiographic contrast and deploy devices such as endografts or replacement heart valves [4,8]. Depending on the procedure, surgeons may navigate these tools within a sheath to further improve vessel access and reduce the risk of vascular injury [4]. The flexible nature of these tools gives rise to complex interactions within the vascular environment characterized by a nonlinear mapping between proximal motions made by the surgeon and resultant tool tip motions. These elements of endovascular navigation require surgeons to undergo a substantial amount of training to acquire proficiency.

Ensuring surgeons' proficiency in endovascular navigation is central to improving postoperative outcomes. Repeated practice is necessary for skill acquisition, and minimally invasive procedures like endovascular surgery may require more or specialized practice. For example, despite observing lower amounts of blood loss and atrial fibrillation during endovascular aortic valve replacement compared to traditional surgical methods, Smith et al. attributed observations of a higher rate of stroke, transient ischemic attacks, and major vascular complications to a protracted learning curve [3]. Similar observations in valvuloplasty and pulmonary valve replacement indicate a relationship between experience level and lower procedure times, as well as a longer time before reoperation is necessary [9,10]. Such findings highlight the need for surgical training that incorporate objective assessment and performance feedback.

Traditional approaches for assessing surgical performance consist of checklists and global rating scales such as the objective structured assessment of technical skill and the Imperial College evaluation of procedural skill for measuring general and procedural skills [11,12]. For endovascular performance assessment, the recommended tool for assessing procedural efficiency and autonomy, fluoroscopic imaging and contrast use, device deployment, and tool manipulation is the global rating assessment device for endovascular skill [13]. These approaches remain subjective and retrospective in nature and require time and resources from several examiners, usually senior-level attending physicians [14].

[1]Corresponding author.
Manuscript received July 30, 2020; final manuscript received November 22, 2020; published online January 28, 2021. Assoc. Editor: Nandini Duraiswamy.

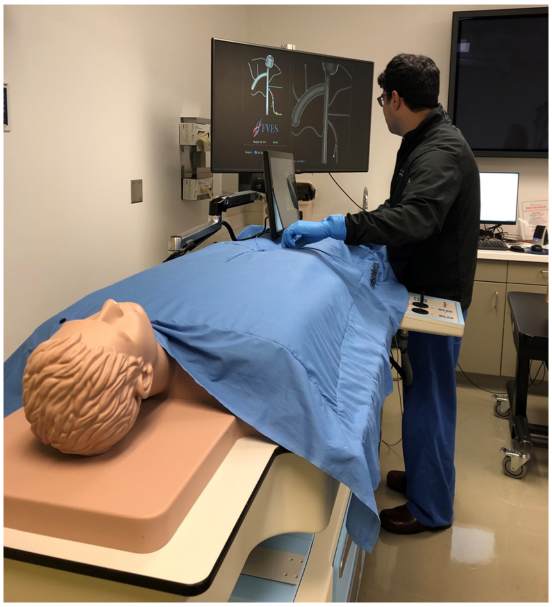

Fig. 1 Participant performing endovascular navigation tasks on an ANGIO Mentor simulator at the MITIE

To address these challenges, there is a growing body of research exploring the use of tool motion data from instrumented or simulated surgical tools as the basis for *quantitative and objective* performance assessment [15–17]. 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 [18–20]. Virtual reality simulators, such as the one shown in Fig. 1, provide a means to acquire tool motion data and test methods for real-time performance feedback.

While there has been a considerable effort along these lines for laparoscopic and robotic surgery, application to the endovascular domain remains underdeveloped. Previous studies have 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 [16,21]. In these works, data analysis was performed after collecting data from the entire navigation task, as challenges in extracting tool tip motions from the various platforms and substantial data processing efforts precluded real-time or near real-time analysis.

1.1 Motivation for Online Performance Feedback. Training and practice can improve manual dexterity, and the provision of performance feedback can improve training outcomes [22]. Most endovascular training uses offline feedback, or feedback that is provided after the conclusion of a surgical task or procedure. For example, the results of assessment with global rating scales are often presented after the completion of navigation tasks. Online performance feedback, defined as feedback that is provided to trainees *as they train*, either in real-time or in near real-time, remains underexplored. In our prior work, we surveyed trainees performing a set of endovascular navigation tasks on a commercial surgical simulator and found that they were interested in receiving both offline and online performance feedback [23]. Over 50% of novice participants indicated their preference for receiving online feedback as a feature included in future iterations of the system. Intermediate and expert participants 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 indicated a preference for some form of feedback over none, while experts preferring either online feedback or no feedback [23].

Online performance evaluation techniques are being tested in other surgical disciplines such as assessment of bone surgery training using virtual reality, real-time assessment of tissue trauma during laparoscopic surgery, and for robotic surgery [22,24–26]. To date, automatic performance evaluation for endovascular techniques has used hidden Markov models as a basis for online evaluation and haptic guidance using proximal guidewire and tool tip motions [20,27].

An alternative technique to using generative models for training and evaluation is to use global performance measures known to be correlated with surgical expertise as a basis for performance feedback. In Jantscher [28], a frequency-domain movement smoothness measure known as spectral arc length (or SPARC [29]) was calculated online and provided to trainees during a mirror-tracing task that emulates endovascular navigation. Performance was evaluated by calculating SPARC for short time intervals throughout the task and intermittently providing a vibrotactile cue to the trainee that corresponded to their level of task performance (good, fair, or poor). Individuals who received this feedback adapted their task performance strategy in beneficial ways, reducing their task completion time while improving their accuracy in tracing the complex shape using the joystick device. Still, it was noted that trainees faced difficulties in interpreting and understanding the movement smoothness-based performance feedback that was provided.

Endovascular navigation occurs in an environment constrained by vessel walls that allow surgeons to generate different motion trajectories that are equally suitable for achieving successful vessel cannulation. This task is fundamentally different than the mirror-tracing task used by Jantscher [28] that resulted in all participants generating the same motion trajectory. It is unclear if the methods for real-time performance assessment and feedback that were successful in the case of the mirror-tracing task can be extended to a less-constrained and longer duration task like endovascular surgical navigation.

1.2 Contributions. Online performance assessment and feedback based on tool movement smoothness have the potential to positively impact endovascular surgical navigation training. In our previous findings, described in detail by Murali et al. [23] and Belvroy et al. [30], we determined that average velocity and idle time (the proportion of the motion profile consisting of idle tool movements [31]) of guidewire motion showed significant differences between experience level. These time-domain metrics also exhibited high linear correlations with SPARC and serve as indirect measures of frequency-domain movement smoothness [23]. However, we did not account for potential differences across our four different navigation tasks, and our analysis was based only on offline (end of trial) computation of metrics for a small number of participants.

This paper makes two primary contributions. First, we demonstrate the effect of task and experience level on two time-domain-based performance measures (average velocity and idle time) and one frequency-domain-based performance measure, SPARC, a variant of the frequency-domain measure of movement smoothness used for offline performance assessment in Estrada et al. [16,21] and online in Jantscher [28]. We show that there is a significant effect of task and experience level for all three metrics calculated offline. The second major contribution of this paper is an examination of the suitability of online estimation methods for each of these metrics since we are interested in providing real-time performance feedback to endovascular trainees. We explore different methods of online estimation by varying time-domain window length and type, and examine correlations between online estimates of each metric to the “gold standard” measure, SPARC, calculated offline.

2 Method

Participants completed a series of endovascular navigation tasks using the ANGIO Mentor system. We computed three performance measures from the velocity profile of guidewire motion (SPARC, average velocity, and idle time) collected during four different navigation tasks. We then applied a linear mixed effects model to verify the validity of each metric as offline measures of performance. We first tested to see if these metrics reflected differences in experience level among our participants, and then whether they reflected differences among the navigation tasks. Next, we examined how well each metric correlated as an online measure of performance with our gold standard metric, offline SPARC. We calculated each metric online by segmenting the tangential velocity profile of the guidewire tip during each navigation task over time using a series of discrete, sliding, and overlapping windows. The series of values calculated during these windows were then averaged and compared to SPARC calculated over the entire velocity profile.

2.1 Participants. Participants 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 75 individuals, (57 male, 18 female, 31 novices, 25 intermediates, and 19 experts) participated in our study.

The number of endovascular procedures performed with and without supervision determined the experience level of each participant, as we have done in our prior work [16,23,30]. Novices were defined as individuals who had performed less than 50 cases; intermediates were defined as individuals who had performed between 50 and 500 cases; and experts were defined as individuals who had performed over 500 cases. 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 [7]. 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 [5].

The novice group consisted of 19 students, seven residents, four fellows, and one industry professional. The intermediate group consisted of 11 residents, 10 fellows, and four attendings. The expert group consisted of six residents, 11 attendings, one fellow, and one physician assistant with experience in vascular surgery. Participants represented a wide range of medical specialties, including anesthesiology, 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 (IRB-FY2019-302, Houston, TX).

2.2 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 (see Fig. 1). The preloaded module containing the virtualized training model used by the fundamentals of endovascular and vascular surgery (FEVS) platform [13] was loaded on the simulator. Various tool geometries and interactions within the virtual environment were simulated using the tool motions recorded at the input by the ANGIO Mentor. The module streamed kinematic data of each tool tip over a TCP network connection at varying sampling rates between 15 and 60 Hz, which was used to compute each performance measure from the tangential velocity profile, as in Fig. 2.

2.3 Procedure. After consenting to participate in the study and prior to starting the first task, participants 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 min rotations during which they completed between one and four target navigation tasks depending on the time available with each participant. Participants recruited from MITIE in Houston completed all four navigation tasks.

The FEVS module consists of a set of eight possible tasks that test various procedural and dexterous tool manipulation abilities. Tasks required either direct navigation to targets or navigation to targets followed by additional procedural steps such as exchanging and introducing various catheters and sheaths. We analyzed performance for only those tasks described in Table 1, since these tasks required participants to directly navigate to targets without any additional tool exchange or tool introduction elements. This set of tasks, illustrated in Fig. 3, consisted of navigating a guidewire and catheter over a right-angle bifurcation, navigating a guidewire, catheter, and sheath into a third-order vessel with posterior takeoff, cannulating a branch vessel extending from an aneurysm, and performing gate cannulation through an aneurysmal segment [13].

Each task required the user to navigate a guidewire, catheter, and if present, a sheath to the color-coded targets shown in Fig. 3. The simulator computer screen displayed basic navigation guidelines before each task, and participants were given approximately 1–2 min to familiarize themselves with this information for

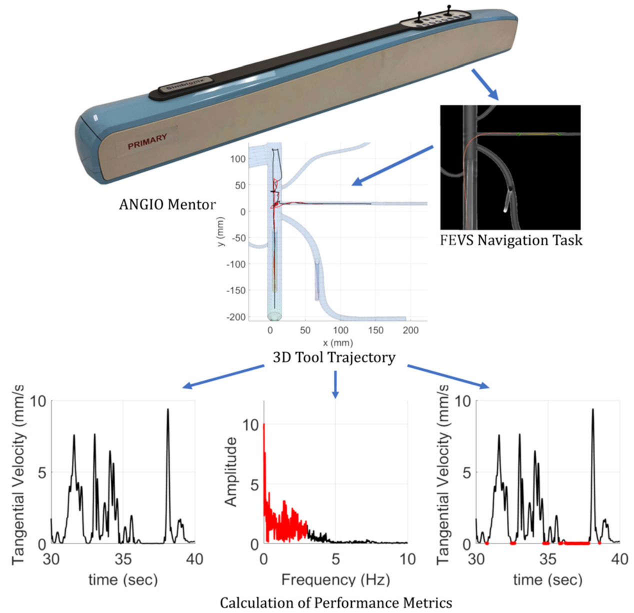

Fig. 2 Process of collecting tool tip kinematics and computing performance metrics from tangential velocity profile. ANGIO Mentor provides tool tip data of entire motion trajectory for a given navigation task.

Table 1 Selected tasks from the FEVS module from Ref. [13] that test endovascular navigation performance without tool exchange or additional procedural steps

Task	Description
1	Navigate up and over a bifurcation
3	Navigate into a third order vessel with posterior takeoff
5	Cannulate a branch vessel extending from an aneurysm
7	Gate cannulation

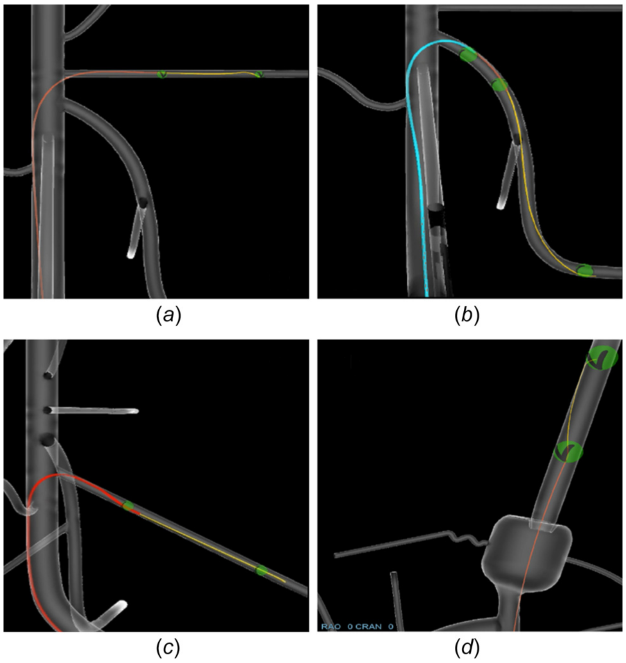

Fig. 3 Four navigation tasks with targets shown as shaded circles: (*a*) right angle bifurcation, (*b*) cannulation of third-order branch vessel, (*c*) cannulation of aneurysmal branch vessel, and (*d*) gate cannulation through aneurysmal segment

navigating each tool to its respective target. The ANGIO Mentor provides some limited haptic feedback to simulate physical interactions between tools and the virtual environment. No other haptic feedback was provided during the execution of the navigation tasks. After participants were ready to proceed with assessment, they performed the navigation task until either successfully reaching the targets or until the simulation timed out (at between 3 and 5 min, depending on task). Almost all participants completed the right angle bifurcation task (task 1, see Fig. 3), and then (time permitting) proceeded to complete additional tasks (tasks 3, 5, 7, see Table 1) [13]. The time limit for task 3 was 5 min while the other tasks had a time limit of 3 min. Most novice and intermediate participants performed one or two tasks while most experts performed two to four tasks. Tasks were not repeated and each session did not exceed 15 min.

After finishing their final navigation task, participants completed an additional custom questionnaire that gathered 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 participants preferred receiving feedback, either during or after each task.

2.4 Data Preprocessing. 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. A total of 199 individual motion trials were generated from participants performing the endovascular navigation tasks, 173 of which were free of critical tool manipulation errors and were used for calculating each performance measure for the online and offline analyses. Removal of trials containing critical failures resulted in data from seven subjects (all novices) being removed from analysis. Outlier removal was not performed.

The FEVS module provided time-series data of X, Y, Z position and the change in position values for each tool present in the task environment. The data provided the full trajectory of each tool from the beginning of the task (illustrated in Fig. 2) until either successful task completion or timeout.

The change in X, Y, and Z positions for each tool was scaled by the time interval between samples to calculate velocity. After this conversion, the data were transformed to a constant sampling frequency of 60 Hz using linear interpolation for frequency analysis and calculation of SPARC. A third-order Savitzky–Golay filter with a window length of 21 samples was then implemented to remove high frequency noise from the tool tip data while preserving the waveform shape of each signal [16]. From the interpolated and filtered data, the tangential velocity profile was calculated, from which each candidate performance measure was calculated.

2.5 Performance Measures

2.5.1 Spectral Arc Length. From the tool tip velocity data, we computed spectral arc length (SPARC), a measure of movement smoothness, that was previously shown to be significantly correlated to experience level for endovascular procedures performed on manual, simulation, and robotic platforms [16,21]. 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 [29]. Additionally, the relatively low computational burden of calculating SPARC shows promise for both online and offline performance evaluation and feedback [32]. SPARC is calculated using Eq. (1)

$$\mathrm{SPARC} = -\int_0^{\omega_c} \left[\left(\frac{1}{\omega_c} \right)^2 + \left(\frac{d\hat{V}(\omega)}{d\omega} \right)^2 \right]^{\frac{1}{2}} d\omega \qquad (1)$$

where $V(\omega)$ is the Fourier magnitude spectrum of the velocity profile $v(t)$ given by the FFT operation. The magnitude spectrum is normalized with its DC magnitude $V(0)$, expressed as $\hat{V}(\omega)$. The cutoff frequency ω_c is determined by an amplitude threshold, which ensures that the metric produces values independent of temporal scaling (i.e., movement profiles of different duration but same shape) [29].

2.5.2 Average Velocity. Average tool-tip velocity was calculated for each tool using the tangential velocity profile. Average tool velocity is another promising metric given its significant correlation to experience level in other catheter-based surgical domains such as transesophageal echocardiography [33]. Average velocity for each tool is calculated by taking the discrete sum of tangential velocity values, given by their index i, which is then divided by the total number of discrete velocity values N, as in Eq. (2)

$$V_{\mathrm{avg}} = \frac{\sum_{i=1}^{N} V_i}{N} \qquad (2)$$

2.5.3 Idle Time. Idle time is defined as the ratio the time during a navigation task in which the surgical tools remain stationary to the total amount of time the tool was present in the task. Idle time was shown to correlate to experience level in open surgery [31], and similar to average velocity, this metric likely provides another measure of cognitive engagement, with higher values evident in individuals with less experience [31]. Idle time was calculated using Eq. (3)

$$T_{\mathrm{idle}} = \frac{\sum_{i=1}^{N} g(i)\Delta t}{t_{\mathrm{tool}}}, \quad g(i) = \begin{cases} 0 \; v(i) > v_0 \\ 1 \; v(i) < v_0 \end{cases} \qquad (3)$$

where the amount of time in which the tools moved at tangential velocities below a threshold value v_0 is provided by a discrete sum using the binary function $g(i)$ that compares the tangential velocity value $v(i)$ at index i with the threshold v_0. This value was multiplied by the sampling interval Δt before being divided by the time between the first instance of the tool entering the simulated environment and task completion t_{tool}. The value of v_0 was defined as 0.5 mm/s to account for motion artifacts such as deceleration of flexible tool tips and deformation of tool tips against the vessel well.

2.6 Data Analysis. After calculating the offline and online values of each performance metric for each navigation task performed across all participants, we performed our analysis in two parts. First, we used a linear mixed effects model to explore whether each offline metric showed significant differences across experience level and navigation task. We then evaluated the suitability of each metric estimated online by comparing the set of online metric values generated by each navigation task with the corresponding value of offline SPARC.

2.6.1 Effects of Expertise and Task. In our previous work, we performed a series of one-way analysis of variance (ANOVA) tests to show that SPARC, average velocity, and idle time are valid offline predictors of surgical expertise [23]. These results, summarized in Tables 2 and 3, did not consider the potential effects of navigation task and subject-specific variability on performance. To better account for these factors, in this paper, we applied a linear mixed effects model using the values of SPARC, average velocity, and idle time calculated from guidewire motion data. Experience level and task were modeled as fixed effects factors and participants were modeled as a random factor. Table 4 provides the between-subject and within-subject factors and their corresponding levels that formed the model. Degrees of freedom in the model were approximated using the Kenward–Roger method.

For each significant main effect, contrasts were applied to determine differences between experience levels and tasks, again using the Kenward–Roger method for determining significance levels. Contrasts compared the novice group with a combined group of intermediate and expert participants, as well as the expert group with the intermediate group. Contrasts for task compared task 1 (right-angle branch cannulation) with task 3 (third-order vessel with posterior takeoff), task 5 (aneurysmal branch cannulation), and task 7 (gate cannulation). We did not test for interactions between experience level and task.

Table 2 ANOVA results and effect sizes for performance metrics calculated from the tangential velocity profile of guidewire motion data

Metric	ANOVA test result	Effect size (Cohen's f)
SPARC	$F(2, 42) = 9.38$; $\boldsymbol{p < 0.001}$	0.67
Average velocity	$F(2, 42) = 10.66$; $\boldsymbol{p < 0.001}$	0.71
Idle time	$F(2, 42) = 8.18$; $\boldsymbol{p = 0.001}$	0.62

Statistically significant p-values (in bold) and large effect sizes for each metric highlight the strong association with experience level. Reproduced from Murali et al. [23].

Table 3 Pearson's r correlation coefficients and accompanying p values from linear regression tests performed in Murali et al. [23]

Metric	Pearson r and p-value
Average velocity	$r(42) = 0.72$; $\boldsymbol{p < 0.001}$
Idle time	$r(42) = 0.70$; $\boldsymbol{p < 0.001}$

Average velocity and idle time are strongly correlated with SPARC and can serve as indirect time-domain measures. Statistically significant values in bold.

Table 4 Summary of fixed and random factors and corresponding levels used by the linear mixed model

Factor	Type	Levels
Task	Fixed	1, 3, 5, 7
Experience level	Fixed	Novice, intermediate, expert
Subject	Random	1–68

2.6.2 Online and Offline Metric Comparison. SPARC, average velocity, and idle time show statistically significant differences across experience levels as offline measures of performance (see Table 2). In addition to serving as two candidate metrics from the time domain, average velocity and idle time may be more suitable for online performance evaluation and feedback, given their strong linear correlations with SPARC and their potential ease of interpretation compared to SPARC. To show the utility of each measure as an online indicator of performance, it is necessary to determine the relationship between each metric calculated online and experience level. A similar analysis to the ANOVAs performed by Murali et al. [23] can be performed by taking advantage of the high correlations between average velocity, idle time, and SPARC from Table 3. These linear correlations allow for the use of offline SPARC to establish the corresponding relationship to experience level.

This comparison can be performed indirectly by exploring the relationship between each online metric and offline SPARC since offline SPARC shows the strongest differences with experience level [23]. If a strong relationship exists between each online metric and offline SPARC, then by extension, the significant differences across experience level observed with the offline metrics in our prior work will likely be present with the online metrics. Linear regression was used to evaluate the online-offline relationship between each continuous valued metric, with higher Pearson correlation coefficients indicating a stronger online-offline relationship.

To compute online measures, we used a set of discrete and sliding windows of 5-15 s lengths. Given the average data sampling rate before interpolation of approximately 20 Hz, each 5 s window contained approximately 100 samples of data, resulting in a frequency resolution of approximately 0.2 Hz, which is likely sufficient for capturing movement frequencies expected for endovascular navigation motions. The upper bound of this range of window sizes was determined by the average completion time across all trials of 102.6 s, which would allow for at least six discrete windows for providing online feedback. The amount of overlap between the sliding windows was varied in 0.5 s increments.

Kinematic data acquired in real time from the virtualized FEVS tasks were used to calculate each candidate metric over a moving window of time, similar to our prior work [28]. A real-time performance feedback scenario can be emulated from the tool tip positions and velocities used for the offline analysis performed previously [23] by calculating each metric across a subset of the overall motion profile defined by a window of fixed time interval. These windows can be either nonoverlapping (discrete) or overlapping (sliding) with different amounts of overlap, as detailed in Fig. 4.

The values for SPARC, average velocity, and idle time from each window were averaged and compared with SPARC calculated across the entire trial. This comparison was performed using robust linear regression to minimize the effect of potential outlier data. These regression tests provided Pearson correlation coefficients that relate the average online values of each performance metric to the offline measure of SPARC shown to correlate with

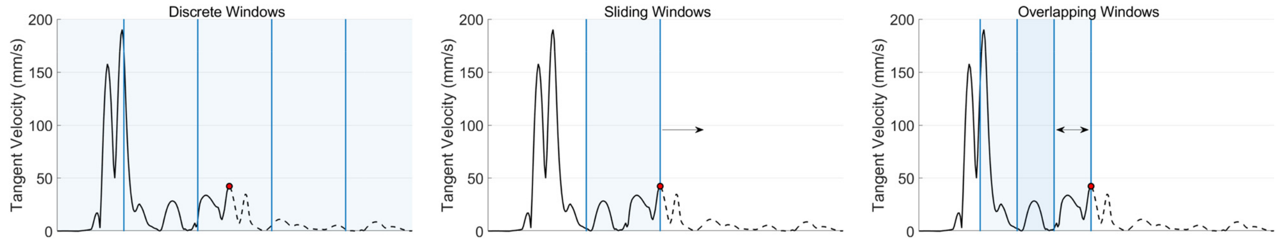

Fig. 4 Discrete, sliding, and overlapping data windows that were used for segmenting velocity profile and calculating online performance measures

experience level from prior studies [23,30,34]. The evaluation of the online performance of each metric was carried out using motion trials corresponding to a single task, given the task dependence of movement smoothness metrics [29]. That each task likely differs in difficulty may result in certain tasks being more amenable for online performance feedback than others.

2.6.3 Performance Thresholds for Feedback Cue Design. Implementation of an online performance feedback protocol similar to that presented in Jantscher [28] requires appropriate thresholds defining high, medium, and low quality of motion. One possible method of determining these thresholds is to use the averaged online values corresponding to the different experience levels used to define novice, intermediate, and expert performance. From the distribution of the online values for each experience level, namely, from the set of error bars using the mean and standard deviation, medium motion quality was defined as the region of values given by the intersection of the error bars of the novice and expert groups. High- and low-motion qualities were defined as the regions above and below this intersecting region, respectively.

Using the average online values of each metric, an appropriate basis for defining performance thresholds requires the assumption that the values calculated across each window of time do not deviate substantially from the average online value, likely due to a combination of both uniform task difficulty and participants using similar motion profiles for performing navigation. This assumption was verified by using the standard deviation of the online values calculated for each motion trial.

3 Results

Using kinematic data collected from a commercial surgical simulator while participants performed a set of endovascular navigation tasks, SPARC, average velocity, and idle time were calculated offline using motion data collected from subjects representing a wide range of surgical experience levels and medical specialties. A linear mixed effects model was used to verify significant group differences across experience level observed in our prior works, and to quantitatively determine the task dependence of movement smoothness metrics and their indirect measures. Then, we evaluated SPARC, average velocity, and idle time of guidewire motion calculated over discrete and sliding windows of varying duration and overlap amount with SPARC calculated offline using the entire tangential velocity profile.

3.1 Differences Between Experience Levels. Significant main effects of experience level were observed for SPARC ($F(2, 63) = 13.87$, $p < 0.001$), average velocity ($F(2, 62) = 8.88$, $p < 0.001$), and idle time ($F(2, 63) = 15.12$, $p < 0.001$) of the guidewire motion, as shown in Fig. 5. From the contrasts, the SPARC values associated with the combined group of intermediates and experts are significantly lower than those of novices ($t(83) = 4.27$; $p < 0.001$). The same trend was present for idle time ($t(88) = 4.43$; $p < 0.001$) and the inverse was true for average velocity ($t(89) = 3.19$; $p = 0.002$). The expert group alone possessed significantly lower SPARC values than intermediates ($t(50) = 3.30$; $p = 0.002$), as well as higher average velocities ($t(46) = 2.89$; $p = 0.006$) and lower idle times ($t(47) = 3.46$; $p = 0.001$).

3.2 Differences Between Tasks. There exists a significant effect of task on performance for SPARC ($F(3, 124) = 13.36$; $p < 0.001$), average velocity ($F(3, 129) = 17.80$; $p < 0.001$), and idle time ($F(3, 127) = 12.25$; $p < 0.001$), illustrated in Fig. 6. The contrasts showed significant differences between tasks 1 and 3 for idle time only ($t(128) = 2.71$; $p = 0.008$). Tasks 1 and 5 showed significant differences for average velocity ($t(128) = 5.02$; $p < 0.001$) and idle time ($t(128) = 3.40$; $p = 0.001$). Significant differences for SPARC ($t(125) = 5.05$; $p < 0.001$), average velocity ($t(131) = 5.49$; $p < 0.001$), and idle time ($t(129) = 2.00$; $p = 0.047$) were present between tasks 1 and 7.

3.3 Correlations Between Online and Offline Metrics. SPARC calculated online resulted in low correlations with SPARC calculated offline using discrete, sliding, and over-lapping windows, with correlation coefficient values ranging between approximately 0.1 and 0.5. A maximum correlation coefficient of approximately 0.6 between online and offline SPARC was observed at an approximately 1–2 s window length using sliding windows.

Discrete windows resulted in a noisy and nonmonotonic trend between correlation coefficients and increasing window size, with no one task producing consistently higher or lower values than the others. Sliding windows resulted in a smoother but nonmonotonic relationship between correlation coefficient values and window size, with values for each task increasing from a minimum value of approximately 0.1 after a window size of approximately 6 s. Correlation coefficients between online and offline SPARC using overlapping windows produced trends similar in shape, magnitude, and noise to those observed using discrete windows. The correlation coefficients between online and offline SPARC for task 3 remained consistently low for discrete, sliding, and overlapping windows.

Average velocity and idle time calculated online produced higher and more consistent correlations with offline SPARC for discrete, sliding, and overlapping windows. Correlation coefficients for online average velocity and idle time remained between approximately 0.6 and 0.8 for each window type, with online average velocity showing an almost constant behavior with increasing window length. Task 1 exhibited the highest correlation coefficient values for online average velocity using discrete windows, while tasks 1, 3, and 7 had comparable values across all window lengths for sliding and overlapping windows. Task 5 produced the lowest correlation coefficients between averaged online average velocity and offline SPARC for discrete, sliding, and overlapping windows.

Correlation coefficients for idle time calculated online versus SPARC offline were comparable in value to those for average velocity calculated online versus SPARC offline for all window types. The lowest correlation values for each window type were observed for data from task 3 while the highest values were observed for data from task 1. Calculating online idle time using sliding windows resulted in a slightly monotonically increasing trend between correlation coefficient value and window size. Tasks 1 and 7 had the highest correlation values using sliding windows, followed by tasks 3 and 5. As with online average velocity,

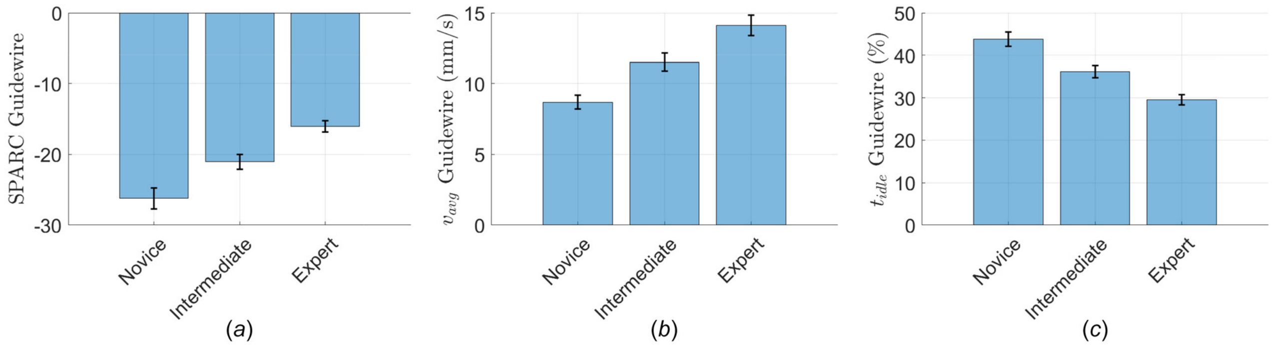

Fig. 5 Bar graphs showing mean of SPARC, average velocity, and idle time calculated from guidewire motion for each experience level. Error bars provide standard error of the mean for each experience level.

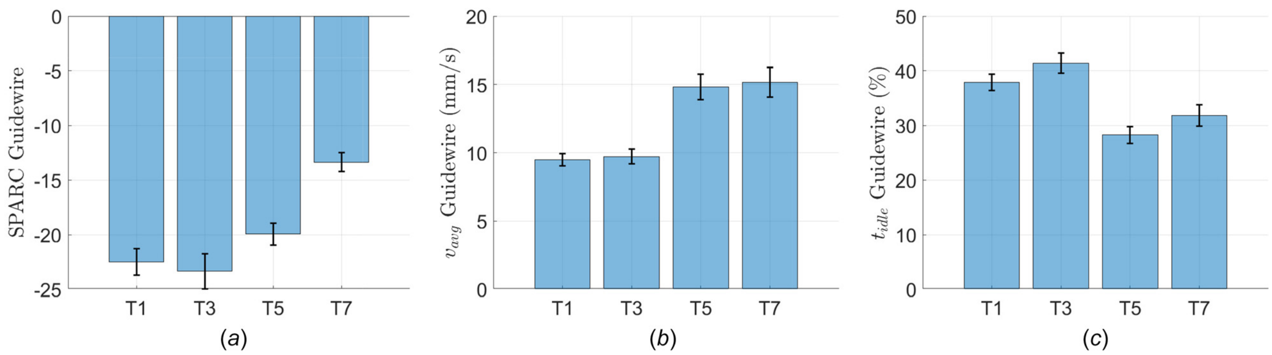

Fig. 6 Bar graphs showing mean of SPARC, average velocity, and idle time calculated from guidewire motion for each navigation task. Error bars provide standard error of the mean for each navigation task.

Table 5 Results from linear mixed effects model contrasts

Contrast	SPARC	Average velocity (mm/s)	Idle time (%)
Novice versus (intermediate, expert)	$t(83) = 4.27$; $\boldsymbol{p < 0.001}$	$t(89) = 3.19$; $\boldsymbol{p = 0.002}$	$t(88) = 4.43$; $\boldsymbol{p < 0.001}$
Intermediate versus Expert	$t(50) = 3.30$; $\boldsymbol{p = 0.002}$	$t(46) = 2.89$; $\boldsymbol{p = 0.006}$	$t(47) = 3.46$; $\boldsymbol{p = 0.001}$
Task 1 versus task 3	$t(125) = 1.29$; $p = 0.20$	$t(129) = 0.16$; $p = 0.87$	$t(128) = 2.71$; $\boldsymbol{p = 0.008}$
Task 1 versus task 5	$t(124) = 0.24$; $p = 0.81$	$t(128) = 5.02$; $\boldsymbol{p < 0.001}$	$t(128) = 3.40$; $\boldsymbol{p = 0.001}$
Task 1 versus task 7	$t(125) = 5.05$; $\boldsymbol{p < 0.001}$	$t(131) = 5.49$; $\boldsymbol{p < 0.001}$	$t(129) = 2.00$; $\boldsymbol{p = 0.047}$

Statistically significant values in bold.

sliding and overlapping windows produced similar behaviors and values with different window lengths and overlap amounts.

4 Discussion

From our offline evaluations of performance, it is clear that each metric shows significant differences across experience level. Differences observed across navigation tasks highlight the task dependent nature of movement smoothness-based performance metrics. These differences may also imply a learning effect and a possible increase in difficulty between subsequent tasks. While SPARC is effective as an offline measure of performance, average velocity and idle time prove to be more effective as online measures, given their higher online–offline correlation coefficients. From the distributions of each online performance metric separated first by task and then by experience level, it is possible to determine performance thresholds that can be used as a basis for providing online performance feedback.

4.1 Assessing Performance Differences Offline: Effect of Expertise and Task. SPARC, average velocity, and idle time each showed significant group differences across different levels of expertise. The contrasts performed as part of the mixed effects model suggest that novices, determined by a caseload of less than 50 endovascular procedures, have significantly higher SPARC and idle time scores, and lower average velocities, than those of intermediate and expert groups combined. This finding is in line with prior studies that have shown the effect of surgical expertise on procedural success rates, complication rates, and completion times [5,7]. Importantly, our approach using direct tool motion data offers a quantitative, objective, and automatic method for assessing expertise.

The significant main effect of task on the values of each performance measure, along with the results of the task-based comparisons from Table 5, provides quantitative evidence that online performance assessment should be conducted on a task-by-task basis. Also, the task dependencies of each metric are verified. The significant differences in SPARC, average velocity, and idle time observed between task 1 and tasks 3, 5, and 7 suggests that a small learning effect may be present between the first and subsequent tasks. Since the majority of participants performed task 1 before proceeding to the other tasks, such a learning effect could reflect the familiarization period that most subjects experienced regarding the simulation environment, tool

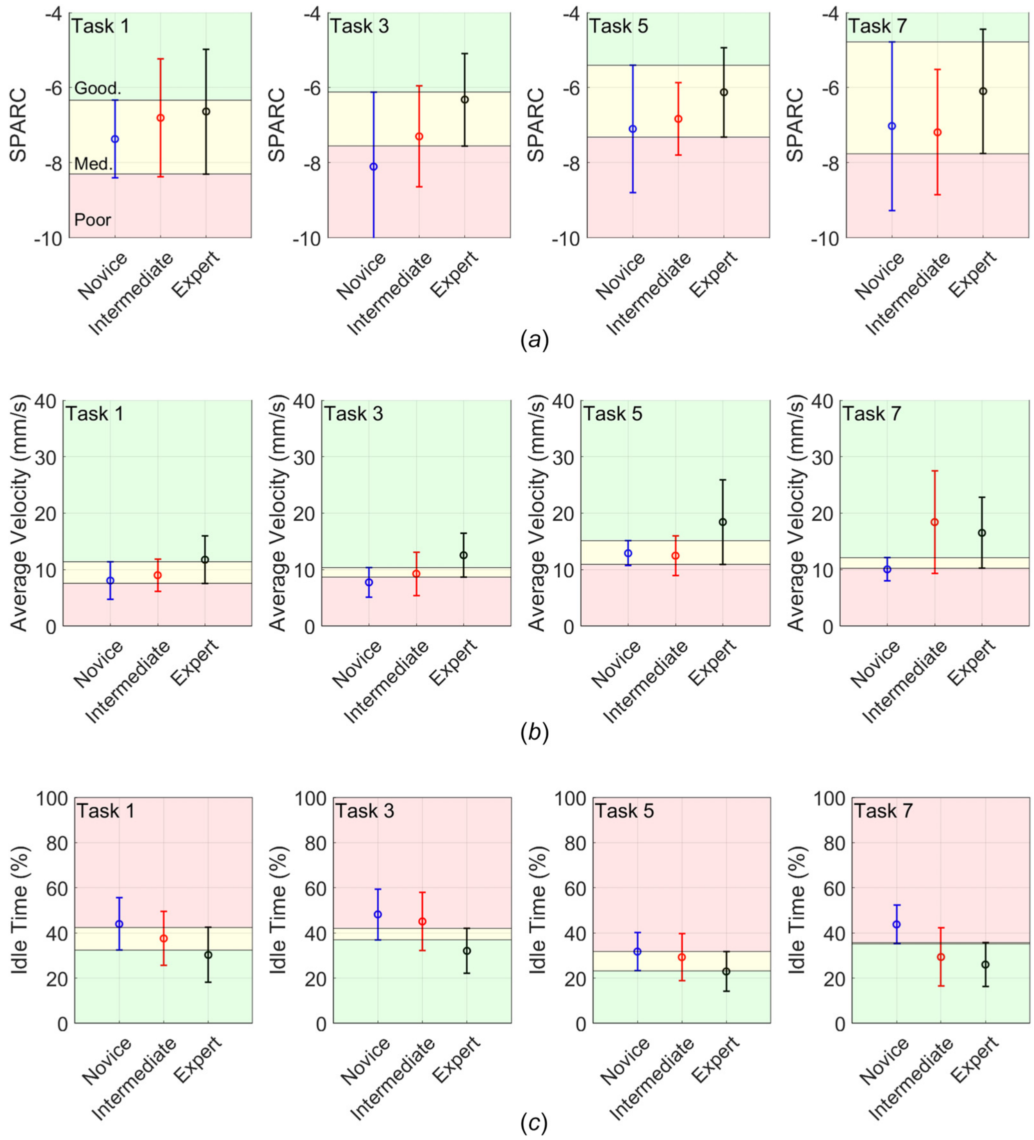

Fig. 7 Performance thresholds selected from averaged online values using 15 s window length with 5 s overlap for each navigation task. For SPARC and average velocity, top region represents good performance, while middle and bottom regions represent medium and poor performance, respectively. For idle time, bottom region represents good performance and top region represents poor performance.

interactions within the FEVS module, and the virtualized visualization controls. Alternatively, given the order of tasks in the FEVS module, it may be possible that tasks are ordered by difficulty level. In our previous work, we observed a linearly increasing trend in SPARC and other movement smoothness metrics across three sessions of performance assessment that may suggest the presence of a learning with these navigation tasks [16]; however, it is unclear whether this trend was due to learning the navigation task itself, or attributable to participants becoming more comfortable with the simulator interface.

4.2 Comparison of Online and Offline Calculation. Objective and quantitative performance metrics such as movement smoothness, average velocity, and idle time have significant potential for offline evaluation of endovascular surgical performance [23,30]. To investigate the utility of these metrics for *online* performance evaluation, the average value of each metric calculated during a navigation task using discrete and sliding windows of different duration and overlap was compared to SPARC calculated offline across the entire motion trial.

The results of the online-offline linear correlations between each metric and SPARC clearly indicate that SPARC, despite providing a robust offline measure of experience level and motion proficiency, is substantially less effective in capturing performance when calculated online. Given its frequency-domain computation and high task dependence [29], SPARC is most effective at comparing discrete motion profiles that employ similar task execution strategies and are of comparable difficulty. The motion profile generated by performing each target navigation task in the FEVS module consists of several smaller subprofiles resulting from the continuous insertion, retraction, and rotation of the guidewire and catheter throughout the task. Segmenting the overall velocity profile in the time-domain would either omit portions of a single subprofile or contain portions of adjacent subprofiles, which would increase the variability of the online SPARC values calculated throughout a navigation task and ultimately obscure any trends evident between online and offline SPARC.

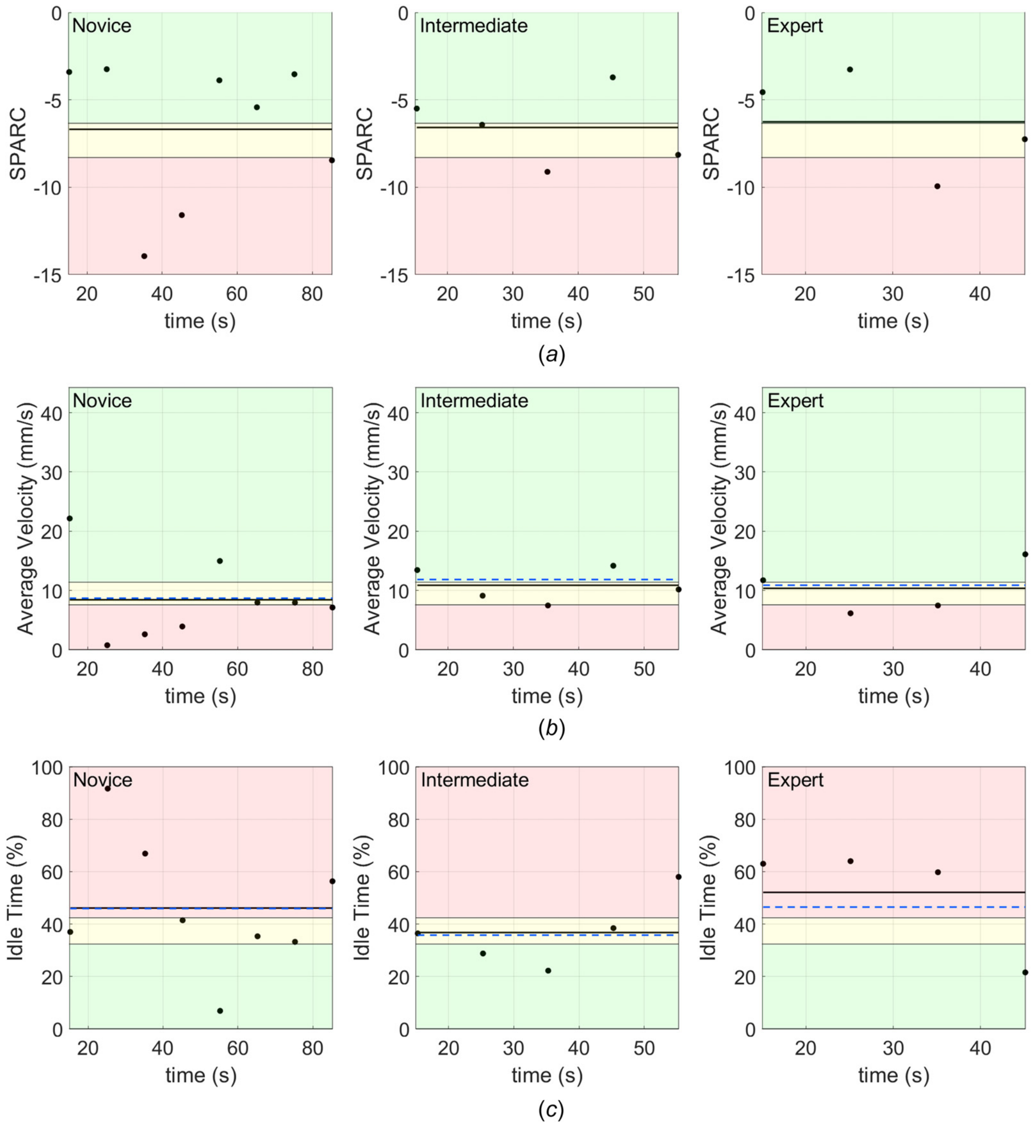

Fig. 8 Distribution of online performance measures from novice, intermediate, and expert participants for FEVS task 1 using 15 s sliding window with 5 s overlap between windows. For SPARC and average velocity, thresholds for high, medium, and poor performance given by the top, middle, and bottom shaded regions, respectively. For idle time, high performance corresponds to bottom region and poor performance corresponds to top region. Averaged online (dashed line) and offline (solid line) values of each metric are also shown.

In contrast, the higher and nearly constant online-offline correlations associated with average velocity and idle time provide a stronger basis for their use as online performance measures. As with the correlations produced by online SPARC, discrete windows produce noisier results, but correlations remain relatively constant and high-valued. The online computation of average velocity and idle time can be expected to produce correlation values roughly equivalent to those observed from their offline computation provided in Table 3. Average velocity shows slightly higher correlations than idle time, as taking the mean of a set of average velocities calculated across moving windows of time would result in a value that correlates highly with the average velocity of the entire motion profile and, by extension, with offline SPARC [23]. Idle time calculated online provides the percentage of the windowed segment in which the tool tip velocity is lower than a threshold defined to account for motion artifacts from tool tip and vessel wall interactions. Overlapping motion subprofiles within a navigation task would likely affect idle time values, but given that its computation does not require a nonlinear operation as with SPARC, idle time can be expected to produce high-valued correlations with offline SPARC. Additionally, the strong correlations exhibited by online average velocity and idle time suggest that these measures may also be correlated to the traditional structured rating scales, given their mutual correlation with offline SPARC [16].

4.3 Future Directions and Recommendations. To support our broader motivation for providing online performance feedback to surgical trainees, our results show that performance thresholds can be defined using the distributions of online SPARC, average velocity, and idle time generated by participants from each experience level. A potential feedback protocol can be illustrated by using a set of overlapping windows of 15 s length and 5 s overlap between windows. Such a window configuration would likely result in trainees receiving an acceptable amount of feedback cues during a navigation task, as shorter window lengths and overlap amounts may overload trainees and negatively affect performance.

After defining a set of online performance threshold regions for each metric and each navigation task (Fig. 7), we can demonstrate the possibility of delivering performance feedback to surgical trainees by calculating the online value of each metric from the most recent data window and determining the appropriate performance region it falls under (Fig. 8). The best way of delivering this information as feedback will likely vary in type (i.e., vibrotactile, visual) and threshold values computed for each navigation task.

The task-dependent nature of movement smoothness metrics and their indirect measures implies that the exact values and thresholds determined for one set of motion tasks will not be applicable for a different set of tasks. Therefore, comparing the average value of a performance metric calculated online with SPARC calculated offline provides a methodology to pinpoint a set of motion-based candidate metrics and to evaluate their utility as online performance measures for providing as feedback. The gold-standard metric to be used for the online–offline correlations may also vary depending on the application domain.

Given the different navigational elements of each task ranging from simple tool insertion/retraction to more complex vessel branch cannulation motions, it is likely that each task is not of constant difficulty. This observation is supported by the standard deviation of the distributions of online SPARC, average velocity, and idle time values within motion tasks, as apparent in Fig. 8, and warrants further work in determining trends in online values for subtasks within navigation tasks. Using the averaged online value of any metric to directly develop performance thresholds for evaluation or feedback requires the assumption that task difficulty is constant throughout a motion task and that online values calculated at each window remains close to the average value across all windows. Thus, while using the average online values for SPARC, average velocity, and idle time for the online-offline comparisons establishes their general utility as online performance measures, variations in online values corresponding to more difficult subtasks, such as during branch cannulation, may preclude their direct use for performance feedback.

Performing online evaluation and feedback using performance thresholds based on the average online values of SPARC, average velocity, and idle time may still result in positive improvements in navigation strategies, as was discussed in Jantscher [28]. Unlike laparoscopic or open surgery, endovascular procedures may be more amenable for such feedback since they involve constrained input motions of insertion, retraction, or rotation of the guidewire and catheter, despite showing large variation in online tool movement smoothness measures within tasks. Difficulties in tool navigation arising from the nonlinear mapping between input and tool tip motions may still require the identification of smaller subtasks in which the performance thresholds could be more appropriately defined.

Given the low and noisy correlations between online and offline SPARC, this metric cannot be used reliably as a basis for online performance feedback, especially for motion tasks that consist of unconstrained movements that generate various types of discrete motion subprofiles. A special case for online computation in which SPARC might perform better would be for navigation tasks that consist of a closed and constrained path, similar to the ones used in Jantscher [28], which would guarantee that participants generate a smaller set of motion subprofiles by following similar motion trajectories. Another possible alternative method of improving the online calculation of SPARC would be to spatially segment motion trajectories into their discrete motion subprofiles [19], in a manner that would guarantee that each value of SPARC would correspond to a single subprofile, with minimal overlap between adjacent subprofiles that reflect different motion strategies or are of varying difficulty.

5 Conclusion

Improved postoperative outcomes are linked to surgical expertise, which provides a strong motivation for research on objective and quantitative performance evaluation. While objective evaluation techniques are more common in other surgical domains, their application in endovascular surgery is sparse. Frequency-domain movement smoothness (SPARC), average tool tip velocity, and idle time are quantitative motion quality metrics calculated from tool tip motion that show significant differences across experience level and navigation task. The capability of these metrics in providing an offline measure of experience level in the midst of differences across task is complemented by their potential as online indicators of performance that can provide a basis for intuitive and robust feedback.

Average velocity and idle time correlate well with SPARC as online and offline measures, and offer promising alternatives to SPARC for delivering to trainees as online feedback of movement smoothness information. These metrics are more amenable for online computation and are more effective at conveying movement smoothness information in a more intuitive manner than SPARC. Transitioning the findings of this paper toward an online evaluation and feedback paradigm will require additional evaluation of the motion tasks to be used for determining subtasks within a navigation task in which the online values across experience levels are different from that of the average, in addition to the most effective method of delivering these performance measures as feedback to trainees.

Acknowledgment

The authors thank Wyatt Crider, Virginia Budzinski, Dr. Malachi Sheahan, and Dr. Mark Mattos for their valuable support with data collection and subject recruiting. We also thank Simbionix for their support with hardware and software throughout the duration of our study.

Funding Data

- National Science Foundation (Grant No. IIS1638073; Funder ID: 10.13039/100000001).
- This work was supported by BOTH the GRFP and the NSF grant.

References

[1] Biancari, F., Catania, A., and D'Andrea, V., 2011, "Elective Endovascular Vs. Open Repair for Abdominal Aortic Aneurysm in Patients Aged 80 Years and Older: Systematic Review and Meta-Analysis," Eur. J. Endovascular Vasc. Surg., **42**(5), pp. 571–576.

[2] Weibers, D. O., 2003, "Unruptured Intracranial Aneurysms: Natural History, Clinical Outcome, and Risks of Surgical and Endovascular Treatment," Lancet, **362**(9378), pp. 103–110.

[3] Smith, C. R., Leon, M. B., Mack, M. J., Miller, D. C., Moses, J. W., Svensson, L. G., Tuzcu, E. M., Webb, J. G., Fontana, G. P., Makkar, R. R., Williams, M., Dewey, T., Kapadia, S., Babaliaros, V., Thourani, V. H., Corso, P., Pichard, A. D., Bavaria, J. E., Herrmann, H. C., Akin, J. J., Anderson, W. N., Wang, D., and Pocock, S. J., 2011, "Transcatheter Versus Surgical Aortic-Valve Replacement in High-Risk Patients," New Engl. J. Med., **364**(23), pp. 2187–2198.

[4] Khoynezhad, A., and Kruse, M., 2009, "Guidewires, Catheters, and Sheaths Used for Thoracic Endografting Procedures," J. Card. Surg., **24**(2), pp. 113–119.

[5] Ahmadi, R., Willfort, A., Lang, W., Schillinger, M., Alt, E., Gschwandtner, M. E., Haumer, M., Maca, T., Ehringer, H., and Minar, E., 2001, "Carotid Artery Stenting: Effect of Learning Curve and Intermediate-Term Morphological Outcome," J. Endovascular Ther., **8**(6), pp. 539–546.

[6] Gurvitch, R., Tay, E. L., Wijesinghe, N., Ye, J., Nietlispach, F., Wood, D. A., Lichtenstein, S., Cheung, A., and Webb, J. G., 2011, "Transcatheter Aortic Valve Implantation: Lessons From the Learning Curve of the First 270 High-Risk Patients," Catheter. Cardiovasc. Interv., **78**(7), pp. 977–984.

[7] Lobato, A. C., Rodriguez-Lopez, J., and Diethrich, E. B., 2002, "Learning Curve for Endovascular Abdominal Aortic Aneurysm Repair: Evaluation of a 277-Patient Single-Center Experience," J. Endovascular Ther., **9**(3), pp. 262–268.

[8] Prinssen, M., Verhoeven, E. L., Buth, J., Cuypers, P. W., van Sambeek, M. R., Balm, R., Buskens, E., Grobbee, D. E., and Blankensteijn, J. D., 2004, "A Randomized Trial Comparing Conventional and Endovascular Repair of Abdominal Aortic Aneurysms," New Engl. J. Med., **351**(16), pp. 1607–1618.

[9] Rihal, C. S., Nishimura, R. A., and Holmes, D. R., 1991, "Percutaneous Balloon Mitral Valvuloplasty: The Learning Curve," Am. Heart J., **122**(6), pp. 1750–1756.

[10] Lurz, P., Coats, L., Khambadkone, S., Nordmeyer, J., Boudjemline, Y., Schievano, S., Muthurangu, V., Lee, T. Y., Parenzan, G., Derrick, G., Cullen, S., Walker, F., Tsang, V., Deanfield, J., Taylor, A. M., and Bonhoeffer, P., 2008, "Percutaneous Pulmonary Valve Implantation: Impact of Evolving Technology and Learning Curve on Clinical Outcome," Circulation, **117**(15), pp. 1964–1972.

[11] Reznick, R., Regehr, G., MacRae, H., Martin, J., and McCulloch, W., 1997, "Testing Technical Skill Via an Innovative "Bench Station" Examination," Am. J. Surg., **173**(3), pp. 226–230.

[12] Jackson, M. J., Pandey, V., and Wolfe, J. H. N., 2003, "Training for Infrainguinal Bypass Surgery," Eur. J. Vasc. Endovascular Surg., **26**(5), pp. 457–466.

[13] Duran, C., Estrada, S., O'Malley, M., Sheahan, M. G., Shames, M. L., Lee, J. T., and Bismuth, J., 2015, "The Model for Fundamentals of Endovascular Surgery (Fevs) Successfully Defines the Competent Endovascular Surgeon," J. Vasc. Surg., **62**(6), pp. 1660–1666.

[14] Moorthy, K., Munz, Y., Sarker, S. K., and Darzi, A., 2003, "Objective Assessment of Technical Skills in Surgery," BMJ, **327**(7422), pp. 1032–1037.

[15] Schwein, A., Kramer, B., Chinna Durai, P., Walker, S., O'Malley, M., Lumsden, A., and Bismuth, J., 2016, "Flexible Robotics With Electromagnetic Tracking Improve Safety and Efficiency During In Vitro Endovascular Navigation," J. Vasc. Surg., **63**(1), pp. 285–286.

[16] Estrada, S., Duran, C., Schulz, D., Bismuth, J., Byrne, M. D., and O'Malley, M. K., 2016, "Smoothness of Surgical Tool Tip Motion Correlates to Skill in Endovascular Tasks," IEEE Trans. Human Mach. Syst., **46**(5), pp. 647–659.

[17] Tausch, T. J., Kowalewski, T. M., White, L. W., McDonough, P. S., Brand, T. C., and Lendvay, T. S., 2012, "Content and Construct Validation of a Robotic Surgery Curriculum Using an Electromagnetic Instrument Tracker," J. Urol., **188**(3), pp. 919–923.

[18] Nisky, I., Okamura, A. M., and Hsieh, M. H., 2014, "Effects of Robotic Manipulators on Movements of Novices and Surgeons," Surg. Endoscopy, **28**(7), pp. 2145–2158.

[19] Krishnan, S., Garg, A., Patil, S., Lea, C., Hager, G., Abbeel, P., and Goldberg, K., 2016, "Transition State Clustering: Unsupervised Surgical Trajectory Segmentation for Robot Learning," Int. J. Rob. Res., **36**(13–14), pp. 1595–1618.

[20] Rafii-Tari, H., Payne, C. J., Liu, J., Riga, C., Bicknell, C., and Yang, G., 2015, "Towards Automated Surgical Skill Evaluation of Endovascular Catheterization Tasks Based on Force and Motion Signatures," IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, May 26–30, pp. 1789–1794.

[21] O'Malley, M. K., Byrne, M. D., Estrada, S., Duran, C., Schulz, D., and Bismuth, J., 2019, "Expert Surgeons Can Smoothly Control Robotic Tools With a Discrete Control Interface," IEEE Trans. Human-Mach. Syst., **49**(4), pp. 388–394.

[22] Malpani, A., Vedula, S. S., Lin, H. C., Hager, G. D., and Taylor, R. H., 2020, "Effect of Real-Time Virtual Reality-Based Teaching Cues on Learning Needle Passing for Robot-Assisted Minimally Invasive Surgery: A Randomized Controlled Trial," Int. J. Comput. Assisted Radiol. Surg., **15**(7), pp. 1187–1194.

[23] Murali, B., Belvroy, V. M., Pandey, S., Byrne, M. D., Bismuth, J., and O'., and Malley, M. K., 2020, "Towards Automated Performance Assessment Using Velocity-Based Motion Quality Metrics," Proceedings of the IEEE International Symposium on Medical Robotics, Atlanta, GA, Apr. 22–24, pp. 1–7.

[24] Davaris, M., Wijewickrema, S., Zhou, Y., Piromchai, P., Bailey, J., Kennedy, G., and O'Leary, S., 2019, "The Importance of Automated Real-Time Performance Feedback in Virtual Reality Temporal Bone Surgery Training," *Artificial Intelligence in Education*, S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, and R. Luckin, eds. (Lecture Notes in Computer Science, Vol. 11625), Springer International Publishing, Cham, pp. 96–109.

[25] Chandler, J. H., Mushtaq, F., Moxley-Wyles, B., West, N. P., Taylor, G. W., and Culmer, P. R., 2017, "Real-Time Assessment of Mechanical Tissue Trauma in Surgery," IEEE Trans. Biomed. Eng., **64**(10), pp. 2384–2393.

[26] Ershad, M., Rege, R., and Majewicz Fey, A., 2019, "Automatic and Near Real-Time Stylistic Behavior Assessment in Robotic Surgery," Int. J. Comput. Assisted Radiol. Surg., **14**(4), pp. 635–643.

[27] Chi, W., Rafii-Tari, H., Payne, C. J., Liu, J., Riga, C., Bicknell, C., and Yang, G., 2017, "A Learning Based Training and Skill Assessment Platform With Haptic Guidance for Endovascular Catheterization," IEEE International Conference on Robotics and Automation (ICRA), Singapore, May 29–June 3, pp. 2357–2363.

[28] Jantscher, W. H., 2018, "Using Real-Time Smoothness Metrics to Deliver Haptic Performance Cues for a Dexterous Task," Master's thesis, Rice University, Houston, TX.

[29] Balasubramanian, S., Melendez-Calderon, A., Roby-Brami, A., and Burdet, E., 2015, "On the Analysis of Movement Smoothness," J. NeuroEng. Rehab., **12**(1), p. 112.

[30] Belvroy, V. M., Murali, B., Sheahan, M. G., O'Malley, M. K., and Bismuth, J., 2020, "Motion Metrics Reliably Differentiate Competency: Fundamentals of Endovascular and Vascular Surgery (FEVS)," J. Vasc. Surg., **72**(6), pp. 2161–2165.

[31] D'Angelo, A.-L. D., Rutherford, D. N., Ray, R. D., Laufer, S., Kwan, C., Cohen, E. R., Mason, A., and Pugh, C. M., 2015, "Idle Time: An Underdeveloped Performance Metric for Assessing Surgical Skill," Am. J. Surg., **209**(4), pp. 645–651.

[32] Jantscher, W. H., Pandey, S., Agarwal, P., Richardson, S. H., Lin, B. R., Byrne, M. D., and O'., and Malley, M. K., 2018, "Toward Improved Surgical Training: Delivering Smoothness Feedback Using Haptic Cues," IEEE Haptics Symposium (HAPTICS), IEEE, San Francisco, CA, Mar. 25–28, pp. 241–246.

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

[34] Estrada, S., O'Malley, M. K., Duran, C., Schulz, D., and Bismuth, J., 2014, "On the Development of Objective Metrics for Surgical Skills Evaluation Based on Tool Motion," IEEE International Conference on Systems, Man, and Cybernetics (SMC), San Diego, CA, Oct. 5–8, pp. 3144–3149.