An Evaluation of Inanimate and Virtual Reality Training for Psychomotor Skill Development in Robot-Assisted Minimally Invasive Surgery

Guido Caccianiga, Andrea Mariani, Elena De Momi, *Senior Member, IEEE*, Gabriela Cantarero, and Jeremy D. Brown, *Member, IEEE*

***Abstract*—Robot-assisted minimally invasive surgery (RAMIS) is gaining widespread adoption in many surgical specialties, despite the lack of a standardized training curriculum. Current training approaches rely heavily on virtual reality simulators, in particular for basic psychomotor and visuomotor skill development. It is not clear, however, whether training in virtual reality is equivalent to inanimate model training. In this manuscript, we seek to compare virtual reality training to inanimate model training, with regard to skill learning and skill transfer. Using a custom-developed needle-driving training task with inanimate and virtual analogs, we investigated the extent to which N=18 participants improved their skill on a given platform post-training, and transferred that skill to the opposite platform. Results indicate that the two approaches are not equivalent, with more salient skill transfer after inanimate training than virtual training. These findings support the claim that training with real physical models is the gold standard, and suggest more inanimate model training be incorporated into training curricula for early psychomotor skill development.**

***Index Terms*—Automated, inanimate, minimally invasive, objective, robot-assisted, sensors, simulation, skill transfer, surgery, training, virtual reality.**

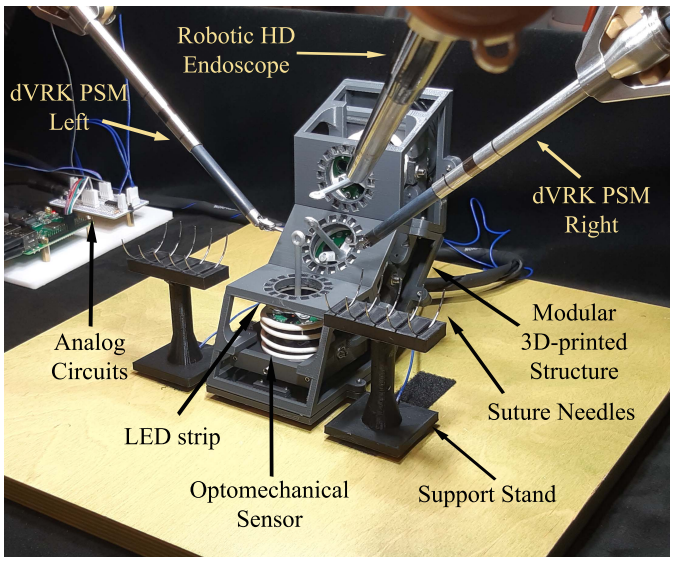


Fig. 1. Inanimate END task: experimental setup is shown with all its components under the dVRK PSMs and the robotic endoscope.

I. Introduction

ROBOT-ASSISTED minimally invasive surgery (RAMIS) is quickly becoming the standard of care for a number of routine and non-routine surgical procedures in various surgical specialties [1]–[4]. Despite its prevalence in both large academic hospital centers and smaller regional hospitals, there exists no standardized training curriculum for novice surgical trainees [5]. Given the increasing constraints on resident work hours and increasing legal pressure on patient safety, a greater portion of training is moving away from the operating theater in the form of simulation-based training [6], [7]. Thus, individual institutions are left to create their own RAMIS training curricula.

Current approaches to RAMIS training utilize a combination of live clinical robot training with inanimate, animal, and cadaver models, and virtual reality (VR) training [8]–[11]. Live robot training with inanimate training models has been shown to correlate well with in-vivo training, and is for many considered the gold standard [11]. At the same time, live robot training can be resource intensive and present logistical problems, especially for hospitals that do not have a dedicated robotic system for training purposes. In addition, live robot training currently necessitates the use of structured human grading for skill assessment, which can be subjective and time-consuming. While their is considerable potential of one day automating skill assessment for live robot training using both the kinematics of the robot [12]–[14] as well as the robot's

Manuscript received December 15, 2019; revised March 22, 2020; accepted April 20, 2020. Date of publication April 28, 2020; date of current version May 20, 2020. This article was recommended for publication by Associate Editor D.-S. Kwon and Editor P. Dario upon evaluation of the reviewers' comments. This work was supported in part by JHU internal fundings and in part by the Tesi All'estero Scholarship of Politecnico di Milano. *(Corresponding author: Guido Caccianiga.)*

Guido Caccianiga was with the Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD 21218 USA. He is now with the Department of Electronics, Information, and Bioengineering, Politecnico di Milano, 20133 Milan, Italy (e-mail: gcaccia1@jh.edu).

Andrea Mariani is with the Biorobotics Institute, Scuola Superiore Sant'Anna, 56025 Pisa, Italy.

Elena De Momi is with the Department of Electronics, Information, and Bioengineering, Politecnico di Milano, 20133 Milan, Italy.

Gabriela Cantarero is with the Department of Physical Medicine and Rehabilitation, Johns Hopkins Medical Institute, Baltimore, MD 21218 USA.

Jeremy D. Brown is with the Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD 21218 USA.

This article has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TMRB.2020.2990692

physical interactions with the surgical environment [15], [16], these advances are still far from clinical readiness.

VR training, on the other hand, allows for repeatable practice of skill-building exercises with easily-measured objective skill metrics, and has also been shown to correlate well with inanimate and in-vivo training [11]. VR training has seen rapid growth in recent years, due in large part to advances in computational power, leading to the development of training platforms like the dV-Trainer by Mimic Technologies [17], the RobotiX Mentor by 3D Systems [18], and the da Vinci SimNow skill simulator by Intuitive Surgical [19]. Still, VR training can at best provide a virtual approximation of a real training environment, which may not match exactly the behavior of the real physical environment.

For either training approach predictive validity is required for clinical adoption, and preliminary evidence suggest that skill obtained in VR and inanimate training do transfer to OR performance [11], [20]–[22]. It should be noted, however, that these assessments generally rely on potentially subjective skill measures such as structured human grading. What is still unclear, however, is whether skill obtained during VR training is equivalent to skill obtained during live inanimate training, and how readily skill obtained through one training approach transfers to the other approach given the manner in which they are interchanged in most clinical training settings.

Several prior studies have attempted to compare inanimate and VR training [23]–[30]. Unfortunately, the experimental methods utilized introduced confounding factors that limit generalization of the experimental findings. Perrenot *et al.* for example trained participants on a set of VR tasks and then evaluated them on a similar inanimate platform [27]. While VR performance was automatically computed using the simulation software, inanimate training performance relied heavily on structured human grading, which produced subjective performance assessments that were not procedure specific. Likewise, Newcomb *et al.* compared the Robotic Training Network's inanimate exercises with the VR tasks available on the da Vinci Skills Simulator [28]. Since each training group performed the exercises only on the assigned platform, no direct comparison between the VR training protocol and the inanimate tasks was possible. In a similar fashion, Brown *et al.*, in an effort to avoid reliance on different scoring systems used the VR simulator as baseline and final evaluation for the two participant groups (VR and inanimate) [29]. The comparison, however, was limited by the varying difficulty of the training tasks, and the fact that the designed protocol favored the VR group on the final testing phase.

It is also worth mentioning that Satava and colleagues developed an evenly structured validation protocol and a multi-functional training platform as part of the Fundamentals of Robotic Surgery validation trial [30]. Two identical VR and inanimate versions of the training task platform were developed, and results indicate that participants significantly improved their performance on both platforms. Still, both the inanimate platform and the avian model used for baseline and post-training evaluation lacked any form of sensorization, forcing all assessments to be performed using structured human grading. The moderate inter-rater reliability scores observed therefore limits a thorough comparison of the two platforms [5].

Recently, the work of Wang *et al.* has come the closest in terms of objectively comparing VR and inanimate training through a human-centered experimental design approach [31]. In their study, participants performed the nearly identical VR and inanimate training task while kinesthetic and physiological measurements were taken. Their results suggest that training in VR environments may be oversimplified (participants exerted less effort), allowing trainees to adopt skill behaviors that hinder performance on the live robot.

To build on these findings and conduct a comprehensive comparison of inanimate and VR approaches to RAMIS training, several requirements must be satisfied. First, the platforms used in either approach must be inanimate and virtual analogs of the same training task, identical in both form and function. Second, both platforms should produce the same metrics for task performance assessment. Given the simplicity with which metrics are generated in virtual simulation, considerable effort needs to be placed on instrumentation of the inanimate platform to produce task performance measures that are both goal-relevant and compatible with the metrics recorded on the VR platform. Finally, regarding the study design, it is essential to assess to what degree participants' skill improves in each approach, as well as the extent to which that skill transfers to the opposite approach.

In this manuscript, we present an experimental platform and user study designed to thoroughly compare inanimate and VR approaches to RAMIS training. In what follows, we discuss the design of the Inanimate and VR training task platforms, followed by an overview of the user study. Results of the user study will then be described in terms of performance improvement post-training (skill learning) and performance comparisons on the opposite platform (skill transfer). Finally, we will discuss the results of our work in the broader context of surgical training. We hypothesize that skill learning will occur for both platforms, and that skill transfer will be more clearly defined for the Inanimate training platform.

II. Virtual and Physical Platforms

A. Task Concept

Our Enhanced Needle Driving (END) task was designed to evaluate needle driving, a motor skill that is a common step in many suturing procedures. The objective of the task is to drive a 20mm radius suture needle through 3 rings following a curved trajectory. The task is rendered in both real and virtual environments as described in the following sections, and the desired trajectory of the needle is defined by three rings positioned at 0, 45, and 90 degrees from the horizontal plane along a 20mm radius circumference. These rings resemble the entry, pass-through, and exit points of the ideal trajectory required to close the incision in a thick tissue flap. To evaluate task performance, we measured any deviations of the needle from the ideal trajectory.

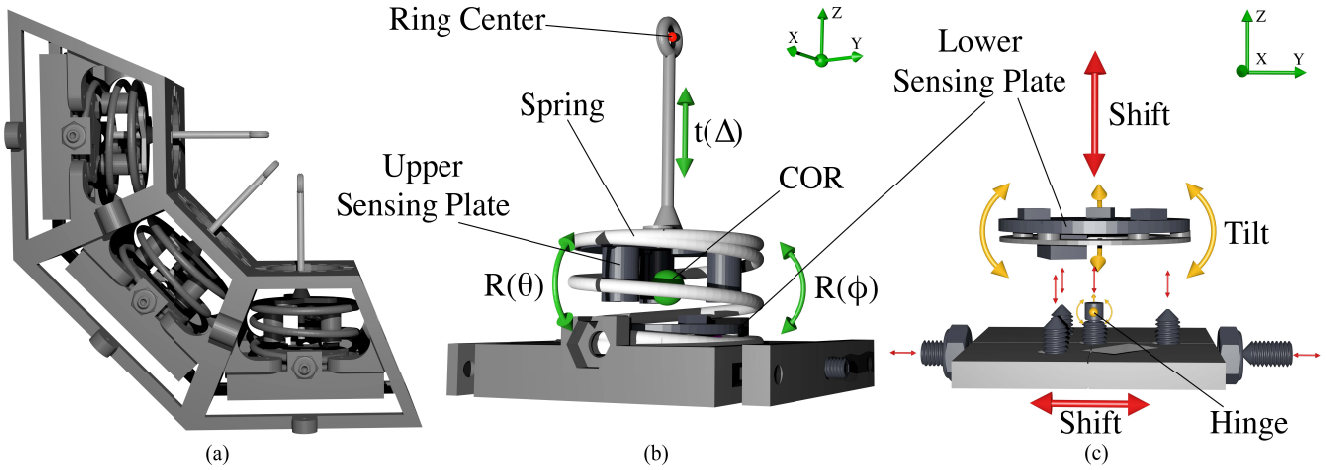


Fig. 2. CAD Models of the Enhanced Needle Driving (END) platform: (a) whole platform consisting of interconnected modules; (b) optomechanical sensor; the degrees of freedom are shown (green arrows) along with the center of rotation (green sphere) between the sensing plates (dark gray); (c) tilt-shift mechanism to calibrate the lower sensing plate's position and orientation.

B. Telerobotic Platform

All research activities reported in the manuscript were performed using the da Vinci Research Kit (dVRK), an open-source telerobotic system derived from the first generation da Vinci Surgical System [32]. The dVRK features a unilateral control architecture that enables teleoperation: the motion of two master tool manipulators (MTMs) at the surgeon's console is replicated by the patient-side manipulators (PSMs), analogously to the clinical da Vinci surgical system. The dVRK also features a robotic endoscope with two optical channels whose images are displayed by a HD stereoscopic viewer at the surgeon's console. In addition, the dVRK supports haptic guidance, data logging, and virtual/mixed reality applications [33]–[35].

C. Inanimate Task

The Inanimate END task (Fig. 1) is composed of a modular 3D printed structure containing three custom optomechanical sensors to measure the 3-dimensional movement of the rings. Visual feedback of the each ring's rotational and translational displacement is provided by a circular LED light strip centered co-axially about the ring assembly. Task progression is monitored using analog circuits generating binary contact signals between the task components of interest. The full dVRK is used for the Inanimate setup along with a Robot Operating System (ROS) [36] interface for data acquisition and control. The task is made available both for left and right hand dominant participants by mirroring the task setup. Complete details of the Inanimate task are discussed in the following sections.

1) Optomechanical Sensors: Each modular custom optomechanical sensor was constructed from the optical sensor of a SpaceNavigator mouse (3Dconnexion Inc.), which is capable of measuring the relative rotations and translation between the two sensing plates of the sensor with a resolution of 170 increments per angular degree and 250 increments per mm, respectively. The lower sensing plate features three IR detectors that measure the intensity of light from three separate IR emitters attached to the upper plate. Fig. 2 shows the CAD models of the END platform as developed: a 3D-printed 45mm spring connects to upper plate to the lower plate, thereby allowing the upper plate to freely move (Fig. 2b). Each ring assembly is rigidly attached to the top of the respective upper plate creating the entry, pass-through, and exit points of the ideal needle trajectory (Fig. 2a).

Given the spring's large diameter/height ratio (1.73:1) and limited number of coils (2), and the length of the ring shaft (65mm), the spring can be modeled as generating a center of rotation (COR) for the ring assembly that can be approximated as centered between the upper and lower sensing plates. This COR translates with the upper assembly when the spring is elongated (Fig. 2b). The resulting kinematics of the ring can thus be approximated by two rotations ($\boldsymbol{R_x}$, $\boldsymbol{R_y}$) around the COR and one translation ($\boldsymbol{t_z}$) along the axis extending from the COR to the ring center. This model was validated by measuring the sensor outputs when forces were applied to the ring. The resulting signals corresponding to the rotation around the ring shaft ($\boldsymbol{R_z}$), and the lateral translations of the COR ($\boldsymbol{t_x}$, $\boldsymbol{t_y}$) were observed to be at least one order of magnitude smaller than the main displacements ($\boldsymbol{R_x}$, $\boldsymbol{R_y}$, $\boldsymbol{t_z}$). Therefore, $\boldsymbol{R_z}$, $\boldsymbol{t_x}$, and $\boldsymbol{t_y}$ can be considered negligible. While the real behavior of the mechanical system may diverge from our model-based assumptions when extreme displacements are applied to the ring assembly, these situations can be considered negligible given that they are outside the working (linear) range of the sensor itself.

The lower plate is attached to a tilt-shift platform (Fig. 2c) that allows for fine tuning of its position and orientation to cancel any orientation offset between the upper and lower plates caused by the different angular configuration of each sensor. Each optomechanical sensor is then enclosed in a 3D-printed housing. The final task is composed of three interconnected modular housings (Fig. 2a).

2) Visual Feedback: For each module, the displacement signals coming from the optomechanical sensor are mapped onto a 16x5050 RGB LED circular strip (Worldsemi Co.) which is located co-axially with respect to the shaft of the ring assembly (see Fig. 1). When providing feedback to the trainee, it is

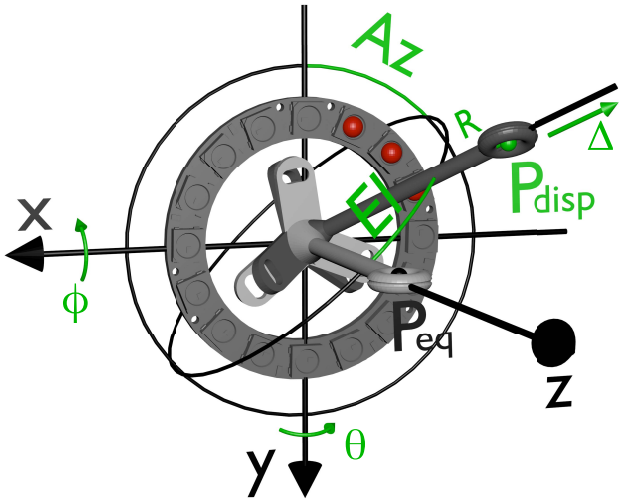


Fig. 3. Conceptual Mapping of the LED Visual Feedback: in this example the ring center is displaced to P_{disp} from equilibrium P_{eq} by a positive and negative rotation of ϕ and θ and a positive translation of Δ. Therefore, the azimuth turns on the central LED of the quadrant, the red color is chosen because of the high elevation angle, and the translation turns on two additional lateral LEDs.

important to give information about the error magnitude and direction. This can be described using the following spherical coordinate framework (Fig. 3): the signed ratio between the rotation errors (ϕ, θ) defines the azimuthal angle (Az), the greater absolute value between ϕ and θ defines the elevation angle (El), and the translation Δ defines the radius (R). The azimuthal angle is mapped to a specific LED out of the 16 that describes the planar direction of the ring's displacement. The elevation angle is compared to two thresholds to determine the color of the LED (off, orange, red), and it describes the maximum angular error. The radius is mapped into the LEDs adjacent to the LED controlled by the azimuthal angle, with Δ compared to three thresholds that turn on additional LEDs (up to three on each side). The resulting visual feedback intuitively displays the rotational and translational errors with separate cues.

A demonstration video of the visual feedback during the task execution can be found in the supplemental material.

3) Contact Circuits: To track task initiation and progression in real-time, contact circuits were placed on the robotic instruments, as well as on the task itself (Fig. 4).

Each daVinci Large Needle Driver [37] was modified by soldering a conductive wire to one of the internal metallic tendon cables creating a monopolar connection with the instrument tip (Fig. 4a). Likewise, a custom 3D-printed needle holder composed of dual-purpose magnets was used to hold the 1/2 round GS-26 taper needles as well as connect them to one end of a voltage divider circuit (Fig. 4b). These two circuits allowed for automatic detection of the contact between instrument and needle during pick-up and hand-to-hand transfer.

To track needle progression during the task, each ring assembly was equipped with two thin conductive wires running along the inner and outer surface of the ring (Fig. 4c). These two wires are connected in parallel to the original SpaceNavigator mouse button circuits. Thus, contact by the

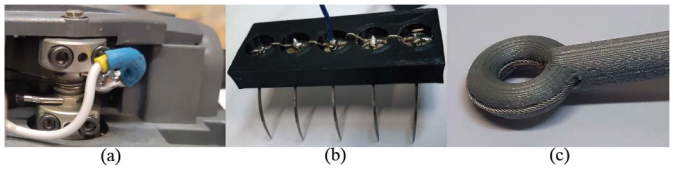


Fig. 4. Inanimate setup contact circuits: (a) wire connection to the Large Needle Driver pulley; (b) magnet and wiring on the needle support; (c) ring contacts on the inner and outer surface.

instrument or needle will short-circuit the button and generate a digital signal.

4) Signal Acquisition and Control: The ring displacement and button contact signals are directly acquired by a Linux-based Raspberry PI through the original SpaceNavigator mouse USB-HID protocol. The analog signals from the voltage divider contact circuits are acquired by A/D chips on the same board. The LED rings are controlled by means of PWM signals generated using the real-time Direct Memory Access and the Adafruit-Blinka Library [38].

The software of the Inanimate platform was developed as a network of ROS nodes generated using the rospy Python [39] client library. This ROS network can easily communicate with other devices such as the dVRK or a Graphical User Interface (See IV-3) through LAN. The whole Inanimate task is therefore portable and it can run without the need of a desktop PC.

D. Virtual Task

The Virtual END task was designed as a virtual analog of the Inanimate END task, both in terms of form and function. The virtual environment was generated using the ATAR framework [40] and it was interfaced with the master console of the dVRK. The simulated task was rendered by the stereo-viewer and the virtual robotic instruments are controlled by the dVRK master tool manipulators. Visco-elastic constraints were used to generate virtual multi-directional hinges that reproduce the movements of the ring assemblies in the physical task. For each simulated object, physical properties were governed by software parameters while the graphical renderings were generated using meshes based on the CAD drawings of the physical task (Fig. 2). Complete details of the Virtual task are discussed in the following sections.

1) Virtual Object Rendering: The CAD models used for 3D printing the Inanimate task platform were exported from Solidworks to generate meshes in the 3D rendering software Blender. After verifying the scaling and orientation, the meshes were loaded in ATAR using a convex hull decomposition approach [41]. The object meshes were automatically decomposed and simplified (where necessary) based on the requirements of reducing the computational load of the graphics engine while also ensuring real-time interactivity and realistic collision detection. The contact between objects was defined using a collision detection function from the Bullet library [42]; this provides pickup, transfer, and needle-ring contact information. Custom meshes were created for the following objects: GS-26 1/2 round taper needle, large needle drivers, task platform (ring assembly, LED-RGB-rings,

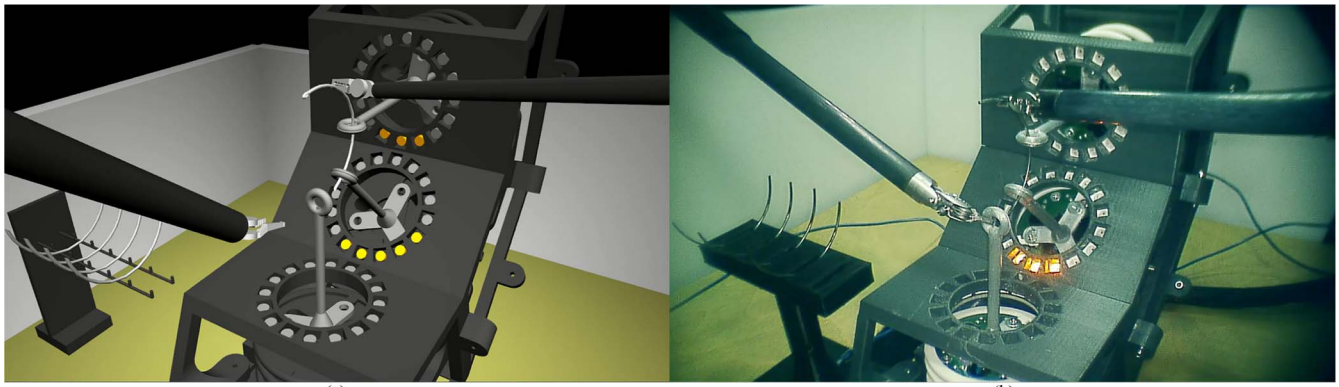


Fig. 5. Virtual (a) and Inanimate (b) tasks as seen from the stereo viewer of the surgeon's console. Proportions, perspective, and background are accurately reproduced between the two platforms.

modular support structures), and needle support stand. The visual feedback replicates the spherical coordinate mapping of the Inanimate platform LEDs, using simulated color-changing lights.

2) Virtual Task Mechanics: Three 6DoF constraints control the movement of the three virtual ring assemblies. Each constraint is modeled as a virtual spring-damper to resemble the dynamics of the optomechanical assembly in the Inanimate task as shown below. Each constraint is fixed on one side to a still reference frame oriented analogous to the real ring assembly in its resting position. The other side of the constraint is attached to the virtual ring assembly and moves accordingly. Two translations and one rotation are locked to recreate the 3DoF ($\boldsymbol{R_x}$, $\boldsymbol{R_y}$, $\boldsymbol{t_z}$) kinematics of the real sensors in the Inanimate task platform. In addition to controlling the virtual ring assembly kinematics, the constraints act as a virtual sensor that measures the displacement between the fixed reference frame and the moving ring.

In order to equate the sensor output between the two platforms, the three optomechanical sensors were calibrated using an external optical measurement system (NDI Polaris). This calibration was then used along with the 3D CAD model of the virtual sensors to ensure parity between the two platforms. The final setup for the Inanimate and Virtual Reality platforms are shown in Fig. 5, in comparison, as seen from the stereoscopic viewer inside the surgeon's console.

E. Data Acquistion and Control

The experimental session was managed through a graphical user interface based on QT widgets that control both the simulation and the real platform. The investigator could set the acquisition variables, the handedness of the task, record data, and home the robot MTMs and PSMs to restore a specific teleoperation workspace. All the ATAR, dVRK, and Inanimate platform variables were recorded at 60Hz.

F. Task Performance Measures

The displacements of the three rings in the task were used to measure deviations from the ideal needle driving trajectory. The sensors (virtual or real) underneath each ring assembly produced three raw displacement signals (ϕ, θ, Δ) that describe the rotational and translational displacement of the ring assembly with respect to its equilibrium position. To reconstruct the position of the ring in the 3-dimensional space $\boldsymbol{P_{disp}}$ (see Fig. 3) the three signals were combined in a kinematic chain as shown below

$$\boldsymbol{P_{disp}} = \boldsymbol{R_x} \cdot \boldsymbol{R_y} \cdot \boldsymbol{t_{ring}}^{\top} \tag{1}$$

where,

$$\boldsymbol{R_x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & -\sin(\phi) \\ 0 & \sin(\phi) & \cos(\phi) \end{bmatrix} \tag{2}$$

$$\boldsymbol{R_y} = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \tag{3}$$

$$\boldsymbol{t_{ring}} = \begin{bmatrix} 0 & 0 & \Delta \end{bmatrix} + \begin{bmatrix} 0 & 0 & L_{ring} \end{bmatrix} \tag{4}$$

and L_{ring} is the distance between the center of rotation (COR) of the ring assembly and the ring center.

The displaced position $\boldsymbol{P_{disp}}$ was then compared to the equilibrium position $\boldsymbol{P_{eq}}$, and the Euclidean distance D between the two centers was calculated as shown below.

$$D = norm\left(\boldsymbol{P_{eq}} - \boldsymbol{P_{disp}}\right) \tag{5}$$

where

$$\boldsymbol{P_{eq}} = \begin{bmatrix} 0 & 0 & L_{ring} \end{bmatrix} \tag{6}$$

The distance signal at each time sample n was calculated for each ring i of the platform as $D_i(n)$. The value of $D_i(n)$ at each sample n can be seen as the amount of stretch that the needle would have caused on real tissue due to deviations from the ideal trajectory. A comparison of the $D_i(n)$ signals for both platforms before training can be seen in Fig. 6. The signals show similar morphology and the same order of magnitude for the VR and Inanimate platforms.

III. Experiment

Using the Inanimate and Virtual task platforms described in Section II, we investigated both skill learning and skill transfer between simulation and dry lab in the needle driving task. All the experiments were conducted in the Laboratory for Computational Sensing and Robotics (LCSR) at Johns Hopkins University (Baltimore, USA).

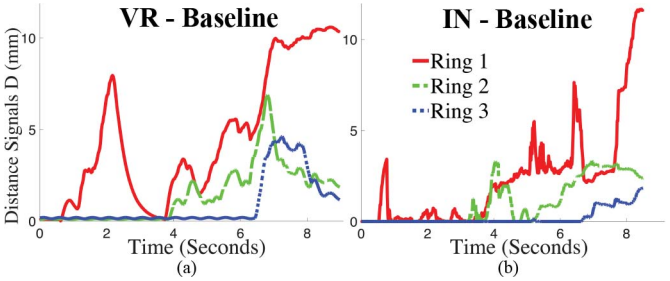


Fig. 6. Distance signals D for a single task execution as recorded from two representative participants (one for each platform) before training. Signals are of the same order of magnitude for the Virtual (a), and the Inanimate (b) tasks.

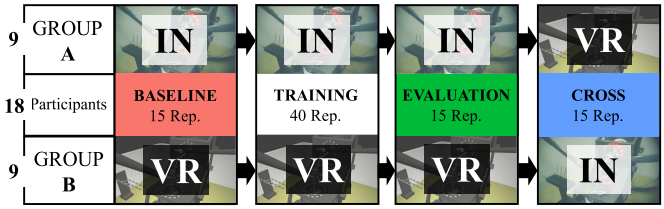


Fig. 7. Description of the study protocol structure.

A. Participants

This study was approved by the Johns Hopkins School of Medicine Institutional Review Board (study #00077792). 18 able-bodied participants were recruited for the study (8 females and 10 males; mean age 26.7 $\pm$ 3.7 years). 17 participants reported being right-hand dominant as assessed using the Edinborough Handedness Survey, one participant reported being left-hand dominant and did the task on the mirrored setup. Among the participants, 5 had medical backgrounds, however, all participants had no prior relevant experience with the da Vinci Surgical System (no involvement in actual surgical cases, user studies, or training). In addition, no participant had prior experience with robot teleoperation. All study activities took place in a single session lasting approximately 120 minutes and participants were compensated at a rate of $20 per hour.

The experiments presented here were part of a larger collaborative study investigating the effect of trans-cranial direct current stimulation (tDCS) on skill learning and skill transfer. This larger-scale study recruited 36 healthy participants and randomly assigned them one of two groups, stimulation and sham. Participants in both groups were connected to the tDCS hardware, however, only participants in the stimulation group received actual stimulation throughout the experiment. Here we are presenting results from only the sham group. In this way, we can focus our discussion on a thorough comparison of the two training approaches without the confounding factor introduced by tDCS. Proper discussion of the comparison between the stimulation and sham groups will be presented in a separate publication.

B. Experimental Protocol

In order to compare the Inanimate and Virtual training platforms, participants completed a single session experiment comprised of four Phases with a specified number of task repetitions for each stage: Baseline (BL - 15 repetitions), Training (TR - 40 repetitions), and Evaluation (EV - 15 repetitions) on one Platform, followed by Cross-evaluation (CR - 15 repetitions) on the other Platform. Participants were randomized (predefined sequence) into two groups (A and B). Group A performed the first three Phases (BL, TR, and EV) on the Inanimate platform and the final Phase (CR) on the Virtual platform. Group B performed the first three Phases (BL, TR, and EV) on the Virtual platform and the final Phase (CR) on the Inanimate platform (Fig. 7).

Participants also completed a brief survey regarding their demographics, handedness, and relevant medical background (used for tDCS eligibility). Participants then sat at the dVRK console and were provided with an overview of the experimental setup and task procedure. Participants were informed that the task objective was to drive the needle through the three rings of the task platform without causing the rings to deviate from their equilibrium positions. Participants were given an explanation of the RGB LED rings so that they could visually track the needle deviations.

In the Baseline, Evaluation, and Cross-evaluation Phases of the experiment, participants were instructed to complete the task in one of three requested amounts of time: Fast (5s), Moderate (15s), or Slow (25s). The three requested times were each presented a total of five times in a randomized fashion for each Phase (BL, EV, and CR) and communicated to the participant before each task execution. The corresponding start signal and countdown auditory cue was triggered by the participant by means of a dVRK foot pedal press. This methodology allowed for sampling of participants' skill level with a wide and even distribution across the speed-accuracy continuum [43]–[45]. In addition, these distributed time samples were used to derive the speed-accuracy function (SAF) to investigate the effect of tDCS on motor learning [46]. As previously mentioned, however, these investigations are not the focus of this manuscript.

It should be noted that in the actual experiment, the experimenter monitored each participant's actual task execution time to detect any behavior that was non-compliant regarding the requested times. Whenever the experimenter observed that the participant exceeded a $\pm$5s interval from the requested time, the requested time was reiterated on the next task repetition. In most cases, non-compliant participants preferred to perform the task at a speed close to the Moderate (15s) requested time. As a result, the experimenter had to over-sample the Fast (5s) and Slow (25s) requested times. In this way, systematically guiding the participants by means of frequent changes in the speed-accuracy tradeoff, together with a continuous supervision of the effective execution time, reduced the risk of outlier behavior. The experimenter also repeated any trial where it was observed that the participant dropped the needle.

C. Metrics and Statistical Analysis

The metric chosen for this study, M_{int}, is the sum of the time-normalized integrals of the distance signals $D_i(n)$ for the three rings. M_{int} therefore represents the cumulative history of the needle deviations across the three rings divided by the task

execution time (Δt), and is computed as:

$$M_{int_k} = \sum_{i=1}^{3} \left(\frac{\sum_n D_i(n)}{\Delta t_k} \right) \tag{7}$$

where i represents the number of rings, and n the number of samples.

M_{int} was calculated for each task execution k. Dividing the integral of the error by the actual time of execution normalizes the performance metric with respect to time. While surgical proficiency typically corresponds to low task error and fast task completion time, the chosen metric here allows for an analysis of each participant's performance across the speed-accuracy tradeoff that is invariant with respect to their actual completion time.

All statistical analyses were performed in R version 3.5.3. Given that our data set featured repeated measures and was unbalanced (see below), a Generalized Linear Mixed Model (GLMM) was chosen to model the between-subject and within-subject effects of the three independent variables (training Platform, experiment Phase, and requested time) on the dependent variable M_{int}. In addition to the three main effects, subjects were treated as a random effect. Data fitting was done using Maximum Likelihood Estimation (MLE). Post-hoc tests were run using simultaneous t-tests with Satterthwaite's method. Bonferroni corrections were applied to account for multiple comparisons in the data set. High p-values are saturated to 1.0 due to the aforementioned correction. A progressive approach was taken to model development wherein the simplest intercept-only model was constructed first, followed by a separate model for each of the three independent variables, and then models comparing the possible 2-way interactions of the independent variables. Finally, a model comparing the 3-way interaction of the three independent variables was constructed. Model selection was based on the Akaike Information Criterion (AIC) as a means of balancing model fit with model complexity.

When analyzing our dataset, we discovered that grouping the data with respect to requested times would have led to unbalanced group sizes given that the experimenter had to over-sample the Fast and Slow requested times for some non-compliant participants (as mentioned in Section III-B). This likely would have impacted the statistical power of our various comparisons. In assessing the distribution of actual task execution times, we observed that our data set could be roughly divided into three distinct groupings based on actual task execution time. We therefore created three new Speed groupings that were defined by two thresholds set at the 33rd and 66th percentile mark of our data set with respect to execution time. The resulting data set is divided into three roughly equal time groups (278 ± 1 samples) and the new Speed groupings are defined as Fast (<11.6s), Moderate (11.6-18.1s), and Slow (>18.1s). Given that the resulting distribution of our dependent variable M_{int} was not normally distributed, a Log_{10} transformation was applied to each data point to restore normality of the model residuals (Shapiro-Wilk Normality Test $W = 0.998$, $p = 0.848$). In our statistical models "Speed" has replaced "requested time" as an independent variable.

TABLE I
LINEAR MIXED MODELS PERFORMANCE COMPARISON INVESTIGATING THE EFFECTS OF INTERCEPT (Z), PLATFORM (P), PHASE (F), SPEED (S), AND THEIR INTERACTIONS

Model	Df	AIC	logLik	p-value
Z	3	450.94	-222.47	
P	4	452.91	-222.45	0.85
F	5	363.15	-176.57	<2E-16***
S	5	297.59	-143.8	<2E-16***
P:F + P:S + F:S	16	147.28	-57.64	<2E-16***
P:F:S	20	146.68	-53.34	0.071

$^{***}p < 0.001$; p-values are referring to comparison with the respective previous row

TABLE II
INITIAL SKILL LEVEL AT BASELINE (BL) FOR THE VIRTUAL (VR) AND INANIMATE (IN) TASK PLATFORMS

Comparison	β	SE	p-value
Intercept	-1.126	0.048	1.34E-54 ***
BL(VR-IN)Slow	-0.140	0.064	0.091
BL(VR-IN)Moderate	-0.114	0.063	0.214
BL(VR-IN)Fast	0.039	0.067	1

$^{***}p < 0.001$. Estimates are based on the Log_{10} data

IV. RESULTS

Table I shows the results of the various models described in Section III-C. A comparison of the first four models demonstrates that both Phase and Speed had a significant fixed effect on the independent variable M_{int} ($p < 0.001$). There was not a significant main effect of Platform on M_{int} ($p = 0.85$). A comparison of the two interaction models demonstrated that both the 2-way and 3-way interactions had a significant effect on the dependent variable M_{int}, and both models show very similar performance. Given that the 3-way interaction model has the lowest AIC and allows for consideration of all three fixed effects simultaneously, this model was chosen for its ability to allow post-hoc hypothesis testing at the lower clustering level (Platform:Phase:Speed).

Fig. 8 shows the distribution of M_{int} for each Platform, Phase, and Speed while Tables II, III, & IV report the model-based estimates for the various between-subject and within-subject comparisons. For clarity, the values reported in Fig. 8 are in non-transformed units. The values reported in Tables II, III, & IV are in Log_{10} units.

A. Initial Skill Level

We found no statistically significant differences between the VR and Inanimate groups at Baseline for the three Speeds. The results are reported in Table II.

B. Skill Learning and Transfer

1) Inanimate Training: For the Inanimate platform, performance during Evaluation showed significantly lower error (higher performance) than at Baseline for the Slow and Moderate Speeds. Performance during Cross-evaluation on the

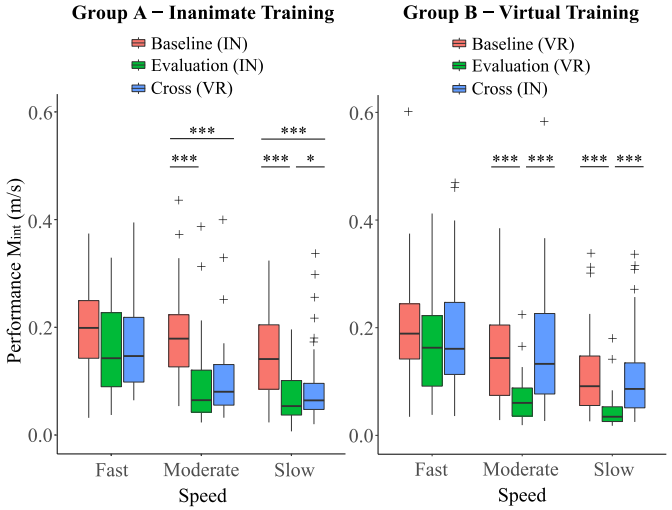


Fig. 8. M_{int} distributions and statistical significance for groups A and B at Fast (<11.6s), Moderate (11.6-18.1s), and Slow (>18.1s) Speeds. ($^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$; + Outliers).

TABLE III
GROUP A - INANIMATE TRAINING - RESULTS: BASELINE (BL), EVALUATION (EV), AND CROSS (CR) ARE COMPARED AT THE SLOW, MODERATE, AND FAST SPEEDS

Comparison	β	SE	$p-$value
IN(EV-BL)Slow	-0.357	0.053	1.68E-10 ***
IN(EV-BL)Moderate	-0.353	0.053	1.51E-10 ***
IN(EV-BL)Fast	-0.093	0.053	0.55
IN(CR-BL)Slow	-0.228	0.057	6.30E-05 ***
IN(CR-BL)Moderate	-0.255	0.053	9.66E-06 ***
IN(CR-BL)Fast	-0.061	0.054	1
IN(CR-EV)Slow	0.128	0.056	0.022 *
IN(CR-EV)Moderate	0.099	0.052	0.41
IN(CR-EV)Fast	0.321	0.053	1

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Estimates are based on the Log_{10} data

Virtual Reality platform was also better (lower error) than Baseline performance on the Inanimate platform for the Slow and Moderate Speeds. Performance during Cross-evaluation (on the Virtual Reality platform) was not significantly different than Evaluation for the Moderate Speed, and only slightly worse (higher error) than Evaluation for the Slow Speed. All comparisons at the Fast Speed were not significant. Complete results are reported in Table III.

2) Virtual Training: For the Virtual Reality platform, performance during Evaluation showed significantly lower error than at Baseline for the Slow and Moderate Speeds. Performance during Cross-evaluation on the Inanimate platform was not statistically different than performance at Baseline on the Virtual Reality platform for all three Speeds. Performance during Cross-evaluation (on the Inanimate platform) was significantly worse (higher error) than during Evaluation on the Virtual Reality platform for the Slow and Moderate Speeds. All comparisons at the Fast Speed were not significant. Complete results are reported in Table IV.

TABLE IV
GROUP B - VIRTUAL TRAINING - RESULTS: BASELINE (BL), EVALUATION (EV), AND CROSS (CR) ARE COMPARED AT THE SLOW, MODERATE, AND FAST SPEEDS

Comparison	β	SE	$p-$value
VR(EV-BL)Slow	-0.375	0.058	7.03E-10 ***
VR(EV-BL)Moderate	-0.303	0.056	7.14E-07 ***
VR(EV-BL)Fast	-0.129	0.058	0.233
VR(CR-BL)Slow	-0.019	0.050	1
VR(CR-BL)Moderate	0.015	0.051	1
VR(CR-BL)Fast	-0.067	0.054	1
VR(CR-EV)Slow	0.357	0.052	4.67E-11 ***
VR(CR-EV)Moderate	0.319	0.054	3.15E-08 ***
VR(CR-EV)Fast	0.062	0.050	1

$^{*}p < 0.05$; $^{***}p < 0.001$. Estimates are based on the Log_{10} data

V. DISCUSSION

In discussing our results we will focus on an overview of our major findings, a comparison of these finding to prior literature, and limitations of these findings.

A. Overview of Major Findings

Overall, our findings suggest that Inanimate and Virtual Reality approaches to training may not be equivalent with respect to basic psychomotor and visuomotor skill development. In particular, while skill learning occurred for both Platforms (Evaluation accuracy was significantly greater than Baseline accuracy) for the Moderate and Slow Speeds, skill transfer occurred differently between the two training platforms. For the Inanimate platform, participants' performance in Cross-evaluation on the VR platform was significantly greater than their Baseline performance on the Inanimate platform and not significantly different from their Evaluation performance at the Moderate Speed. The same trend holds true for the Slow Speed, however, we did observe that Cross-evaluation performance was significantly less, albeit marginally less, than Evaluation performance. We feel that this moderate level of significant difference (p=0.02) is likely caused by the presence of outliers. For the VR platform, participants' Cross-evaluation performance on the Inanimate platform was not significantly different than their Baseline performance on the VR platform and significantly worse than their Evaluation performance at both the Moderate and Slow Speeds.

These trends, however, do not hold at the Fast Speed; participants' performance did not change significantly between Baseline, Evaluation, or Cross-evaluation for either Platform. When choosing the three different requested times, we tried to pick three speeds that we felt would cover the widest range of the speed/accuracy tradeoff. We felt that our choices (5s, 15s, and 25s) were in line with other research investigating the speed/accuracy tradeoff in motor control tasks [43]–[45]. That said, our findings revealed that attempts to perform the task at the fastest speed likely represented the lower-bound of the speed-accuracy tradeoff for participants. At this extreme,

it is clear that participants prioritized speed over accuracy, thereby diminishing the potential for learning to occur. It is possible that with increased practice learning could have take place. A definitive answer, however, is beyond the scope of this manuscript. While a full analysis of the that tradeoff is not the focus of this manuscript, its impact on our results is nonetheless interesting.

B. Comparison to Existing Literature

While the findings presented here are only applicable to the needle driving task used in the experiment, they provide to the best of the authors' knowledge, the first truly objective comparison of inanimate and VR RAMIS training in terms of skill learning and skill transfer. Several studies have utilized functionally similar tasks to investigate skill learning for inanimate and VR training approaches [25], [27], [28] and have found that both platforms led to post-training performance improvements, with inanimate robot performance slightly outperforming VR performance. Our findings align well with these results in that we observed skill learning for both Inanimate and VR platforms. In addition, numerous studies have investigated the skill transfer process from virtual training to real-world practice [47]–[50]. Results from these studies, however, tend to be variable due to the difficulty of developing Inanimate and VR tasks with equal levels of complexity and methods for performance assessment. Our results also align well with these findings, which in general, found no statistical observations of skill transfer from VR to inanimate practice. Our findings are further supported by the work of Brown *et al.* who investigated skill transfer from inanimate practice to VR performance and found that inanimate practice led to comparable performance on a VR simulator [29]. Finally, the recent work of Wang *et al.* is noteworthy as they used a human-centred approach to investigate differences between VR and inanimate training and found that the training on the live robot was more difficult for participants than training in VR [31]. Given our presented findings, it is worth considering whether this increase in difficulty leads to lack of skill transfer for our VR group.

We chose the task of needle driving as it is a key step in most suturing procedures, and adequately exemplifies the psychomotor and visuomotor skills required of any surgeon. Our Inanimate and Virtual platforms captured the essence of needle driving in real tissue, in particular the objective of following an ideal trajectory from needle insertion to needle exit. In addition, both platforms captured and displayed deviations from this trajectory in an manner consonant with the way real tissue might stretch under such deviations. From a functional perspective, this task fits well with similar inanimate and virtual task used in early psychomotor skill development such as peg transfer. At the same time, the sensorized nature of both platforms fits well with one of growing trends in surgical training research, the use of metrics from instrumented training platforms to produce automated, objective assessments of skill [15], [16], [51]–[53]. That both training platforms highlighted in this manuscript are capable of producing the same metrics expands the potential for the development of robust, standardized approaches to surgical training. Even though our present findings suggest a disparity between inanimate and VR approaches, the data from this study and others may prove useful in creating skill-equivalence mappings that are capable of equating performance between the two approaches.

C. Limitations

Although the results presented in this manuscript have positive implications for RAMIS training, there are few limitations that need to be addressed in future investigations. First, despite the clear relevance to clinical practice, the results discussed in the manuscript have not been validated with a clinical participant pool consisting of novice and expert surgeons. Thus, the face and construct validity of either platform has yet to be established. In addition, while our training platform was designed to measure deviations from the ideal needle driving trajectory, these results lack the validity to predict how performance in our task translates to real tissue. Second, the use of open-sourced hardware and software such as the dVRK and ATAR framework leave open the question of how these results would transfer to commercial telesurgical platforms such as the da Vinci Surgical System and the da Vinci Surgical Simulator. Third, with respect to the study design, we were not able to measure each participants' Baseline performance on both platforms. This information would have produced useful insights for interpreting the skill transfer results. It is also worth considering how these findings might change if participants were given unconstrained time to perform the task. Fourth, regarding the experimental setup, the optomechanical sensors had a limited sensing range causing signal saturation for extreme ring displacements. Likewise, the suturing task and camera angle were only tested in one configuration. Thus it is unclear whether the current results would be sustained in different configurations or if participants could change the orientation themselves. Finally, the ATAR simulation framework, while functional, was not the most realistic simulation. While other more realistic simulation platforms exist, they are generally proprietary, which would limit the ability to design custom training task as we have done here.

VI. Conclusion and Future Work

In this study, we have investigated the equivalency of inanimate and virtual training for robot-assisted minimally invasive surgery (RAMIS). To that end, we developed an enhanced needle driving task (END) with analogous Inanimate and Virtual Reality platforms that were each capable of generating the same objective, robust, and realtime measures of task performance. Utilizing the da Vinci Research Kit (dVRK) telerobotic platform, a user study was conducted to assess participants' performance improvement on a given platform following training (skill learning), as well as the extent to which that improvement was sustained on the opposite platform (skill transfer). In this way, we were able to assess skill development and skill transfer without the typical confounding factors associated with different telerobotic platforms, different training tasks, or different skill performance measures. Our results indicate that the two approaches produce different training outcomes, with inanimate training resulting in more salient

skill transfer than VR training. These findings suggest that training with the real robot using inanimate models remains the gold standard for fundamental psychomotor and visuomotor skill development in robot-assisted minimally invasive surgery (RAMIS).

While the presented work provides empirical evidence of the potential disparity that exists between inanimate and virtual RAMIS training, it has also presented a unique experimental platform that can be used for future investigations. Initial investigations should seek to verify the findings presented here with clinically relevant participants and clinically relevant hardware. This will likely require improving and modifying the Inanimate training platform to work with clinical surgical systems. Such improvements could make use of non-contact optical measurement techniques for tracking task performance. There are also a number of other follow-up experiments that can be conducted to assess the robustness of these results. This includes longitudinal study protocols that use multiple training and post-training sessions to assess any asymptotic behavior of either skill learning or skill transfer. In addition, future work should consider the use of alternative feedback mechanisms such as haptic feedback, and its role in skill development and skill retention.

ACKNOWLEDGMENT

The authors would like to thank: Anton Deguet, for the support during the dVRK setup; Pablo Celnik, for the support during the task and protocol design; Manuel Anaya, for the support during the data collection; Mohit Singhala and Neha Thomas, for the support during the statistical analysis.

REFERENCES

[1] P. Gorphe, "A contemporary review of evidence for transoral robotic surgery in laryngeal cancer," *Front. Oncol.*, vol. 8, p. 121, Apr. 2018.

[2] S. Maeso-Martinez *et al.*, "Efficacy of the da Vinci surgical system in abdominal surgery compared with that of laparoscopy a systematic review and meta-analysis," *Ann. Surg.*, vol. 252, no. 2, pp. 254–262, 2010.

[3] A. L. Smith, K. M. Schneider, and P. D. Berens, "Survey of obstetrics and gynecology residents' training and opinions on robotic surgery," *J. Robot. Surg.*, vol. 4, no. 1, pp. 23–27, 2010. [Online]. Available: https://doi.org/10.1007/s11701-010-0176-0

[4] L. A. McGuinness and B. P. Rai, "Robotics in urology," *Ann. Roy. College Surgeons England*, vol. 100, no. 6s, pp. 45–54, 2018. [Online]. Available: https://doi.org/10.1308/rcsann.supp1.38

[5] R. Satava *et al.*, "Proving the effectiveness of the fundamentals of robotic surgery (FRS) skills curriculum: A single-blinded, multispecialty, multi-institutional randomized control trial," *Ann. Surg.*, to be published.

[6] C. B. Barden, M. C. Specht, M. D. McCarter, J. M. Daly, and T. J. Fahey, III, "Effects of limited work hours on surgical training," *J. Amer. College Surgeons*, vol. 195, no. 4, pp. 531–538, 2002. [Online]. Available: https://doi.org/10.1016/S1072-7515(02)01242-5

[7] J. Y. Lee, P. Mucksavage, D. C. Kerbl, V. B. Huynh, M. Etafy, and E. M. McDougall, "Validation study of a virtual reality robotic simulatorŮrole as an assessment tool?" *J. Urol.*, vol. 187, no. 3, pp. 998–1002, Mar. 2012. [Online]. Available: https://doi.org/10.1016/j.juro.2011.10.160

[8] E. I. George, R. Smith, J. S. Levy, and T. C. Brand, *Simulation in Robotic Surgery BT—Comprehensive Healthcare Simulation: Surgery and Surgical Subspecialties*, D. Stefanidis, J. R. Korndorffer, Jr., and R. Sweet, Eds. Cham, Switzerland: Springer Int., 2019, pp. 191–220. [Online]. Available: https://doi.org/10.1007/978-3-319-98276-2_17

[9] M. D. Weintraub and S. V. Kheyfets, "Training in robotic urologic surgery," in *Robotics in Genitourinary Surgery*, 2nd ed. Cham, Switzerland: Springer Int., Sep. 2018. pp. 163–173.

[10] M. A. Aghazadeh *et al.*, "External validation of global evaluative assessment of robotic skills (GEARS)," *Surg. Endoscopy*, vol. 29, no. 11, pp. 3261–3266, 2015. [Online]. Available: https://doi.org/10.1007/s00464-015-4070-8

[11] A. J. Hung, I. S. Jayaratna, K. Teruya, M. M. Desai, I. S. Gill, and A. C. Goh, "Comparative assessment of three standardized robotic surgery training methods," *BJU Int.*, vol. 112, no. 6, pp. 864–871, 2013. [Online]. Available: https://doi.org/10.1111/bju.12045

[12] C. E. Reiley, H. C. Lin, D. D. Yuh, and G. D. Hager, "Review of methods for objective surgical skill evaluation," *Surg. Endoscopy*, vol. 25, no. 2, pp. 356–366, 2011. [Online]. Available: https://doi.org/10.1007/s00464-010-1190-z

[13] H. C. Lin, I. Shafran, D. Yuh, and G. D. Hager, "Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions," *Comput.-Aided Surg.*, vol. 11, no. 5, pp. 220–230, 2006. [Online]. Available: https://doi.org/10.3109/10929080600989189

[14] K. Liang, Y. Xing, J. Li, S. Wang, A. Li, and J. Li, "Motion control skill assessment based on kinematic analysis of robotic end-effector movements," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 14, no. 1, Feb. 2018, Art. no. e1845. [Online]. Available: https://doi.org/10.1002/rcs.1845

[15] E. D. Gomez, R. Aggarwal, W. McMahan, K. Bark, and K. J. Kuchenbecker, "Objective assessment of robotic surgical skill using instrument contact vibrations," *Surg. Endoscopy*, vol. 30, no. 4, pp. 1419–1431, 2016. [Online]. Available: https://doi.org/10.1007/s00464-015-4346-z

[16] K. Brown, N. Mosley, and J. Tierney, "Battle of the Bots: A comparison of the standard da Vinci and the da Vinci surgical skills simulator in surgical skills acquisition," *J. Robot. Surg.*, vol. 11, no. 2, pp. 159–162, 2017.

[17] *Mimic Technologies, Inc., WA 98104, USA*. Accessed: Mar. 22, 2020. [Online]. Available: https://mimicsimulation.com/dv-trainer/

[18] *3D Systems Healthcare, CO 80127, USA*. Accessed: Mar. 22, 2020. [Online]. Available: https://simbionix.com/simulators/robotix-mentor/

[19] (2019). *Intuitive Surgical, Inc., CA, USA*. [Online]. Available: https://www.intuitive.com/en-us/-/media/Project/Intuitive-surgical/files/pdf/1047821-usra-simnow-leave-behind-r3b.pdf?la=en&hash=12D3DB0F44212FEC5C0E79DCA2564F05

[20] J. S. Shim *et al.*, "Predictive validation of a robotic virtual reality simulator," *Urology*, vol. 122, Dec. 2018, pp. 32–36.

[21] P. Culligan, E. Gurshumov, C. Lewis, J. Priestley, J. Komar, and C. Salamon, "Predictive validity of a training protocol using a robotic surgery simulator," *Female Pelvic Med. Reconstruct. Surg.*, vol. 20, no. 1, pp. 48–51, 2014.

[22] A. J. Hung *et al.*, "Concurrent and predictive validation of a novel robotic surgery simulator: A prospective, randomized study," *J. Urol.*, vol. 187, no. 2, pp. 630–637, Feb. 2012. [Online]. Available: https://doi.org/10.1016/j.juro.2011.09.154

[23] A. I. Tergas, S. B. Sheth, I. C. Green, R. L. Giuntoli, A. D. Winder, and A. N. Fader, "A pilot study of surgical training using a virtual robotic surgery simulator," *J. Soc. Laparoendoscopic Surgeons*, vol. 17, no. 2, pp. 219–226, 2013. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/23925015

[24] R. Korets *et al.*, "Validating the use of the MIMIC dV-trainer for robotic surgery skill acquisition among urology residents," *Urology*, vol. 78, no. 6, pp. 1326–1330, 2011. [Online]. Available: https://doi.org/10.1016/j.urology.2011.07.1426

[25] M. J. Amirian, S. M. Lindner, E. J. Trabulsi, and C. D. Lallas, "Surgical suturing training with virtual reality simulation versus dry lab practice: An evaluation of performance improvement, content, and face validity," *J. Robot. Surg.*, vol. 8, no. 4, pp. 329–335, 2014.

[26] M. A. Lerner, M. Ayalew, W. J. Peine, and C. P. Sundaram, "Does training on a virtual reality robotic simulator improve performance on the da Vinci surgical system?" *J. Endourol.*, vol. 24, no. 3, pp. 467–472, 2010.

[27] C. Perrenot *et al.*, "The virtual reality simulator dV-trainer is a valid assessment tool for robotic surgical skills," *Surg. Endoscopy*, vol. 26, no. 9, pp. 2587–2593, 2012.

[28] L. K. Newcomb *et al.*, "Correlation of virtual reality simulation and dry lab robotic technical skills," *J. Minimally Invasive Gynecol.*, vol. 25, no. 4, pp. 689–696, 2018.

[29] J. D. Brown, C. E. O'Brien, S. C. Leung, K. R. Dumon, D. I. Lee, and K. J. Kuchenbecker, "Using contact forces and robot arm accelerations to automatically rate surgeon skill at peg transfer," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 9, pp. 2263–2275, Sep. 2017.

[30] R. Smith, V. Patel, and R. Satava, "Fundamentals of robotic surgery: A course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 10, no. 3, pp. 379–384, 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/rcs.1559

[31] Z. Wang *et al.* (Mar. 2020). *A Comparative Human-Centric Analysis of Virtual Reality and Dry Lab Training Tasks on the da Vinci Surgical Platform*. [Online]. Available: https://doi.org/10.1142/S2424905X19420078%5C

[32] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, "An open-source research kit for the da Vinci surgical system," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2014, pp. 6434–6439.

[33] A. Mariani, E. Pellegrini, N. Enayati, P. Kazanzides, M. Vidotto, and E. D. Momi, "Design and evaluation of a performance-based adaptive curriculum for robotic surgical training: A pilot study," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2018, pp. 2162–2165.

[34] N. Enayati *et al.*, "Robotic assistance-as-needed for enhanced visuomotor learning in surgical robotics training: An experimental study," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, 2018, pp. 6631–6636.

[35] L. Qian, A. Deguet, and P. Kazanzides, "dVRK-XR: Mixed reality extension for da Vinci research kit," in *Proc. 12th Hamlyn Symp. Med. Robot. Imperial College*, 2019, pp. 93–94.

[36] M. Quigley *et al.*, "ROS: An open-source robot operating system," in *Proc. ICRA Workshop Open Source Softw.*, 2009, p. 5.

[37] Intuitive Surgical. *EndoWrist Instruments*. Accessed: Mar. 22, 2020. [Online]. Available: https://www.intuitivesurgical.com/products/instruments/

[38] A. Industries. *Adafruit Blinka Library*. Accessed: Mar. 22, 2020. [Online]. Available: https://pypi.org/project/Adafruit-Blinka/

[39] P. S. Foundation. *Python Library*. Accessed: Mar. 22, 2020. [Online]. Available: https://www.python.org/

[40] N. Enayati, A. Mariani, E. Pellegrini, T. Chupin, G. Ferrigno, and E. D. Momi, "A framework for assisted tele-operation with augmented reality," in *Proc. CRAS Joint Workshop New Technol. Comput. Robot Assist. Surg.*, 2017, p. 107.

[41] K. Mamou and F. Ghorbel, "A simple and efficient approach for 3D mesh approximate convex decomposition," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, 2009, pp. 3501–3504.

[42] E. Coumans. (2013). *Bullet Physics Library*. [Online]. Available: http://bulletphysics.org/

[43] Y. Itaguchi and K. Fukuzawa, "Influence of speed and accuracy constraints on motor learning for a trajectory-based movement," *J. Motor Behav.*, vol. 50, pp. 1–11, Nov. 2017.

[44] R. P. Heitz, "The speed-accuracy tradeoff: History, physiology, methodology, and behavior," *Front. Neurosci.*, vol. 8, p. 150, Jun. 2014. [Online]. Available: https://www.ncbi.nlm.nih.gov/pubmed/24966810

[45] J. H. Chien *et al.*, "Accuracy and speed trade-off in robot-assisted surgery," *Int. J. Med. Robot. Comput. Assist. Surg.*, vol. 6, no. 3, pp. 324–329, 2010. [Online]. Available: https://doi.org/10.1002/rcs.336

[46] J. Reis *et al.*, "Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 5, pp. 1590–1595, 2009.

[47] M. V. Vargas *et al.*, "Transferability of virtual reality, simulation-based, robotic suturing skills to a live porcine model in novice surgeons: A single-blind randomized controlled trial," *J. Minimally Invasive Gynecol.*, vol. 24, no. 3, pp. 420–425, 2017, doi: 10.1016/j.jmig.2016.12.016.

[48] J. Hoogenes *et al.*, "A randomized comparison of 2 robotic virtual reality simulators and evaluation of trainees' skills transfer to a simulated robotic urethrovesical anastomosis task," *Urology*, vol. 111, pp. 110–115, Jan. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0090429517310567

[49] S. V. Whitehurst *et al.*, "Comparison of two simulation systems to support robotic-assisted surgical training: A pilot study (swine model)," *J. Minimally Invasive Gynecol.*, vol. 22, no. 3, pp. 483–488, 2015, doi: 10.1016/j.jmig.2014.12.160.

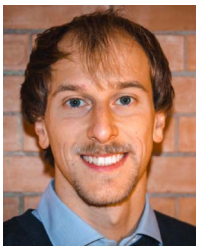
[50] C. M. Vaccaro, C. C. Crisp, A. N. Fellner, C. Jackson, S. D. Kleeman, and J. Pavelka, "Robotic virtual reality simulation plus standard robotic orientation versus standard robotic orientation alone: A randomized controlled trial," *Female Pelvic Med. Reconstruct. Surg.*, vol. 19, no. 5, pp. 266–270, 2013.

[51] T. J. Tausch, T. M. Kowalewski, L. W. White, P. S. McDonough, T. C. Brand, and T. S. Lendvay, "Content and construct validation of a robotic surgery curriculum using an electromagnetic instrument tracker," *J. Urol.*, vol. 188, no. 3, pp. 919–923, 2012, doi: 10.1016/j.juro.2012.05.005.

[52] T. S. Lendvay *et al.*, "Virtual reality robotic surgery warm-up improves task performance in a dry laboratory environment: A prospective randomized controlled study," *J. Amer. College Surgeons*, vol. 216, no. 6, pp. 1181–1192, 2013, doi: 10.1016/j.jamcollsurg.2013.02.012.

[53] S.-K. Jun, M. S. Narayanan, A. Eddib, P. Singhal, S. Garimella, and V. Krovi, "Minimally invasive surgical skill assessment by video-motion analysis," in *Proc. 5th Hamlyn Symp. Med. Robot.*, 2012, p. 2.

Guido Caccianiga received the B.Sc. degree in mechanical engineering and the M.Sc. degree in biomedical engineering from the Politecnico di Milano in 2016 and 2019, respectively. He was a Visiting Scholar with the Laboratory for Computational Sensing and Robotics, Johns Hopkins University, Baltimore, MD, USA. He is currently a Research Engineering Intern with the Electronic Information and Bioengineering Department, Politecnico di Milano. He has been selected as a 2020 Ph.D. student with the International Max Planck Research School for Intelligent Systems, Stuttgart. His research focuses on the integration of hardware and software solutions for robot-assisted surgery with specific mention to sensorization, haptics, virtual reality, and user interface technologies.



Andrea Mariani received the B.Sc. and M.Sc. degrees (Hons.) in biomedical engineering from Politecnico di Milano in 2015 and 2018, respectively, and the Alta Scuola Politecnica Diploma degree (Hons.) in 2018. He is currently pursuing the Ph.D. degree in biorobotics with Scuola Superiore Sant'Anna, Pisa. He attended a visiting period at Johns Hopkins University, Baltimore, MD, USA. His research focuses on simulation-based training for robot-assisted surgery and high-intensity ultrasound therapeutic technologies. He has been involved in international projects, including an Intuitive Technology Research Grant, FUTURA2020, and the 2020 KUKA Innovation Awards.

Elena De Momi (Senior Member, IEEE) received the M.Sc. degree in biomedical engineering and the Ph.D. degree in bioengineering from Politecnico di Milano in 2002 and 2006, respectively, where she is currently an Associate Professor with the Electronic Information and Bioengineering Department. She is a Co-Founder of the Neuroengineering and Medical Robotics Laboratory, in 2008, being responsible of the Medical Robotics Section. She participated in several EU funded projects in the field of surgical robotics (ROBOCAST, ACTIVE, and EuRoSurge, where she was a PI for partner POLIMI). She is currently a PI for POLIMI of the EDEN2020 Project, aimed at developing a neurosurgery drug delivery system and of the ATLAS MSCA-ITN-2018-EJD, and a Coordinator of the MSCA-IF-2017—Individual Fellowships. She has been an Evaluator and a Reviewer for the European Commission in FP6, FP7, and H2020. Her academic interests include computer vision and image processing, artificial intelligence, augmented reality and simulators, teleoperation, haptics, medical robotics, and human–robot interaction. She is currently an Associate Editor of the *Journal of Medical Robotics Research*, the *International Journal of Advanced Robotic Systems*, *Frontiers in Robotics and AI*, and *Medical & Biological Engineering & Computing*. Since 2016, she has been an Associate Editor of IEEE ICRA, IROS, and BioRob, and an Area Chair of MICCAI. She is currently a Publication Co-Chair of ICRA 2019.

Gabriela Cantarero received the Ph.D. degree in neuroscience from the Johns Hopkins School of Medicine and postdoctoral training with the Walter Reed Army Institute of Research. She is an Assistant Professor with the Johns Hopkins Department of Physical Medicine and Rehabilitation. She is a Co-Director for the Center for Movement Studies and a Scientific Director of the Non-Invasive Brain Stimulation Clinical Program. Her current work focuses on studying the effects of brain stimulation and behavior on motor neurophysiology in humans, understanding the neurophysiological mechanisms underlying human motor learning in both healthy individuals and those after damage to the brain, and optimizing motor learning and recovery using noninvasive brain stimulation techniques.

Jeremy D. Brown (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in mechanical engineering from the University of Michigan at Ann Arbor, MI, USA, in 2008, 2012, and 2014, respectively, and the B.Sc. degree in applied physics from Morehouse College, Atlanta, GA, USA. He completed a Postdoctoral Research Fellowship with the University of Pennsylvania, Philadelphia, PA, USA, from 2014 to 2016. He is currently a John C. Malone Assistant Professor with the Department of Mechanical Engineering, Johns Hopkins University, Baltimore, MD, USA. His research focuses on the interface between humans and robots with a specific focus on medical applications and haptic feedback. He directs the Haptics and Medical Robotics Laboratory, which is part of the Laboratory for Computational Sensing and Robotics. He received several awards, including the National Science Foundation CRII Award, the National Science Foundation Graduate Research Fellowship, the Best Student Paper Award from the IEEE Haptics Symposium in 2012, and the Penn Postdoctoral Fellowship for Academic Diversity.