Design and Evaluation of a Performance-based Adaptive Curriculum for Robotic Surgical Training: a Pilot Study *

Andrea Mariani ◆ [1], Edoardo Pellegrini ◆ [1], Nima Enayati[1], Peter Kazanzides[2], Marco Vidotto[1] and Elena De Momi[1]

Abstract*— Training with simulation systems has become a primary alternative for learning the fundamental skills of robotic surgery. However, there exists no consensus regarding a standard training curriculum: sessions defined a priori by expert trainers or self-directed by the trainees feature lack of consistency. This study proposes an adaptive approach that structures the curriculum on the basis of an objective assessment of the trainee's performance. The work comprised an experimental session with 12 participants performing training on virtual reality tasks with the da Vinci Research Kit surgical console. Half of the subjects self-managed their training session, while the others underwent the adaptive training. The final performance of the latter trainees was found to be higher compared to the former (p=0.002), showing how outcome-based, dynamic designs could constitute a promising advance in robotic surgical training.

I. Introduction

Robot-Assisted Minimally Invasive Surgery (RAMIS) has become adopted in a wide area of surgical interventions including urology, gynecology and general abdominal surgery [1]. This technique results in a surgical approach that is easier for surgeons and safer for patients [2]. Nevertheless, RAMIS implies new control modalities, where the user has to manipulate a couple of robotic masters in order to position the surgical instruments inside the patient. This teleoperative approach detrimentally involves the need to learn how to deal with the new dynamics of the manipulators [3] and how to compensate for the absence of haptic feedback [4]. Together with the complexity of surgery as a sensorimotor task, these drawbacks lead to a distinct necessity of efficient training modalities.

Ideally, a training platform has to transfer the required surgical skills as efficiently and safely as possible, together with allowing the learning assessment of the trainee. Simulation-based education for the development of general surgical skills has come to the forefront in recent years. In fact, its features are well matched to the requirements previously listed. Firstly, simulators allow to frame realistic, versatile training environments where the trainee can gain technical skills that have been shown to be directly transferable to the patient-based settings [5]. Secondly, if robotic platforms generally provide access to motion data that can be exploited to assess the trainee's proficiency [2], simulators add the possibility of objective performance evaluations by measurement of intra-task metrics [6]. Finally, simulation-based training allows trainees to learn the required skills without risking the patient safety and in a cost-efficient way [7].

The exploitation of simulators does not cancel the need for defining a training curriculum. Several alternatives have been proposed in recent years [8][9] but there exists no consensus regarding a standard curriculum. Additionally, simulators have shifted the paradigm of supervised learning towards a self-directed training, with the consequent reduction of program costs [10]. Nevertheless, the success of this approach relies on the ability of the trainee to correctly perform self-assessment. Especially in the early stages of training, this condition is not totally accomplished [11]. In order to overcome the restrictions of a static schedule and the lack of consistency of a self-managed training, adaptive approaches have been proposed in other educational fields [12]. Adaptive training can be defined as a training modality in which the task is varied as a function of how well the trainee performs [13]. This adaptivity should also increase the training efficiency since effective learning takes place only when training occurs at an appropriate level of challenge for the trainee [14].

Our study proposes an adaptive training curriculum for robotic surgery that automatically schedules the training session on the basis of an objective assessment of the trainee's performances. The research question can be summarized as follows: *can an adaptive training achieve better final performances compared to a self-managed training?*

II. Methods

A. Experimental Setup

The robotic platform employed in this study was the surgeon console of a da Vinci Research Kit (dVRK, Intuitive Surgical, CA, USA). This console includes a foot-pedal tray, a stereo viewer and two master manipulators.

The VR simulation environment was designed using our recently developed Assisted Teleoperation with Augmented Reality (ATAR) framework [15]. This framework is built using Robot Operating System (ROS) nodes. The core simulation node generates the graphics and physics of the virtual objects, along with the task logic. The virtual tools are

* Research partially supported by an Intuitive Surgical Technology Research Grant (to N. Enayati and E. De Momi) and by the European Union's Horizon 2020 research and innovation program under grant agreement No 732515.

◆ These authors contributed equally to the work.

1. Andrea Mariani, Edoardo Pellegrini, Nima Enayati, Marco Vidotto and Elena De Momi are with the Department of Electronics, Information and Bioengineering, Politecnico di Milano, 20133, Italy (e-mail: andrea20.mariani@mail.polimi.it, edoardo.pellegrini@mail.polimi.it, nima.enayati@polimi.it, marco.vidotto@polimi.it and elena.demomi@polimi.it).

2. Peter Kazanzides is with the Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA (e-mail: pkaz@jhu.edu).

978-1-5386-3646-6/18/$31.00 ©2018 IEEE

simulated as kinematic objects and their pose is constrained to that of the master devices by the node which is responsible for the teleoperation. Through a GUI node, the operator can select the task to perform, as well as control the recording of the current session data (which is performed at 30 Hz).

B. Training Curriculum

Exploiting this setup, we developed a curriculum composed of elementary and complex tasks. The former aim at training a single fundamental skill of robotic surgery, while the latter involve multiple skills at once. The starting point was a complex bimanual visuo-motor task that we analyzed in a previous study [16]. This is a classic steady-hand game often included in surgical robotics training curricula. This task was then broken down into its essential components. In particular, this choice was guided by the analysis of a skill deconstruction list generated by robotic operations observation and interviews with experts [17]. Depth perception, hand-eye coordination, wrist articulation and object transfer were identified as the elementary skills involved in the steady hand exercise and four elementary tasks were developed in order to train each skill respectively. A full description of skills and tasks is reported in Table I.

C. Performance Measurement

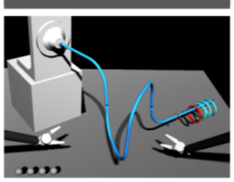
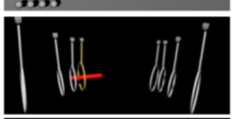
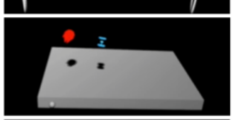
The performances were objectively assessed on the basis of task metrics (Table I, right side). All the metrics were expressed as values ranging from 0 to 1. To achieve this goal, a metric m was normalized and saturated as:

$$m^* = sat(0, \frac{m_{best} - m}{m_{best} - m_{worst}}, 1) \quad (1)$$

where $sat(a, x, b)$ saturates the input x between the minimum a and the maximum b. The ideal value of each metric m_{best} was determined by taking into consideration the performances of the authors (regarded as experts after the long practice on each task), while the lower boundary m_{worst} was tuned according to the novice users' data. Finally, the performance P_i in each elementary task i was computed as a weighted sum of the associated metrics; as far as it concerns the complex task, both an evaluation S_i of each involved skill i (using again a weighted sum) and of the overall performance P_{tot} (as the average of the four skill performances) were derived. These weights are reported in Table I (right side).

D. Acquisition Protocol

In order to test the research hypothesis, a user study was carried out with volunteer users at the Johns Hopkins University (Baltimore, US). The study population consisted in 12 non-medical participants (aged between 22 and 41 years old, 9 males and 3 females, all right-handed) with none to little experience with robotic teleoperation. The study was approved by the JHU Homewood Institutional Review Board (HIRB00000701). The users were randomly divided into two groups: a *control group*, performing the self-managed training, and an *experimental group*, undergoing the adaptive training. The total time was selected as a training constraint to have comparable protocols between the two groups. In particular, this pilot study focused on a single session characterized by three parts. Firstly, the subjects were introduced to the dVRK console, the experimental protocol and the final evaluation modalities, as well as they were shown videos of a successful execution of each task; they also performed two repetitions of the complex task in order to assess that all the trainees initially belonged to the same statistical population in terms of performances. Secondly, the users underwent 45 minutes of training: the ones belonging to the control group directly chose their exercises and they were provided with their percentage

TYPE	PICTURE	NAME	DESCRIPTION		METRIC WEIGHTS				
					m_1	m_2	m_3	m_4	m_5
COMPLEX		STEADY HAND TASK	A user moves a ring along a curved wire pathway, while attempting to avoid the ring and wire making contact and keeping the ring's plane perpendicular to the wire's tangent. In each repetition, the subject has to grasp a ring from the right side of the wire with the right hand tool, carry it to the middle of the wire and transfer the ring to the left tool and finally carry the ring to the left end of the wire.	DEPTH PERCEPTION	7	1	1	1	0
				HAND-EYE COORDINATION	1	2	0	6	1
				WRIST ARTICULATION	1	0	9	0	0
				OBJECT TRANSFER	0	3	0	0	7
ELEMENTARY		DEPTH PERCEPTION	*Capability to perceive the depth of the objects in the scene and their position in perspective*	The user controls a cylinder that moves in the space. The purpose is to insert the cylinder sequentially into rings placed at different depth levels.	7	2	0	1	0
		HAND-EYE COORDINATION	*Ability to position the tools and interact with objects accurately*	The task consists in controlling the position of an object and keeping it as close as possible to a target that moves along a random path in the space.	1	2	0	7	0
		WRIST ARTICULATION	*Dexterity in orienting the tools in the space and align them with a specific direction*	The user controls the orientation of one virtual hand and he has to rotate his wrist in order to keep it overlapped to a second, concentric and randomly rotating hand.	2	0	8	0	0
		OBJECT TRANSFER	*Proficiency in transferring the grip of an object between the tools*	The subject has to grasp a ring from the right tower and to place it on the target tower after performing an instrument-to-instrument exchange.	2	1	0	0	7

TASK METRICS LEGEND

INDEX	NAME	DESCRIPTION	UNIT
m_1	DEPTH ERROR	Distance between tool and target along the perspective direction	*mm*
m_2	TIME	Time needed to complete the task	*sec*
m_3	POSITION ERROR	Linear distance between the actual and the ideal position of the manipulator	*mm*
m_4	ORIENTATION ERROR	Angular difference between the actual and ideal pose of the manipulator	*rad*
m_5	EXCHANGE ERROR	Number of drops while performing an object transfer	/

Table I. Tasks composing the training curriculum (left) and associated metrics (right). A detailed description of complex (blue) and elementary (red) tasks is provided. The weights associated to each metric (defined in the gray table) in the different tasks are used to compute the user's performance. They range between 0 and 10, and their sum is equal to 10 for each task.

performance after each task repetition; the experimental group was provided with the task to perform according to the adaptive algorithm (described in the following paragraph). Finally, all the subjects went through a final test, consisting of 4 repetitions of the complex task. The adaptive logic automatically modulated the training curriculum of the experimental group as a function of the performances. The training session of these users was split into 3 units. These were defined as 15 minutes training slots, where the user performed one repetition of the complex task at the beginning and then elementary tasks till the end of that unit. Once the elementary training started, the adaptive algorithm had to output the elementary task that the user should perform in order to fill the gaps in the different surgical skills. The proposed method aimed at scheduling the training on the basis of a *priority index* (φ_i, i=skill index={1,2,3,4}) that took into account the user's initial performance in the complex task, together with the influence of the on-going elementary training. In a formal way,

$$\bar{\imath} = elementary\ task\ to\ perform = argmax(\varphi_i). \quad (2)$$

To initialize this index, the performances in the first complex task repetition were exploited:

$$\varphi_i = 1 - S_i \quad \in [0,1] \quad (3)$$

where

$$S_i = \frac{\sum \alpha_j m_j}{\sum \alpha_j} \in [0,1] \quad (4)$$

is the i^{th} skill evaluation in the complex task and α_j is the weight associated to the j^{th} metric m_j in the complex task. Once the first selected elementary task was performed, in order to update the priority index, the result in the elementary training was considered as well as the correlation of the performed elementary task with other skills and with the complex task itself. To achieve this goal, we introduced the concepts of *skill connection* and *structural significance* respectively. These were quantified by computing the correlation of metric weights of the two tasks under analysis. The *skill connection* was defined as:

$$\rho_{ij} = \frac{\sum_m \beta_m^i \beta_m^j}{\sqrt{\sum_m {\beta_m^i}^2 \sum_m {\beta_m^j}^2}} \in [0,1] \quad (5)$$

where β_m^i is the weight associated to the m^{th} metric in the i^{th} elementary task, while β_m^j is the weight associated to the m^{th} metric in the j^{th} elementary task. As far as it concerned the *structural significance*,

$$\sigma_i = \frac{\sum_m \alpha_m^i \beta_m^i}{\sqrt{\sum_m {\alpha_m^i}^2 \sum_m {\beta_m^i}^2}} \in [0,1] \quad (6)$$

where α_m^i is the weight associated to the m^{th} metric in the complex task to evaluate the i^{th} skill and β_m^i is the weight associated to the m^{th} metric in the i^{th} elementary task. In the end, the priority index was updated as

$$\varphi_i^{new} = \varphi_i^{old}\left(1 - \sigma_i\, \rho_{ij}\, P_j\right) \in [0,1] \quad (7)$$

where j is the index of the elementary task that the user performed last time and P_j the associated performance.

E. *Result Analysis*

Due to the small sample size, non-parametric statistical significance tests were used to compare the training effects on the two groups. The Wilcoxon rank sum test was employed with the performance measures as dependent variables and training groups as independent factors. Statistically significant effects were assessed at $p < 0.05$. The statistical analysis was performed in MATLAB using the command *ranksum()*.

III. Results and Discussion

Figure 1 shows the task composition of the training session of each group, averaged across all the participants. The experimental users performed a higher number of tasks during the training compared to the control group (24 vs. 13, p=0.009). This can be explained considering that those subjects had just to perform the task the adaptive algorithm selected for them, while the users belonging to the control group had to think about which task to execute. Considering the time constraints related to the usage of simulators by medical residents, these results highlight that an automatic schedule of the training allows to take full advantage of the available time with the device.

Moving to the global performances in the complex task before and after training, Figure 2.a presents the results of the two groups. Both the groups were characterized by an increase of performances and a reduction of their variability. The initial baseline assessment resulted in the absence of statistically significant difference between the control and experimental users (p=0.94). On the other hand, the analysis of the final outcomes led to the conclusion that the adaptive training allowed to achieve higher final performances (0.65 vs. 0.48, p=0.002). Additionally, there was a statistically significant improvement in the experimental group after training (p=0.002), while the control group featured a non-significant learning (p=0.31).

In order to better understand these outcomes, the final skill

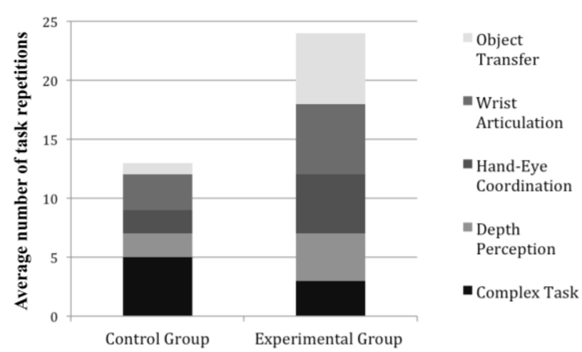


Figure 1. Histograms of the training session composition of the control (left) and experimental (right) group.

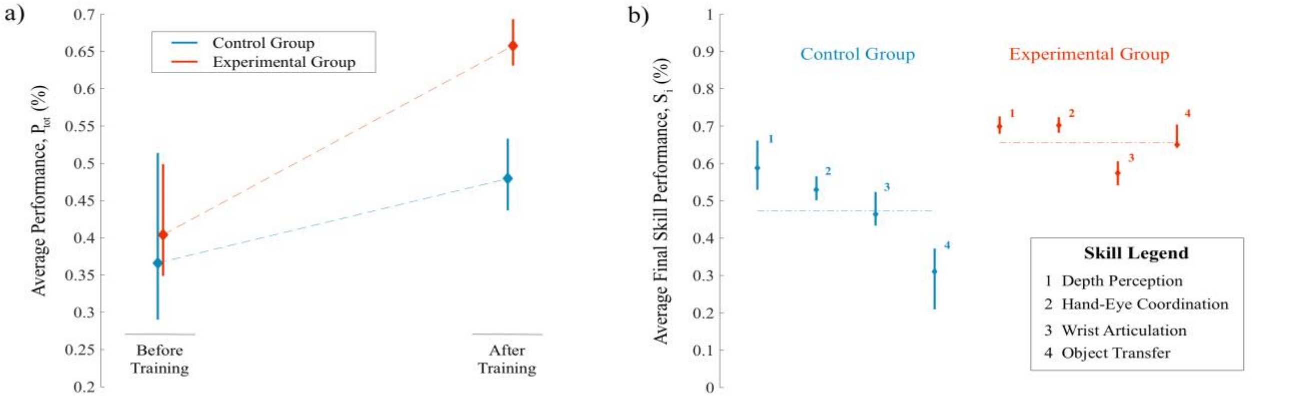


Figure 2. [a, left] The average performances across the initial (before training) and final (after training) complex task repetitions are displayed for the control group (blue) and experimental group (red). The diamond markers represent the median, while the vertical bars show the 25th and 75th percentiles. The dotted line graphically presents the learning rate. [b, right] The graph shows the average skill performances across the final (after training) complex task repetitions of the two groups. Markers and bars have the same meaning, while the dotted line is in correspondence of the median overall performance.

performances were analyzed (Figure 2.b). The performances were significantly higher in the experimental group across all the elementary skills (depth perception, p=0.04; hand-eye coordination, p=0.004; wrist articulation, p=0.009; object transfer, p=0.002). Furthermore, the users who underwent the adaptive training showed lower variability of final performances both intra skill and inter skills. These results can be interpreted by assuming that the adaptive training achieved its goal of structuring the curriculum to fill the subject's gaps in the each skill. On the other side, the control group presented the biggest final gap in the object transfer; if we go back to their training session histogram (Figure 1), the corresponding elementary task was the one that the users performed the least, showing a failure in their self-management of the training session. At the same time, the higher final performances of the experimental users can be linked to the reduction of the subject's cognitive load during training, when the subject could focus just on the exercise and not also on the choice of the task.

IV. Conclusions and Future Work

Analyzing simulation-based training in robot-assisted surgery, this study compared a user-managed approach to an adaptive training where the curriculum was automatically and dynamically structured on the basis of the trainee's performances. The results were promising, showing that the final performances of the latter trainees were significantly higher and their skill learning featured more uniformity and higher end points compared to the former. The research focused just on the initial part of the trainees' learning curve and longer or multiple session experiments should be performed. At the same time, a larger sample size could lead to a better statistical determinism. Finally, further analyses should involve a third acquisition group which performs a training supervised by an expert in order to assess the efficacy of the adaptive algorithm with respect to a subjective coaching.

References

[1] S. Maeso, M. Reza, J. A. Mayol 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.* 252: 254-262, 2010.

[2] K. Moorthy, Y. Munz, A. Dosis, J. Hernandez, S. Martin, F. Bello, and A. Darzi, *Dexterity enhancement with robotic surgery*, *Surgical Endoscopy and Other Interventional Techniques*, vol. 18, no. 5, pp. 790-795, 2004.

[3] I. Nisky, M. H. Hsieh, and A. M. Okamura, *Uncontrolled manifold analysis of arm joint angle variability during robotic teleoperation and freehand movement of surgeons and novices*, *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 12, pp. 2869-2881, 2014.

[4] A. M. Okamura, *Haptic feedback in robot-assisted minimally invasive surgery*, *Current Opinion in Urology*, vol. 19, no. 1, pp. 102-107, 2009.

[5] S. R. Dawe, J. A. Windsor, J. A. Broeders, P. C. Cregan, P. J. Hewett, G. J. Maddern, *Systematic review of skills transfer after surgical simulation-based training: laparoscopic cholecystectomy and endoscopy*, *BR J Surg;* 101:1063–76, 2014.

[6] N. Enayati, G. Ferrigno, and E. De Momi, *Skill-based human–robot cooperation in tele-operated path tracking*, *Autonomous Robots*, 2017.

[7] D. M. Gaba , *The future vision of simulation in health care*, *Quality & Safety in Health Care* 13(suppl. 1):i2–i10, 2004.

[8] R. Smith, V. Patel, 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*; 10(3): 379–84, 2014.

[9] Chitwood, W. R., Nifong, L. W., Chapman, W. H. H., Felger, J. E., Bailey, B. M., Ballint, T., Albrecht, R. A., *Robotic Surgical Training in an Academic Institution*, *Annals of Surgery*, 234(4), 475–486, 2001.

[10] J. MacDonald, et al., *Self-assessment in simulation-based surgical skills training, The American Journal of Surgery*, 185(4), 319 – 322, 2003.

[11] D. A. Davis, P. E. Mazmanian, M. Fordis, R. Van Harrison, K. E. Thorpe, L. Perrier, *Accuracy of Physician Self-assessment Compared With Observed Measures of Competence: A Systematic Review*. *JAMA*; 296(9): 1094–1102, 2006.

[12] C. Shih, H. Y. D. Lu, L. Sun, J.-T. Huang, J. Packard, *An adaptive training program for tone acquisition*. In Speech Prosody, 100981, Chicago, pp. 1–4, 2010.

[13] C. R. Kelley, *What is adaptive training?*, *Human Factors*, 11(6), 547–556., 1969.

[14] M. A. Guadagnoli, D. L. Timothy, *Challenge Point: A Framework for Conceptualizing the Effects of Various Practice Conditions in Motor Learning, Journal of Motor Behavior*; 36(2): 212–24, 2004.

[15] N. Enayati, A. Mariani, E. Pellegrini, T. Chupin, G. Ferrigno, and E. De Momi, "A Framework for Assisted Tele-operation with Augmented Reality," in *CRAS: Joint Workshop on New Technologies for Computer/Robot Assisted Surgery*, 2017.

[16] N. Enayati, A. M. Okamura, A. Mariani, E. Pellegrini, M. Coad, G. Ferrigno, E. De Momi, *Robotic Assistance-as-Needed for Enhanced Visuomotor Learning in Surgical Robotics Training: An Experimental Study* , (Accepted to) ICRA 2018.

[17] G. Dulan, R. V. Rege, D. C. Hogg et al., *Developing a comprehensive, proficiencybased training program for robotic surgery*, *Surgery*;152:477–88, 2012.