Assessment of Laparoscopic Skills Based on Force and Motion Parameters

Tim Horeman*, Jenny Dankelman, Frank Willem Jansen, and John J. van den Dobbelsteen

***Abstract*—Box trainers equipped with sensors may help in acquiring objective information about a trainee's performance while performing training tasks with real instruments. The main aim of this study is to investigate the added value of force parameters with respect to commonly used motion and time parameters such as path length, motion volume, and task time. Two new dynamic bimanual positioning tasks were developed that not only requiring adequate motion control but also appropriate force control successful completion. Force and motion data for these tasks were studied for three groups of participants with different experience levels in laparoscopy (i.e., 11 novices, 19 intermediates, and 12 experts). In total, 10 of the 13 parameters showed a significant difference between groups. When the data from the significant motion, time, and force parameters are used for classification, it is possible to identify the skills level of the participants with 100% accuracy. Furthermore, the force parameters of many individuals in the intermediate group exceeded the maximum values in the novice and expert group. The relatively high forces used by the intermediates argue for the inclusion of training and assessment of force application during tissue handling in future laparoscopic skills training programs.**

***Index Terms*—Box trainers, force and motion surgical trainer (ForMoST), force feedback, laparoscopy, objective assessment, training methods.**

I. Introduction

A. Training in Laparoscopic Surgery

In laparoscopic procedures, the moment arm between incision point and instrument fluctuates during the procedure as the force exerted by the tip on the tissue depends highly on the insertion depth of the instruments. These and other instrument handling difficulties make it essential to train this type of surgical skills before the approach is used in the operation theatre. One common training method for laparoscopic surgery is facilitated by a so-called "box trainer." In a box trainer, one can perform training tasks using real laparoscopic instruments (see Fig. 1). Today, most of the available training tasks focus on improving the trainee's eye–hand coordination [1]. Assessment of the trainee's performance can be either subjective when based on the interpretation of the tutor or objective when quantitative measures are used. Objective scoring methods can be based on time, errors, instrument motions, or forces exerted on the instruments or training task. In many studies, assessment is based on task errors and task time [2]. In other studies, task time in combination with assessment parameters extracted from instrument motions are used for discrimination between experts, intermediates, and novices [3]. In the study of Rosen *et al*. and our own previous study, it was found that assessment based on interaction force between instrument tip and environment alone gives similar results [4]–[6].

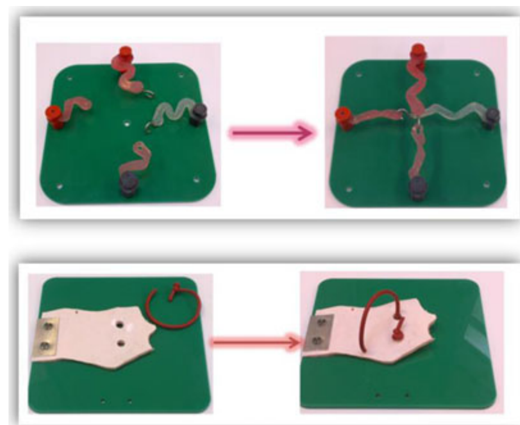

Fig. 1. Top: connection of two vertical and two horizontal flexible elements in Task 1. Bottom: placement of a silicone wire through two holes in Task 2.

All aforementioned studies indicate that skills assessment based on motion, force information, or performance time is possible. Therefore, the question arises whether a multitude of sensor systems has added value when the discriminating power of force, motion, and time parameters is comparable. The studies of Chamarra *et al*. [3] showed that the correlation of motion parameters such as time, path length (PL), and motion smoothness is high. This indicates that fast performance on a training task likely results in a good motion parameter score which is in line with the opinion of some experts that measuring task time is sufficient. However, our previous study regarding assessment based on force parameters shows that task time is not representative for force application skills [5], [7].

Considering that force parameters are, other than motion parameters, indicators for tissue damage and therefore patient

Manuscript received August 1, 2013; revised October 1, 2013; accepted November 2, 2013. Date of publication November 7, 2013; date of current version February 14, 2014. *Asterisk indicates corresponding author*.

*T. Horeman is with the Department of Biomechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft 2628 CD, Zuid Holland, The Netherlands (e-mail: t.horeman@tudelft.nl).

J. Dankelman and J. J. van den Dobbelsteen are with the Department of Biomechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, Delft 2628 CD, Zuid Holland, The Netherlands (e-mail: j.dankelman@tudelft.nl; J.J.vandenDobbelsteen@tudelft.nl).

F. W. Jansen is with the Department of Gynecology, Leiden University Medical Center, 2300 RC Leiden, Zuid Holland, The Netherlands (e-mail: F.W.Jansen@LUMC.nl).

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

Digital Object Identifier 10.1109/TBME.2013.2290052

safety, monitoring the presence of dangerous excessive forces during training is recommendable [8]. Although force parameters are not correlated to task time, it is possible that force parameters are correlated to some motion parameters making the measurement of forces obsolete. To determine whether concurrent measurements of force and motion has added value for the assessment of laparoscopic skills we studied time, force, and motion parameters in training tasks that represent tissue manipulation in surgery.

In order to find differences in tissue and instrument handling between groups with different skills levels, new training tasks are required particularly for training of instrument motion during tissue manipulation. Those standardized tasks should combine the strong aspects of the existing tasks (i.e., delicate position control) and require active control of two hands. Moreover, to mimic real *in-vivo* situations, force control should be part of the training as well as technical insight of the instrument actions necessary to complete the task efficiently.

B. Aims of This Study

The first goal is to identify whether motion parameters are correlated to force parameters. The second goal is to determine whether a combination of force, time, and motion parameters can be used for classification of the skills of a trainee.

II. Materials and Methods

A. Participants

Based on the number of available residents and supervising surgeons in the educational courses, 42 participants with three different levels of experience in laparoscopy took part in this study. The number of participants is similar to the number used during comparable studies [3], [5]. The expert group ($n = 11$) consisted of surgeons and gynecologists that performed over 100 laparoscopic procedures. The intermediates group ($n = 19$) consisted of residents during their specialization in gynecology. All of them succeeded one or more laparoscopic training sessions in eye–hand coordination. The group of novices ($n = 12$) consisted of first- and second-year medical students with no experience in laparoscopic surgery or laparoscopic training. Each participant was asked to answer a short questionnaire with detailed information about prior experience in laparoscopy. All of the participants were right handed.

B. Two New Dynamic Position Tasks

One of the standards that is commonly used to train basic eye–hand coordination in laparoscopy is the fundamental laparoscopic skills (FLS) skills testing system. For this testing system, five tasks (peg transfer, pattern cutting, endoloop, extracorporeal, and intracorporeal suturing) are used for skills assessment [9], [10]. Although those tasks are proven effective for training of technical skills they focus mainly on laparoscopic eye–hand coordination rather than tissue manipulation skills. Since tissue manipulation is one of the most important issues in surgery, objective structured assessment of technical skills have been developed to assess the trainee in his performance in daily practice [11]. In order to measure safe tissue handling in box trainers however, additional tasks and performance metrics need to be developed. In this study, we developed tasks that focus on efficient and well-controlled bimanual tissue handling. We used elastic materials that mimic tissues inside the abdomen (see Fig. 1).

Task 1—Tissue attachment under traction: This task is made from four different elastic elements with different elastic properties. All elements have equal lengths but the stiffness is different due to differences in shape and thickness. Therefore, good force balance requires that the elements are connected slightly outside the task middle point. If only visual information is used to complete the task and force feedback is mainly ignored, is expected that higher forces are exerted on the task than necessary. Moreover, smart positioning of the elements in both instruments and a good strategy is required for efficient handling. Completion of the task with only one instrument is not possible.

Task 2—Placement of a silicone wire: This task is made from two elastic elements with different elastic properties. A piece of artificial tissue with two holes at one side is connected to the task's ground plate on the other side. In order to drive the 2-mm thick elastic wire through the holes the artificial tissue needs to be turned and twisted for proper sight on the task. This is best achieved using both instruments in parallel as the use of only one instrument may result in fluctuating traction forces.

The new dynamic position tasks are based on an action analysis of manipulations during tissue dissection in real laparoscopic procedures. In tissue manipulation, often both laparoscopic grippers are used for the positioning of tissue in view of the camera. The orientation of the tissue inside the grippers is crucial for good inspection. Precise navigation of tissue under tractive force is common during, for instance, laparoscopic sterilization. In female sterilization, the ovarian tube needs to be positioned perpendicular to the laparoscopic camera and stretched for precise placement of a clip or ring. Placement of a clip on a stretched ovarian tube requires precise alignment of instruments and tissue. These manipulations are comparable with stretching the "worm" before placing it over a small pin of the opposite "worm" in our task. During this two-handed action, it is essential that the tractive force, generated by both instruments, remains low and constant when the clip is applied [7].

C. Test Protocol

The participants performed tasks inside a box trainer equipped with two 5-mm trocars and one 11-mm trocar (Endopath XCEL, Johnson & Johnson), two grippers (Endopath Ethicon Endosurgery, Johnson & Johnson), and a USB camera system (see Fig. 2). The tasks were mounted on top of a custom-made 3DOF force sensor (see Fig. 3). The top plate of the training box is nontransparent and the USB camera is used for visualization of the task on a computer screen. The order in which the tasks were performed was randomized for each participant.

Before the measurements started, a picture was shown to the participants to explain how to complete the two tasks. If a problem occurred in the first 2 min of the task, a new measurement was started for the next attempt and all recorded data

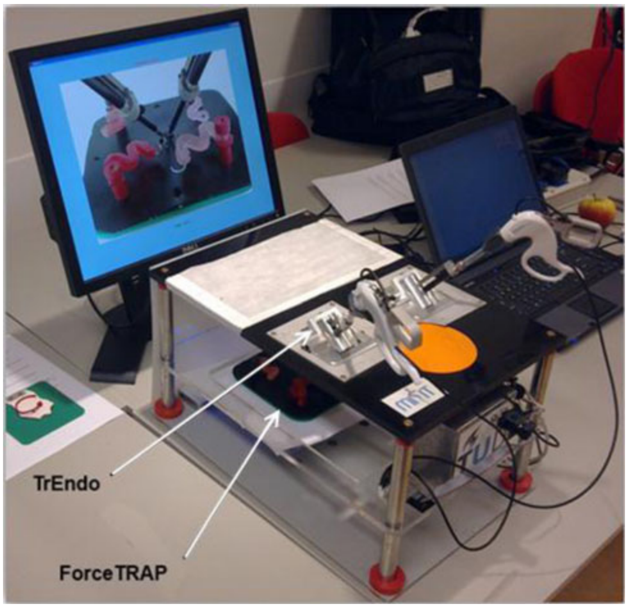

Fig. 2. ForMoST system measures force and motion with TrEndo and ForceTRAP. The image of the task is displayed on a computer screen.

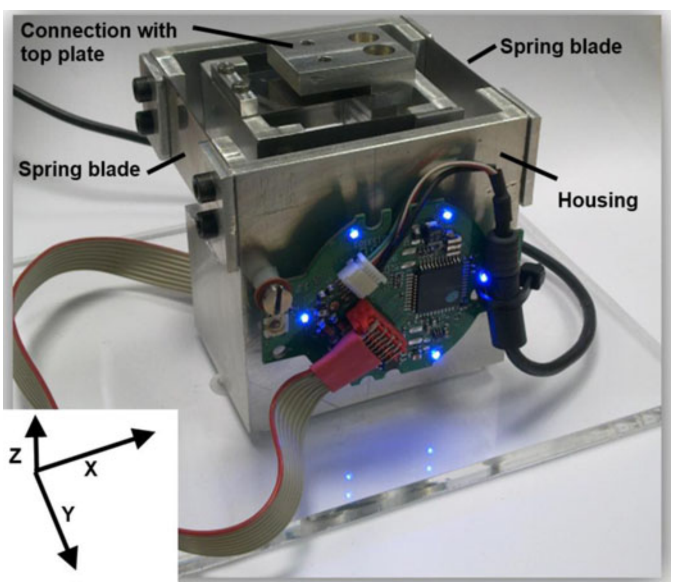

Fig. 3. Prototype of the 3-D ForceTRAP that is fixated between training task and bottom plate of ForMoST. The ForceTRAP is built from three parallelogram mechanisms and components of a 3-D connection mouse. The three parallelogram mechanisms prevent rotations around the sensor's midpoint. Therefore, accurate force measurements become possible even if forces are exerted further away from the sensors midpoint.

was deleted. If problems occurred after 2 min, the participant was removed from the study. Problems that can occur during the measurements are identified as: breaking of one of four artificial tissues due to excessive forces in Task 1 and falling out of sight of the silicon wire in Task 2. If necessary, participants received additional verbal instructions during the tasks. All participants performed only one of the two tasks in order to prevent learning effects. At the start of Task 2, the thread was positioned on a predefined location in the right upper corner of the ground plate of the task by the experimenter so that the starting conditions were the same across participants and trials. If it took students more than 15 min to complete the task at the first trial, they were excluded from the study and the data were removed.

D. Task Measurement Setup

The TrEndo and new 3-D force measurement platform (ForceTRAP) were integrated in a force and motion surgical trainer (ForMoST). The ForceTRAP is based on three parallel mechanism and uses the optical sensor unit of a commercially available optoelectronic device for sensing [5], [6]. Fig. 3 shows the ForceTRAP as it is placed between the task and bottom plate of the box. In this sensor, the first of three parallelogram mechanism consists of the housing that is connected with two spring blade to a U-profile that can only move in the X-direction. On this U-profile, of the second parallelogram is fixed that allows only movement in the Y-direction of the opposite U-profile. Finally, a third parallelogram is fixed between the second parallelogram and the optical sensor unit. Together, the three parallelograms allow movement of the optical sensor unit in X, Y, and Z and do not allow the sensor unit to rotate in any direction. The calibrated device has an accuracy of 0.1 N and threshold of 0.3 N. A more detailed description of the calibration including pictures of the setup and function fitting can be found in a previous study about the force platform [6].

Custom made software was written in MATLAB (2012b) to record the sensor output to a computer at a sample frequency of 30 Hz. A user interface allowed the experimenter to show the different training tasks with description on the training screen, to start and stop the USB video camera and to store data under a predefined filename. Finally, the user interface allows the experimenter to mark specific events during a measurement. The timestamp of these button presses was recorded alongside the sensor data and used to link written remarks to the recorded force and motion data.

E. Performance Parameters

To use motion and force information for skills assessment based on classification, performance parameters are required. The nature of a performance parameter depends on the surgical action it needs to reflect in a surgical training task. Seven existing and two new parameters, based on force, motion, and time are used to measure performance on the new training tasks. The parameters were chosen partly because of their discriminating power in earlier studies [5], [6] and partly based on the opinion of experienced surgeons.

1) Force Related Parameters: Max absolute force (MAF): The maximal force found in a trial indicating jerks or punches in instrument–tissue interactions [4].

Mean absolute nonzero force): Indicating the averaged mean absolute force of periods during training the absolute force is not nonzero [3].

Force volume (FV): Indicating the volume of an ellipsoid spanned around the standard deviations (SD) of the force along the three main principal components (PC's). The largest SD found in the 3-D force defines the orientation of PC1. The second largest SD defines the orientation of PC2 perpendicular to the first. PC3 oriented perpendicular to PC1 and PC2 [5].

$$V = \frac{4}{3\pi}(\mathrm{stdF_{pc1}} \cdot \mathrm{stdF_{pc2}} \cdot \mathrm{stdF_{pc3}}) \tag{1.1}$$

where V is the volume, $\mathrm{stdF_{pc1}}$ is the standard deviation of force along PC1-axis, $\mathrm{stdF_{pc2}}$ is the standard deviation of force along PC2-axis, and $\mathrm{stdF_{pc3}}$ is the standard deviation of force along PC3-axis.

2) Motion Related Parameters: Path length (PL-left and PL-right): Indicating the length of the 3-D instrument tip trajectory and is used as a measure to determine the efficiency of instrument motion for both instruments [5].

Motion volume (MV-left and MV-right): MV is a measure for the space required by the trainee to complete the task. Different from PL, MV is influenced by the direction the instrument tip moved in a 3-D space [5]. For calculation, (1.1) is used with left or right instrument motion data instead of force data.

Mean distance between tips (MDBT): The MDBT indicates if both instruments are in the area of interest:

$$T2TD = \sqrt{(x_l - x_r)^2 + (y_l - y_r)^2 + (z_l - z_r)^2} \tag{1.2}$$

where dist is the absolute distance between two tips, x_l is the position left tip on local x-axis, x_r is the position right tip on local x-axis, y_l is the position left tip on local x-axis, y_r is the position right tip on local x-axis, z_l is the position left tip on local x-axis, and z_r is the position right tip on local x-axis.

3) Force and Time Related Parameters: Max force area (MFA)—where the MAF parameter indicates the highest measured absolute force, the MFA is defined as the largest period with the highest absolute force between t_1 and t_2. In earlier study, MAF was referred to as force peak [5].

$$\mathrm{MFA} = \int_{t_1}^{t_2} |F| \mathrm{d}t \tag{1.3}$$

where F is the absolute force, t is the starting time of absolute force peak, and t is the stopping time of absolute force peak.

4) Motion and Time Related Parameters: Out of view time (OVT): Indicating the time that the instrument tips were not visual on the screen. In this new parameter, the local Z-axis is pointing upward from the middle of the task. After transformation, the new global Z-axis is pointing from the task midpoint toward the midpoint of the camera. The global X-axis (left and right in the box) remains the same while the Y-axis is oriented perpendicular to X global and Z global. The total time per instrument the max absolute distance between u (1.4) and midpoint of training task is exceeded is a measure for OVT:

$$u = \sqrt{x_g^2 + y_g^2 + z_g^2}$$
$$u_{\max} - u > 0 \tag{1.4}$$

where u is the shortest distance between tip and midpoint on task, $u_{\max}$ is the max allowed shortest distance between Z-axis and u, x_g is the position tip in global x coordinates, y_g is the position tip in global y coordinates, and z_g is the position tip in global z coordinates. The visual area between camera and task is shaped like a cone, orienting from the lens. The shape of the area that should not be left by the tip of the instruments is defined as a globe to simplify the algorithm and to minimize calculation power.

5) Time Related Parameters: Task time (T): Indicating the period of time elapsed between the start of a training and the first second after the task was completed.

F. Statistical Tests

A one-way ANOVA with Bonferroni Post-Hoc test (SPSS 17) was used to determine statistical differences between the experience level of groups. A p-value of less than 0.05 (two tailed) is considered to be significant. In the Pearson correlation matrices (SPSS 17), a correlation between parameters with $p < 0.05$ (two tailed) was considered significant.

G. Correlation Between Force and Motion Parameters

Pearson correlation matrices were used to investigate the relation between all motion and force parameters [4]. If high correlations are found between motion and force parameters, many contacts between tip and tissue are expected. If there are no correlations found, instrument motions are not directly related to the task and partly performed without contact between tip and tissue.

H. PCA, Classifier, and Leave-One-Out-Cross Validation (LOOCV)

Based on the classification methods used in our earlier study [4], the amount of correctly classified subjects (e.g., LOOCV score) is determined for the experts versus novices, experts versus intermediates, and intermediates versus novices. To investigate whether certain combinations of parameter categories (e.g., motion parameters, force parameters, or task-time parameter) give a better LOOCV outcome, we determined the LOOCV score for the combination task-time, force parameters, and motion parameters, the combination of force and motion parameters, the combination of task-time and motion parameters, and finally of task-time and force parameters. Only when at least one significantly different parameter is found for each of the included categories, it is possible to perform the analysis.

1) PC Analysis: In this study, the PCA analysis was used to calculate new PCs for the significant parameters of both tasks (princom.m, MATLAB 2008b). For each group of highly correlated parameters in the correlation matrix, PC analysis (PCA) was used to find new representative parameters for each group of correlated parameters. PCA orders the newly calculated PCs based on the amount of variance they explain. The first PC explains the most variance whereas the succeeding PC's explain the rest of the variance in decreasing order. For this study, we sum up the number of PC's from top down until a minimum of 75% of the total variance in the data is explained. Since the variance of the used parameters is extremely heterogeneous, all data was first normalized before PCA was applied. The data of each parameter were normalized according to:

$$Z = \frac{x - \mu}{\sigma} \tag{1.5}$$

where Z is the standard force parameter score, x is the raw force parameter score to be standardized, μ is the mean force parameter value, and σ is the standard deviation of force parameter.

2) Classifier: The two PCs that explain minimal 75% of the variance of the data from the participants are now used as input for the classifier (classify.m, MATLAB 2008b). The classifier describes the boundary between two groups with different skills levels with use of only two parameters.

3) Leave-One-Out-Cross Validation: To obtain the number of participants that can be correctly classified based on the data, an LOOCV program was written in MATLAB 2008b. For each LOOCV case, the training set consists of the data of all minus one participant while the data of that single participant is selected as a test case. The data of all participants are used once as test case resulting in a number of LOOCV cases equal to the amount of participants. During each LOOCV case, the skills level of the test case is predicted based on its location with respect to the boundary as determined by the LDA. Since the real experience level of each test case is known, the predicted outcome of each LOOCV case can be correct or incorrect. The percentage correctly classified LOOCV cases indicates how reliable new participants are classified based on the used dataset and force parameters. A more detailed description of LOOCV for classification can be found in our previous study [5].

III. Results

A. Statistical Difference Between Parameters

All participants were able to complete each of the two tasks within 15 min. The results per parameter are represented in Fig. 7-top for Task 1 and Fig. 7-bottom for Task 2. The presence of a p-value in a graph indicates a statistical difference between groups.

1) Task 1, Tissue Connection Under Traction: For Task 1, the MAF, STD force, task time, PL-left, and PL-right were found to be significantly different between the novice and expert group. Between the intermediate and expert group, the MAF, task time, PL-left, PL-right, MFA, and MDBT were found to be significantly different. For none of the parameters significant differences were found between the intermediate and novice group.

2) Task 2: Placement of a Silicone Wire: For Task 2, the task time, PL-left and PL-right, and OVT left were found to be significantly different between the novice and expert group. Between the intermediate and expert group the MAF, task time and FV, were found to be significantly different. Between the intermediate and novice group the FV, MDBT, PL-left, OVT left, and MV-right were found to be significantly different.

B. Correlation Between Force and Motion Parameters

Fig. 4 shows the correlation between force, motion parameters, and task time for the dynamic position task (Task 2). The yellow square in each of the three tables indicates the area where correlating force and motion parameters can be found. The top matrix of Fig. 4 shows that 12 of the 35 yellow blocks changed color meaning that 37% of the parameters were correlated in the

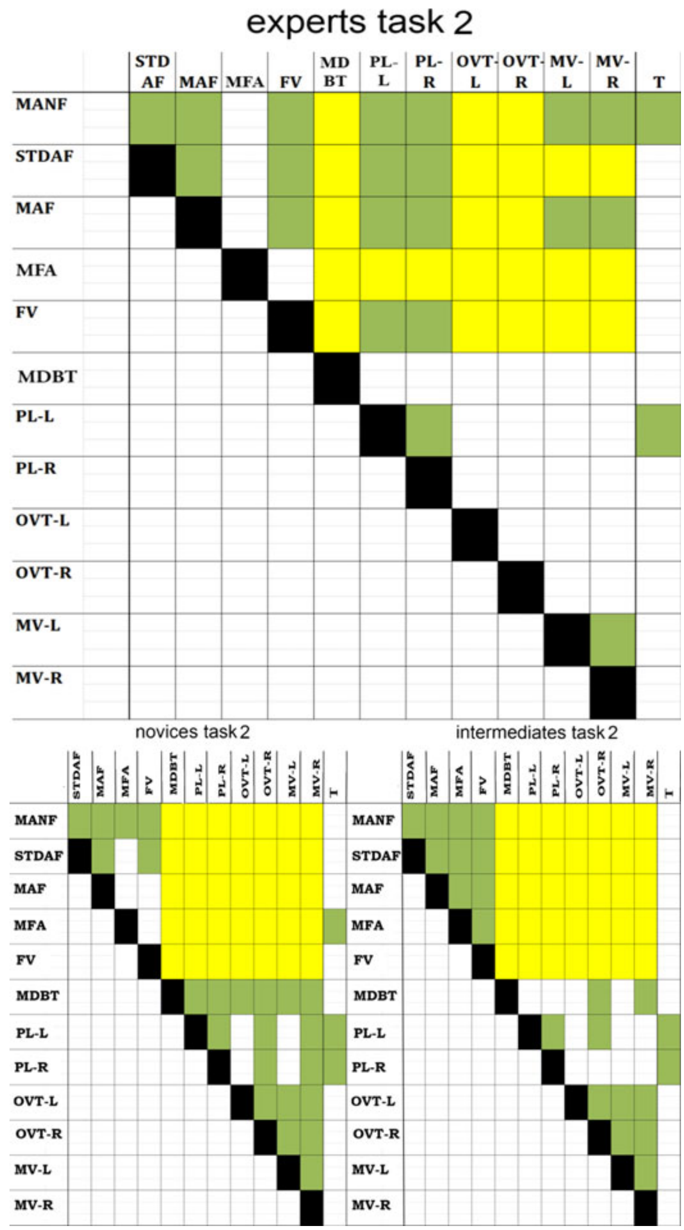

Fig. 4. Correlation matrices of experts, intermediates, and novices for Task 2. A green block indicates correlation between the parameter above and parameter left of the block ($p < 0.05$). If green blocks are found in the yellow area, the motion parameters above the block is correlated to a force parameter left of the block.

expert group for Task 2. The lower matrix in Fig. 4 shows that none of the blocks in the yellow area's changed color indicating that there is no correlation found between motion and force parameters for Task 2 in the intermediate group and novice group. For the tissue connection task (Task 1), 8.6% of the force and motion parameters were correlated in the novice group and none in the intermediate and experts group. Looking at the correlation between task time and motion parameters a correlation was found in each group between the PL of the left instrument and task time. Correlation between task time and the PL of the right instrument was found for both tasks in the intermediate groups, Task 1 in the expert group and Task 2 in the novice group.

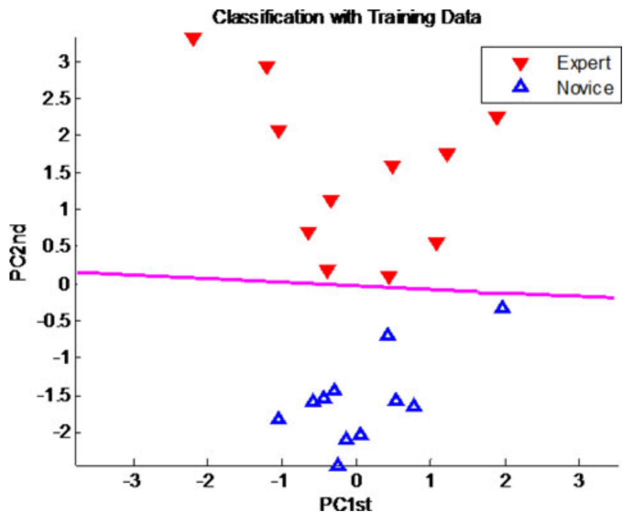

Fig. 5. LDA performed on the expert and novice data for Task 1 with all significant parameters. PC1st and PC2nd, largest and second largest PC in arbitrary units. Magenta line, boarder line as determined by the LDA. In this example, 100% of the participants were correctly assigned with the LOOCV.

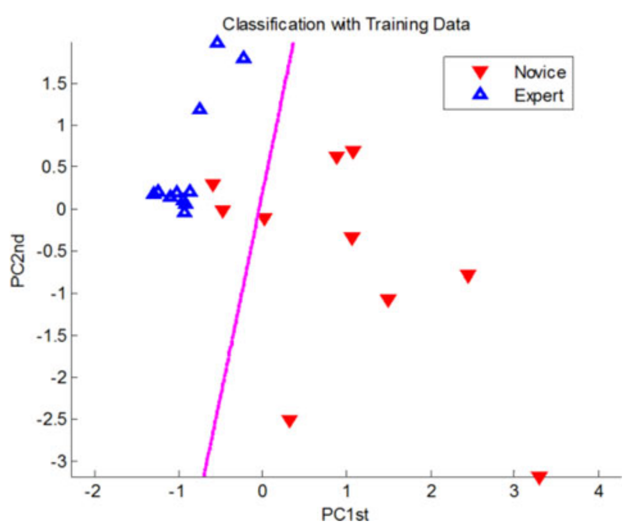

Fig. 6. LDA performed on the expert and novice data for Task 2 with all significant parameters. PC1st and PC2nd, largest and second largest PC in arbitrary units. Magenta line, boarder line as determined by the LDA. In this example, 91% of the participants were correctly assigned with the LOOCV.

C. PCA, Classifier, and LOOCV

Fig. 5 shows the distribution of experts and novices in Task 1 when the first two PCs (PC1st and PC2nd) are calculated from the LDA based on the significantly different MAF, STD force, task time, PL-left, and PL-right parameter data. With this dataset it is possible to discriminate between a novice and expert level with 100% accuracy. Fig. 6 shows the distribution of experts and novices when the significant different task time, PL-left and PL-right and OVT left data is used in Task 2. With this dataset it is possible to discriminate between the novice and expert level with 91% accuracy.

For Task 1, none of the parameters showed significant difference between the novice and intermediate groups and therefore the LOOCV was not performed. Since there was no significant difference in motion parameters between the expert and intermediate group of Task 2, the LOOCV was performed with the significant different force parameters (e.g., MAF and FV) and task time. Task time was not significantly different between the novice and intermediate group. Therefore, an LOOCV was performed with only force and motion parameters (e.g., FV, MDBT, PL-left, OVT left, and MV-right) and task time. Table I shows the outcome of the LOOCV after comparison of different skills levels with different sets of significant parameters.

TABLE I
PERCENTAGE CORRECTLY ASSIGNED PARTICIPANTS FOR EACH COMBINATION OF GROUPS FOR BOTH TASKS

	Novice-Experts		Intermediates-Experts		Novice-Intermediates	
	Task 1	Task 2	Task 1	Task 2	Task 1	Task 2
Motion Force Task-Time	100%	91%	70%	- *1	- *2	- *3
Motion Force	87%	91%	80%	- *1	- *2	77%
Motion Task-Time	91%	87%	87%	- *1	- *2	- *3
Force Task-Time	100%	91%	87%	76%	- *2	- *3

*1 There were no significant motion parameters found.
*2 There were no significant parameters found.
*3 Task time parameter was not different between groups.

IV. DISCUSSION

A. Aims of This Study

The first goal was to identify whether motion parameters are correlated to force parameters in Tasks 1 and 2. The correlation matrices indicated that a correlation between motion and force parameters was found only in the expert group of Task 2. This argues that motion and time parameters alone cannot be used to asses a student's tissue handling skills.

The second goal was to determine whether a combination of force, time, and motion parameters can be used for classification of the skills of a trainee. Especially, with Task 1 that required adequate force control besides motion control, it is possible to distinguish between novices and experts with 100% accuracy. For Task 2, it is still possible to discriminate between intermediates and experts with accuracies up to 91% and 87% depending on the set of parameters that is used as input for the LOOCV. Interestingly, if motion and force parameters and task time are used to distinguish between intermediates and experts in Task 1, the success rate declines. This can be explained since there is no correlation between motion and force parameter in the compared groups. Due to the nature of the used PCA, it is possible that the discriminative power of the force and motion parameters counteracts the discriminative power of the task time parameters when they are not correlated

For each of the part of this study (i.e., new tasks, performance parameters, correlation, and classification), the results are discussed in more detail in the following paragraphs.

B. Discrimination Power of Tasks

In order to find differences in tissue and instrument handling between groups with different skills levels, new training tasks were developed particularly for training of instrument motion and tissue manipulation. The results indicate that 7 out of 13 parameters were significantly different between skills levels for Task 1 and 8 out of 13 parameters for Task 2. Although the discriminating power of the used parameters is high in this study, Table I shows that the discrimination power varies over each combination of parameters in each task. Task 1 requires adequate motion and force control and therefore the highest LOOCV outcome was found if force parameters were used in the analysis. The LOOCV for Task 2, which required mainly adequate motion control, gave the best results if motion parameters were part of the analysis. For both tasks, enough parameters were found to be statistically different in order to classify between a novice or intermediate level and an expert level with more than 87% accuracy.

C. Observations

The FV shows high differences between the intermediates and novices as well as intermediates and experts in both tasks. In general, a high FV as seen in the intermediate group, results from fast increasing and decreasing forces (i.e., force spikes) in multiple directions. One explanation for this relatively large difference is that students in the intermediate group are more convinced about their motion control but less skilled as they may think. Since they are not familiar with these new tasks, the high values in FV could indicate that fast increasing forces result from a slow reaction on unexpected restrictions in movements (i.e., contact with ground plate or stretched tissues). As also observed, the novices seemed imposed by the given instruction to handle all instruments and tissues with great care and move their instruments carefully in order to prevent potential damage to the task. The experts, however, have better understanding of the developed training tasks and focus on adequate force and motion control thereby preventing sudden collisions between the solid parts of the tasks and instruments.

D. Correlation Between Force and Motion Parameters

Compared with the intermediate and novice group, that showed almost no correlation between force and motion parameters in both tasks, a correlation of 37% for Task 2 was found in the expert group. In the expert group of Task 1, no correlation was found between motion and force parameters. In general, if a high correlation between force and motion occurs, better control is assumed due to more efficient instrument handling and therefore less unintended force exertion during performance. In the novice and intermediate group, however, many instrument motions and tissue manipulations are accidental and caused by the unfamiliarity with mirror and scaling effects or depth perception difficulties. Since more instrument motions and tissue manipulations are not intended or not effective, less correlation is found.

For both tasks, higher correlation was expected in the expert groups compared with the novice and intermediate groups. However, this is only partly true since no correlation was found for Task 1 in the expert group. One reason could be that Task 2 was more familiar for the expert surgeons that all were highly experienced with manipulating tissue. The experts recognized the step to step approach and created clear vision on the backside of the silicone tissue before the thread was inserted. The novices and intermediates basically ignored this first critical step and started to manipulate the tissue without a clear strategy in mind till an opportunity occurred. In other cases, some novices and intermediates tried to push the thread in the hole without clear vision. Therefore, the clear uniform approach of the surgeons could explain a higher correlation in this task.

Compared with the clear uniform approach that was observed in the expert group during Task 2, more different strategies were used to solve Task 1. This could explain why a high correlation between force and motion parameters was not observed.

E. Skills Classification

The results in Table I indicate that both tasks can be used to classify the difference in skills levels between novices and experts with high accuracy. Compared with the standard tasks used by Chmarra *et al.* [3] for classification, the new dynamic position task show slightly higher LOOCV results (±90% versus ±80%). If the difference between skills levels becomes smaller, so is the discrimination power of the used LOOCV method. Looking at the skills levels of the intermediates in this study, it is still possible to discriminate them from the expert group with acceptable accuracy but not from the novice group. Fortunately, for the assessing teacher, it is mainly interesting to know whether an expert level is reached. If not, additional training is required.

Besides skills level, also the nature of the training task determines the outcome of the LOOCV. In general, the LOOCV outcome for Task 1 is slightly better compared with Task 2. Looking at the individual results of the parameters in Fig. 7, it becomes clear that mainly the force parameters have more discriminative power for Task 1. This task requires well-controlled manipulation force for adequate completion whereas Task 2 is performed efficiently when no force is exerted. Since instrument positioning with tissue under traction in Task 1 was not trained by the intermediates and novices, adequate force control was found difficult explaining the difference in scoring with experts on force parameters. Furthermore, all instrument motions (incl. motion errors) with tissue under traction in Task 1 always resulted in force data whereas Task 2 only records force data when motion or manipulation errors occurred. Therefore, the force data of Task 1 gives force parameters with potentially more discriminating power. Table I indicates that the discriminating

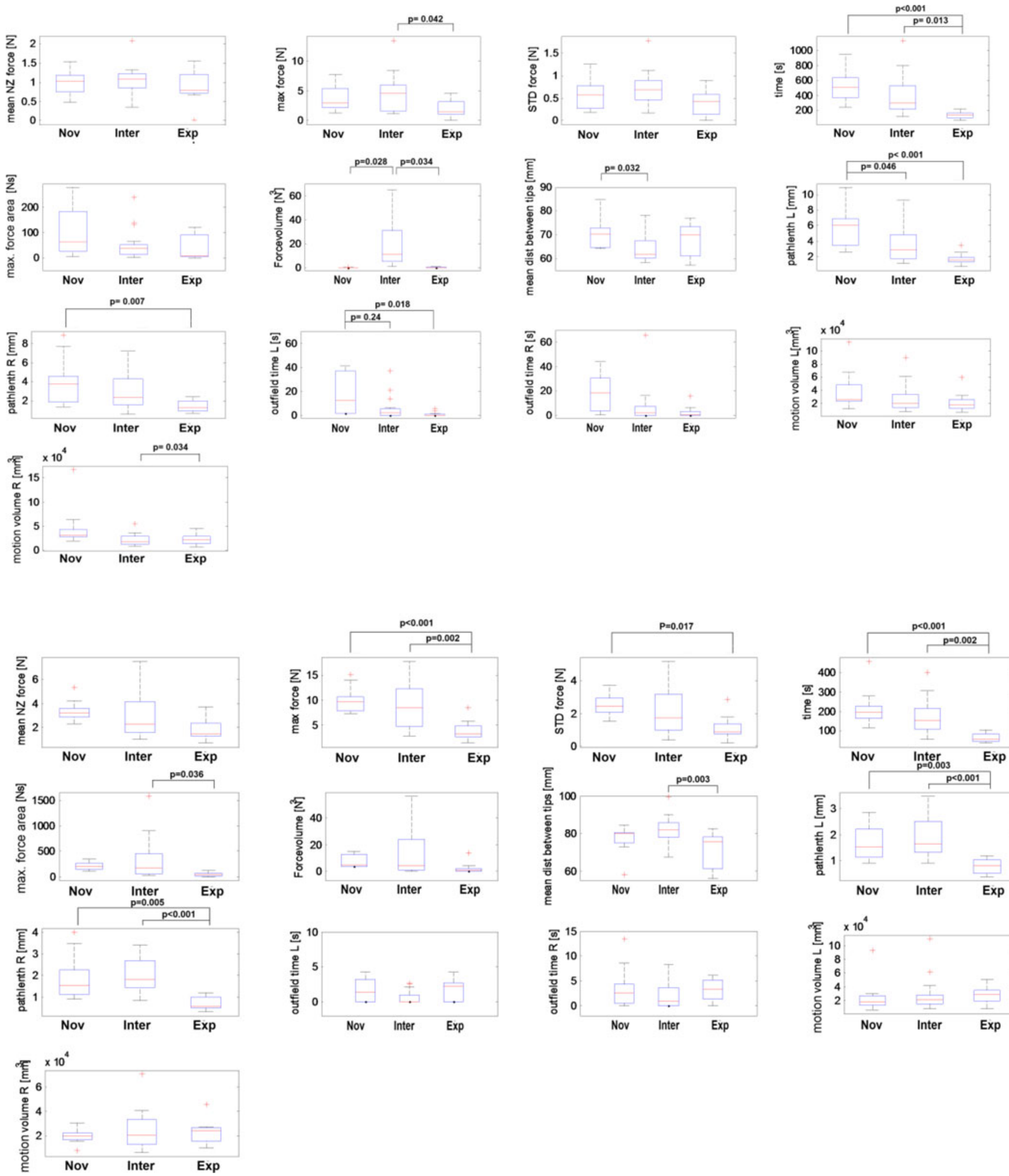

Fig. 7. Boxplot representation of the parameter results for (top) Task 1 and (bottom) Task 2. Each graph represents the results for the novice, intermediate, and expert group.

power of the selected parameters is linked to the actual skills levels in a group. For example, where multiple significant different motion parameters were found for the expert and novice comparison of Task 1, none were found for Task 2. The force parameters, however, proved useful for both tasks in this comparison.

F. Fundamental Training Tasks

For further study, it is interesting to investigate the discriminative power of standard FLS tasks. When the guidelines and instructions are considered, the results can be compared with other studies performed with the FLS tasks. Although an additional set of tasks was developed especially for the assessment of basic tissue handling during a dynamic position task it is interesting to investigate the discriminative power of the tasks used in FLS when motion and force data is used. Besides the suture tasks, also the circle cutting task requires well-controlled traction during cutting. Due to this nature, both the suture task and circle cutting task could reflect important differences in force control during delicate tissue handling.

V. Conclusion

A new set of dynamic position tasks was developed that requires not only motion control but also adequate force control for good results. If the data from the motion, time, and force parameters are used for classification, it is possible to distinguish the skills level of a novice or expert with an accuracy up to 100%. The results indicate that the tissue manipulation forces of many intermediates exceed the levels of the novices indicating that the focus on task time and instrument motion alone during a skills training course has a negative influence on the tissue handling skills of some students. The relatively high forces used by the intermediates in combination with the apparent lack of correlation between force and motion parameters argues for the inclusion of training and assessment of force application in tissue handling in laparoscopic skills training programs.

Acknowledgment

The authors would like to thank F. van Delft, a graduation student from the Hogeschool Zuyd in Heerlen, for doing all the necessary measurements during his internship. They also thank all participating students, residents, and surgeons for taking part in this study.

References

[1] R. Campo, C. Reising, Y. Van Belle, J. Nassif, P. O'Donovan, and C. R. Molinas, "A valid model for testing and training laparoscopic psychomotor skills," *Gynecological Surg.*, vol. 7, no. 2, pp. 133–141, 2010.

[2] W. Kolkman, M. A. J. van de Put, R. Wolterbeek, J. B. M. Z. Trimbos, and F. W. Jansen, "Laparoscopic skills simulator: Construct validity and establishment of performance standards for residency training," *Gynecological Surg.*, vol. 5, no. 2, pp. 109–114, 2007.

[3] M. K. Chmarra, S. Klein, J. C. F. de Winter, F. W. Jansen, and J. Dankelman, "Objective classification of residents based on their psychomotor laparoscopic skills," *Surg. Endosc.*, vol. 24, pp. 1031–1039, 2010.

[4] C. Richards, J. Rosen, B. Hannaford, C. Pellegrini, and M. N. Sinanan, "Skills evaluation in minimally invasive surgery using force/torque signatures," *Surg. Endosc.*, vol. 14, no. 9, pp. 791–798, 2000.

[5] T. Horeman, J. Dankelman, J. J. van den Dobbelsteen, S. P. Rodrigues, and F. W. Jansen, "Force parameters for skills assessment in laparoscopy," *IEEE Trans. Haptics*, vol. 5, no. 4, pp. 312–322, 2012.

[6] T. Horeman, S. P. Rodrigues, F. W. Jansen, J. Dankelman, and J. J. Dobbelsteen, "Force measurement platform for training and assessment of laparoscopic skills," *Surg. Endosc.*, vol. 24, no. 5, pp. 3102–3108, 2009.

[7] T. Horeman, D. D. Kurteva, P. Valdastri, F. W. Jansen, J. J. van den Dobbelsteen, and J. Dankelman, "The influence of instrument configuration on tissue handling force in laparoscopy," *Surg. Innov.*, vol. 20, no. 3, pp. 260–267, 2012.

[8] R. B. Singapogu, S. DuBose, L. O. Long, D. E. Smith, T. C. Burg, C. C. Pagano, and K. J. Burg, "Salient haptic skills trainer: Initial validation of a novel simulator for training force-based laparoscopic surgical skills," *Surg. Endosc.*, vol. 27, no. 5, pp. 1653–1661, 2013.

[9] A. M. Derossis, J. Bothwell, H. H. Sigman, and G. M. Fried, "The effect of practice on performance in a laparoscopic simulator," *Surg. Endosc.*, vol. 12, pp. 1117–1120, 1998.

[10] J. H. Peters, G. M. Fried, L. L. Swanstrom, N. J. Soper, L. F. Silli, B. Schirmer, and K. Hoffman, "Development and validation of a comprehensive program of education and assessment of basic fundamentals of laparoscopic surgery," *Surgery*, vol. 135, no. 1, pp. 21–27, 2004, SAGES FLS Commitee.

[11] E. Hiemstra, W. Kolkman, R. Wolterbeek, B. Trimbos, and F. W. Jansen, "Value of an objective assessment tool in the operating room," *Can. J. Surg.*, vol. 54, no. 2, pp. 116–122, 2011.

Tim Horeman was born in Haarlem, Holland, in 1980. He received the M.Sc. degree in biomedical engineering from the Delft University of Technology (DUT), Delft, The Netherlands, in 2008, with specialization in minimally invasive surgery and medical safety. He is currently working toward the Ph.D. degree at Minimally Invasive Surgery and Interventional Technology Group, Faculty of Mechanical, Maritime, and Materials Engineering, DUT.

His research interests include endoscopic instrument development and training and assessment of technical skills in minimally invasive surgery.

Jenny Dankelman received the degree in mathematics, with specialization in system and control engineering at the University of Groningen, Groningen, The Netherlands, in 1984 and the Ph.D. degree at the Man–Machine Systems Group, Faculty of Mechanical, Maritime, and Materials Engineering, Delft University of Technology (DUT) , in 1989, on her research of the dynamics of the coronary circulation.

Since 1992, she has been a Researcher at the Man–Machine Systems Group, DUT. In 2001, she became Professor in minimally invasive surgery and interventional techniques at the Department of Biomechanical Engineering. Her interests and research projects include the fields of medical instruments, skills training, and patient safety with a focus on minimally invasive techniques.

Frank Willem Jansen received the medical degree at the University of Groningen, Groningen, The Netherlands, the residency in obstetrics and gynecology at the Leiden University, Leiden, The Netherlands, and the Ph.D. degree in laparoscopic surgery in gynecology in 1997.

He is a Senior Consultant and Head of the section Minimally Invasive Surgery (MIS) in gynecology at the Leiden University Medical Center, Leiden, The Netherlands. Since 2009, he has been appointed as a Professor, both at the Delft University of Technology (MISIT Group) and the Leiden University (Medicine). His research interests include technology and quality assessment studies (implementation, complication, and patient safety) and skills training in MIS.

John J. van den Dobbelsteen received the degree in psychophysics at the University of Groningen, Groningen, The Netherlands, and the Ph.D. degree at the Department of Neuroscience, Erasmus MC, Rotterdam, for his research into human motor control, motor learning, and sensorimotor integration.

In 2005, he started as a Research Fellow at the Department of Biomechanical Engineering of the Delft University of Technology, Delft, The Netherlands, where he is currently an Assistant Professor in the research group Minimally Invasive Surgery and Interventional Techniques. His research projects include the development of medical trainers and measurement devices to study forces and motion during instrument–tissue interaction in minimally invasive surgery, interventional radiology and anesthesiology.