Surgical Skill Assessment Using Motion Quality and Smoothness

Ahmad Ghasemloonia, PhD,,† Yaser Maddahi, PhD,* Kourosh Zareinia, PhD,* Sanju Lama, PhD,* Joseph C. Dort, MD,† and Garnette R. Sutherland, MD**

*Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada; and †Department of Surgery, Arnie Charbonneau Cancer Institute, Section of Otolaryngology—Head & Neck Surgery, Cumming School of Medicine, University of Calgary, Calgary, Alberta, Canada

OBJECTIVES: This article presents a quantitative technique to assess motion quality and smoothness during the performance of micromanipulation tasks common to surgical maneuvers. The objective is to investigate the effectiveness of the jerk index, a derivative of acceleration with respect to time, as a kinetostatic measure for assessment of surgical performance.

DESIGN: A surgical forceps was instrumented with a position tracker and accelerometer that allowed measurement of position and acceleration relative to tool motion. Participants were asked to perform peg-in-hole tasks on a modified O'Connor Dexterity board and a Tweezer Dexterity pegboard (placed inside a skull). Normalized jerk index was calculated for each individual task to compare smoothness of each group.

SETTING: This study was conducted at Project neuroArm, Cumming School of Medicine, the University of Calgary.

PARTICIPANTS: Four groups of participants (surgeons, surgery residents, engineers, and gamers) participated in the tests.

RESULTS: Results showed that the surgeons exhibited better jerk index performance in all tasks. Moreover, the residents experienced motions closer to the surgeons compared to the engineers and gamers. One-way analysis of variance test indicated a significant difference between the mean values of normalized jerk indices among 4 groups during the performance of all tasks. Moreover, the mean value of the normalized jerk index significantly varied for each group from one task to another.

CONCLUSIONS: Normalized jerk index as an independent parameter with respect to time and amplitude is an indicator of motion smoothness and can be used to assess hand motion dexterity of surgeons. Furthermore, the method provides a quantifiable metrics for trainee assessment and proficiency, particularly relevant as surgical training shifts toward a competency-based paradigm. (J Surg Ed ■:■■■-■■■. © 2016 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: surgical skill, motion, smoothness, jerk index, acceleration, dexterity

ACGME COMPETENCIES: Medical Knowledge, Practice-Based Learning and Improvement

INTRODUCTION AND LITERATURE REVIEW

Currently, surgical trainees acquire technical skills through years of hands-on training in the operating room (OR) in an apprenticeship model, supplemented by written anatomy examinations, tutorials, and laboratory-based surgical skill courses using cadavers or models. More recently, virtual reality (VR)–based simulations have been included in this paradigm. Although knowledge of anatomy can be assessed by written or oral examination, assessment of technical skill such as dexterity and tool handling remains more subjective than objective.

Assessment of surgical skills in both open surgery and minimally invasive surgery is necessary to ensure patient safety and provides information for residents to enhance their skills before operating on patients. Surgical training is shifting to a competency-based education paradigm[1,2] and assessing surgical competence has several potential benefits, including improved safety of surgical training processes, enhanced accreditation of specialists, and maintenance of

Correspondence: Inquiries to Garnette R. Sutherland, Department of Clinical Neurosciences, Cumming School of Medicine, University of Calgary, RM 1C58 Health Research Innovation Centre, 3280 Hospital Dr. NW, Calgary, Alberta, Canada T2N 4Z6; e-mail: garnette@ucalgary.ca

 1931-7204/$30.00 http://dx.doi.org/10.1016/j.jsurg.2016.10.006

public confidence in the surgical profession. Currently, most of the assessment methods for dexterity of residents are based on subjective evaluation by an expert surgeon observing the residents conducting different surgical tasks. The assessment for the same task could vary depending on who evaluates the performance and could even be biased.[3,4] This article proposes a quantitative method for assessment of the motion quality and smoothness during the performance of micromanipulation tasks and explores the correlation between the level of surgical training and the proposed metrics.

Skills that are currently evaluated include respect for tissue, aggressive or smooth motion, and instrumentation handling.[3-5] The subjective assessment of trainees by their preceptors has evolved to make the assessment process more standardized. Score sheets are often used for subjective assessment; however, the scoring procedure is dependent on the assessment conducted by experts. The Objective Structured Assessment of Technical Skill rating system,[6] the Global Operative Assessment of Laparoscopic Skills,[7] and the Fundamental Laparoscopic Skills program[8] are assessment methods that incorporate checklists to calculate an overall score for dexterity. The methods include some objective performance metrics including time and number of errors, as well as some subjective measures that still require an expert to judge the performance. A complete review of assessment of technical skills is included in the publication by Grantcharov et al.[9] Accurate, quantitative evaluation, and improved training efficiency are demands imposed by new training paradigms that have reduced work hours and training resources for surgical residents.[10] Virtual simulators are one of the ways to achieve accurate and unbiased quantitative measurement of certain aspects of surgical performance.[11] Objective assessment of surgical skills reduces the need for subjective evaluation and provides information for improvement of specific surgical tasks.[12,13]

Objective feedback of technical skills is essential to the structured learning of surgical technique and provides essential feedback for trainees. This type of evaluation, while evolving, has not been widely adopted into clinical practice because of expensive instrumentation required and lack of reliable objective measures of technical performance.[12] Objective assessment methods can be categorized into procedure-specific checklists, global, rating scales, motion analysis, VR simulators, and automated video-kinematic assessment.[13] A review of objective assessment techniques has been conducted by Moorthy et al.[13] and van Hoove et al.[12] They concluded that objective feedback of technical skills is crucial to the structured learning of surgical skills. They highlighted the progress in the methods of objective assessment of technical skills to provide objective feedback and help residents to improve their skill. They also discussed the potentials of VR simulators as an objective assessment method. Additionally, it was concluded that most methods of skills assessment that are valid for measuring training progress could also be used for examination or credentialing. Moreover, different methods of skills assessment are appropriate in different assessment scenarios. However, as they mentioned, further research is required to address the limitations and determine the link between objective assessment of technical skills and measured parameters such as complication and recurrence rates and postoperative pain.

Most of the surgical tools developed for objective assessment measure kinetostatic characteristics of the surgical tool and the surgeon's hand during the performance of surgical operations. These objective measures include time, trajectory traveled by the instrument, number of movements, peak forces, and mean values of velocity and acceleration. Although the earlier measures could be considered as a set of indicators to evaluate surgical dexterity and skills without inclusion of any biased judgment, they do not provide direct feedback for motion quality or smoothness and are not sufficient to indicate a resident's surgical skill in tool handling and quality of their hand motion for the tasks involving tactical approach. To address this gap, frequency analysis and peak acceleration have recently been considered as new metrics to incorporate effect of the hand motion in both time and frequency domains.[14] The addition of data analysis in the frequency domain helps to evaluate the smooth motion that is normally violated by jerky motion, tremor, and hesitant motion.

Motion smoothness in handling a surgical tool is an essential skill that surgical residents need to acquire before operating on patients. Motion smoothness is an indicator of skilled coordinated hand motion.[15] Smoothness or gracefulness is usually quantified based on the rate of change in acceleration (third derivative of the tool/surgeon's hand position) or curvature of the path, where a low curvature shows a straight line to the target and curvature values close to 1 means abrupt changes in curvature and jerky movements.[3] Motion smoothness is also measured as a cumulative number of sudden accelerations and decelerations. Jerk index is a scalar value that quantifies the smoothness and is influenced by the duration of the task and movement amplitude. As the jerk index is measured over a time interval, the integral of the mean squared magnitude of jerk is conventionally considered. Jerk measure is a time- and amplitude-dependent factor and represents the smoothness of the motion and tool handling during the performance of a given task. The smoothest motion results in a lower jerk index. To eliminate the effects of time and amplitude, the jerk index needs to be normalized by multiplying [(duration interval)^5/(3D path length)^2] to the jerk index.[16] It has been proved that normalized jerk cost can quantify smoothness and coordination.[16] This is because of the fact that smoothness is an indicator of movement quality that should be independent of speed and distance. A review of normalization procedure of jerk index is discussed by Hogan et al.[17]

Jerk index has been used as a quantitative metric in studies related to laparoscopic skills,[16] upper limb motor function,[18] movement smoothness in golf,[19] and irregularity of jaw movement during chewing.[20] Cotin et al.[21] used the smoothness as a metric for evaluating laparoscopic skills, which showed that experts had a smoother motion compared to novice participants. Judkins et al.[3] measured curvature of the motion signal (the tendency of the motion to keep the trajectory as a straight line) in robotic surgery. Median values of curvature were used to determine smoothness and objective evaluation of novices and experts in the performance of robotic surgical training tasks. Chmarra et al.[22] defined smoothness as a motion metric to objectively classify residents according to their basic laparoscopic skills. The smoothness metric was able to distinguish skill level as experienced, intermediate, or novice. Also, in their study, time, path length, depth perception, and motion smoothness were positively correlated. Sakata et al.[23] used the jerk square mean value as the average of the square of the joint angle third derivative value to study age-related changes in smoothness of lower extremity joint movement during load lifting. These results verified that smoothness in the hip and ankle joints during lifting decreases with advancing age. Carpinella et al.[18] used jerk index to study arm impairment in multiple sclerosis for a sensitive quantification of arm function. Action Research Arm Test tasks executed by patients with multiple sclerosis were significantly less smooth (jerk increased) with respect to controls. Choi et al.[19] studied kinematic evaluation of movement smoothness in golf with the normalized jerk cost. The jerk index analysis verified that skilled golfers have smoother swings than unskilled golfers.

To investigate smoothness using the jerk index, digitized acceleration and position recordings are converted into reliable quantitative metrics that correlate experimental data to surgical performance. Normally, this conversion requires complicated postprocessing of the digitized data to correlate, for instance, the metrics with hand motion dexterity and surgical skills, that is, smooth motion. In addition, artificial neural network algorithms, with their own shortcomings, are required to train the systems using experienced surgeons. Furthermore, the experimental environment and the surgical tasks should realistically simulate conditions inside the OR during the performance of surgical procedures and should accurately measure the quantitative metrics.

This study focuses on investigating the potential application of the normalized jerk index to compare the dexterity and smoothness of motion during the performance of 3 sets of micromanipulation tasks that simulated picking up and placing cotton strips in brain surgery.[24] This study shows how participants with different levels of expertise exhibit different performance while conducting the same micromanipulation task. The measure used was the normalized jerk index that was determined by integrating the square of the absolute jerk over time. Three pegboards were designed to test smoothness of the motion and the participant dexterity during completion of each task. In addition, effect of the tool handling (unanchored and anchored wrist) on the task performance was investigated.

METHOD

Experimental Test Rig and Instrumentation

A bipolar forceps (Codman & Shurtleff Inc., MA, USA) was augmented by an accelerometer (Mide Technology, MA, USA) to record acceleration and by a position tracker (Polhemus, VT, USA) to record the path of the tool motion. The acceleration signals were recorded at a sampling rate of 120 Hz and the position signals at a sampling rate of 120 Hz. The data were transferred to the computer for postprocessing of signals. The instrumented forceps are shown in Figure 1.

As observed in Figure 2, 2 different test platforms were designed to investigate the feasibility of analyzing smoothness and dexterity of hand motion with the normalized jerk index. The modified versions of O'Connor Finger Dexterity and Tweezer Dexterity tests were chosen, as they are capable of testing both manual dexterity and the ability to use fine instruments. The first test board was a normal O'Connor Finger Dexterity board with 100 holes (Fig. 3A). The holes are rectangular with randomly distributed directions. Nine holes are numbered in the board, and participants were asked to place a pin in each hole with the bipolar forceps (task A). For the next 2 tests, a smaller O'Connor Dexterity board was mounted inside a skull (5 cm below the surface) to mimic brain surgery conditions, as shown in Figure 3B and C. The participants placed 9 rectangular pins inside the 9 holes of the board. This test was conducted in 2 different conditions: first with unanchored condition (Fig. 3C) where no pivot point was allowed for resting the hand (task B), and for the second test, the same test was conducted with anchored hands where the participants were allowed to anchor their hand over the skull or over the test table (task C), similar to how surgeons perform some surgical procedures to have a more-steady hand.

Test Procedure

Four different groups of subjects participated in this study: surgeons, residents, engineers, and gamers. Surgeons included 2 neurosurgeons (5 and 25 y in practice), 1 otolaryngologist (25 y in practice), and 1 general surgeon

FIGURE 1. The augmented bipolar forceps (1) instrumented by accelerometer (2) and position tracker (3).

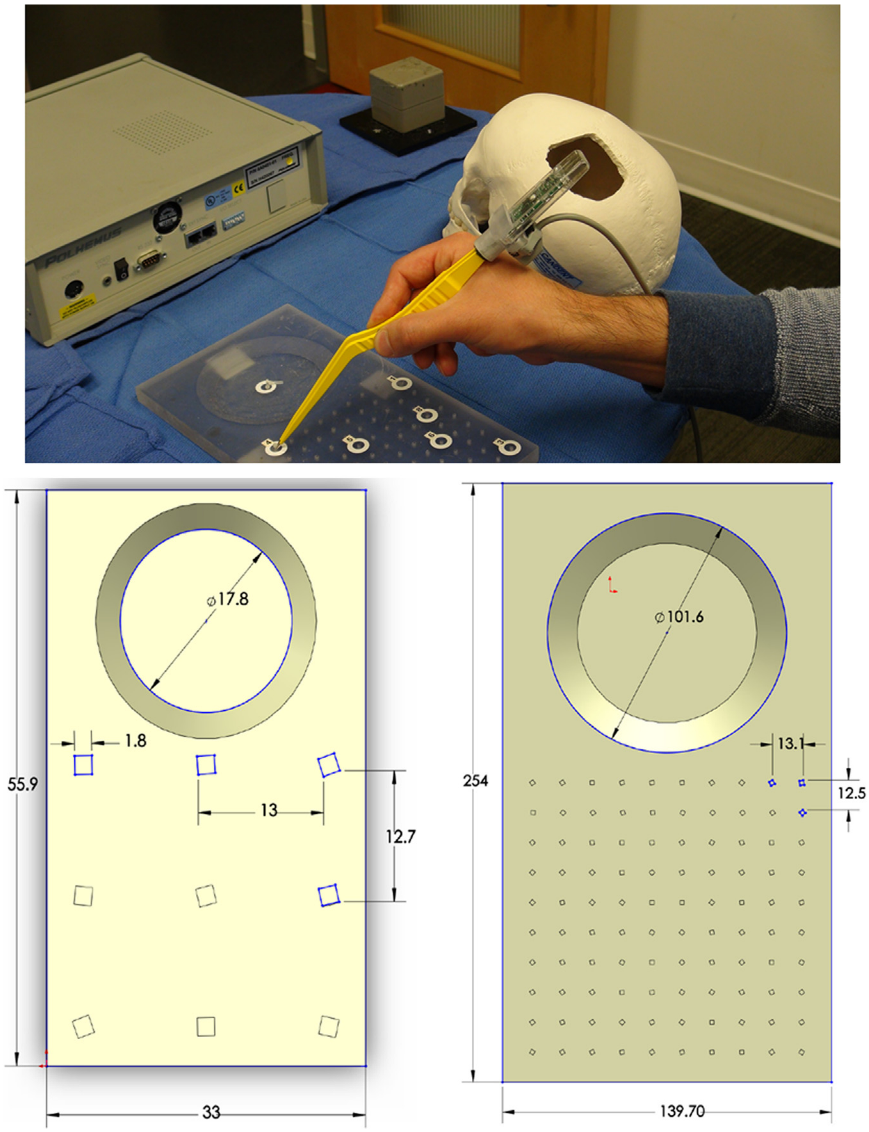
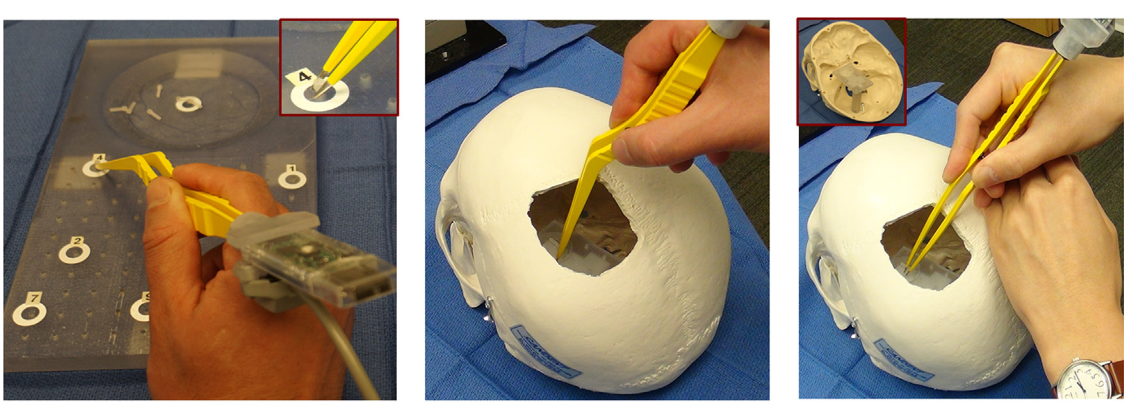


FIGURE 2. Experimental test rig developed to conduct the experiments. Two pegboards (numbered and designed to test the dexterity and smoothness of the motion while conducting 3 micromanipulation tasks.



Task A Task B Task C

FIGURE 3. Three micromanipulation tasks: task A—O'Connor Finger Dexterity test with unanchored configuration of the wrist, task B—Tweezer Dexterity test with unanchored configuration, and task C—Tweezer Dexterity test with anchored configuration.

(25 y in practice). Engineer group consisted of mechanical, electrical, and biomedical engineers (25, 30, 35, and 45 y old), and residents included 2 neurosurgery residents (postgraduate year 1 [PGY1] and PGY3) and 2 otolaryngology residents (PGY2 and PGY4). Gamers (a participant in a computer or role-playing game) were all undergraduate students.

For each participant, a video demonstration of the performance of each task was provided, followed by a training period of not less than 15 minutes to familiarize each participant with the setup before conducting the real experiments. Participants consisted of 4 surgeons, 4 residents, 4 gamers, and 4 engineers. Once the participant felt comfortable in handling the bipolar forceps and performing the experiment, the task A (big board), task B (small board —unanchored), and task C (small board—anchored) were conducted. Each participant conducted 9 trial runs for each task. Therefore, 4 (groups) × 4 (participants) × 3 (tasks) × 9 (trials) = 432 trial runs were collected. The experiments were performed without any preference in sequence by surgeons, engineers, neurosurgery, and otolaryngology residents, and gamers.

DATA PROCESSING

The recorded signals for each participant were transferred to a computer for postprocessing. In the first step, the signals were segmented to compute the jerk index for each motion. Nine signals are stored for the big board experiment, 9 signals for the small board–anchored experiment and 9 signals for small board–unanchored test. In the next step, as gravity acceleration is included in the acceleration signals, the gravity component was filtered by implementing a high pass fourth order Butterworth filter with a cutoff frequency of 0.2 Hz. This resulted in 3-dimensional acceleration signals without the gravity component.[25]

To determine the jerk index of each signal, the length of the pathway needs to be determined as follows[26]:

$$L_{\text{pathway}} = \int_{t_1}^{t_2} \sqrt{\left(\frac{\partial x}{\partial t}\right)^2 + \left(\frac{\partial y}{\partial t}\right)^2 + \left(\frac{\partial z}{\partial t}\right)^2}\, dt \tag{1}$$

where x, y, and z are position coordinates along the Cartesian coordinate system, and t_1 and t_2 are the start and end time of the task. The jerk index can be determined using[20]:

$$J = \int_{t_1}^{t_2} \left(\frac{\partial^3 x}{\partial t^3}\right)^2 + \left(\frac{\partial^3 y}{\partial t^3}\right)^2 + \left(\frac{\partial^3 z}{\partial t^3}\right)^2 dt \tag{2}$$

As mentioned in the Literature Review section, in order to compare the smoothness, the effect of duration and amplitude must be removed from the jerk index, or the jerk index should be normalized. The normalized jerk cost can be determined as follows:

$$J_{\text{normalized}} = \int_{t_1}^{t_2} \left(\frac{\partial^3 x}{\partial t^3}\right)^2 + \left(\frac{\partial^3 y}{\partial t^3}\right)^2 + \left(\frac{\partial^3 z}{\partial t^3}\right)^2 dt \times \frac{(t_2 - t_1)^5}{\left(\int_{t_1}^{t_2} \sqrt{\left(\frac{\partial x}{\partial t}\right)^2 + \left(\frac{\partial y}{\partial t}\right)^2 + \left(\frac{\partial z}{\partial t}\right)^2}\, dt\right)^2} \tag{3}$$

These equations are applied on our recorded digitized data at a sampling frequency of 120 Hz while the highest frequency component in the power spectral density plots was 13 Hz.

STATISTICAL ANALYSIS

Mean values ± standard deviation ($\mu \pm \sigma$) of quantified jerk indices were reported to quantitatively compare the task performance of the participants in the 4 groups: surgeons, neurosurgery and otolaryngology residents, engineers, and gamers. Results of the 2 experiments, using the large pegboard and the small pegboard, were compared using the statistical information of each group. In addition, analysis of variance (ANOVA) was employed to investigate the differences or similarities of the measured indices among all 4 groups. For each group, the type of task was compared for the participant to see if the level of difficulty could affect the amount of the jerk index. With respect to the second task, the ways each participant held the surgical tool was also investigated to observe how handling the tool affected the value of the jerk index. The procedure with $p < 0.05$ was considered as a significant trial.[27]

The jerk index was considered as the random variable to compute the values of kurtosis (to estimate the peakedness of the probability distribution of the jerk index) and skewness (to estimate the asymmetry of the distribution).[28] Each set of jerk indices was recognized to differ or skew to a significant degree if the absolute values of kurtosis or skewness are more than 2.[29] The kurtosis (k) and skewness (s) of the distribution diagrams were defined as

$$k = \frac{\frac{1}{n}\sum_{i=1}^{n} (\delta r_i - \mu)^4}{\left(\frac{1}{n}\sum_{i=1}^{n} (\delta r_i - \mu)^2\right)^2} - 3 \tag{4}$$

$$s = \frac{\frac{1}{n}\sum_{i=1}^{n} (\delta r_i - \mu)^3}{\left(\frac{1}{n}\sum_{i=1}^{n} (\delta r_i - \mu)^2\right)^{1.5}} \tag{5}$$

where δr_i and μ represent the sample value of the jerk index and the corresponding mean value, respectively. n is the number of indices considered for investigating the form of the probability distribution.

RESULTS AND DISCUSSION

A typical acceleration (3D) and position (3D) of picking a pin in the O'Connor Finger Dexterity test are depicted in Figure 4. The normalized jerk index based on the procedure defined in Data Processing section was calculated for each motion.

Table 1 shows the mean values of jerk indices of each group of participants. Each number provides the information of all 36 trial runs (4 participants × 9 trials) for each group. Box and whisker plots were used to investigate the distributional characteristics of jerk indices obtained from each experiment. Figure 5 shows a comprehensive range of jerk indices (from minimum to maximum values) for each group of participants and all 3 tasks A, B, and C.

As observed in Table 1 and Figure 5, in all 3 tasks, the performance of the assigned tasks by surgeons was accomplished with a lower jerk index. In contradistinction, gamers (in tasks A and B) and engineers (in task C) show the highest jerk index. The results showed that even the maximum values of the jerk indices, quantified from residents and surgeons, were still less than the minimum values of those of engineers and gamers. Therefore, there is a significant difference between the levels of jerk index in engineers/gamers and residents/surgeons, indicating that the residents and surgeons were able to perform the tasks better than gamers and engineers. Because gamers have a high degree of manual dexterity, we expected their performance to be much closer to that of surgeons and residents. We also found that the variability of the data among surgeons was much less than that found in the other groups. Values of kurtosis and skewness are also listed in Table 1. Both kurtosis and skewness would later be used to determine the type of the distribution that fits the quantified jerk indices.

One-way ANOVA was conducted to evaluate whether type of the task accomplished (i.e., the difficulty of the task) as well as how the tool was held (anchored and unanchored) had a significant effect on the measured jerk index. For the 4 groups of participants, in task A, there was a significant difference in the mean values of jerk indices. More specifically, a p value (p) of less than 10^{-5}, while the F value (F) and critical value of F (F_{crit}) were 3.67 and 3.47, respectively ($p \ll 0.05$, $F = 3.67 > F_{crit} = 3.47$. Similarly, in the performance of task B, the p value was much less than 0.05 ($p \ll 0.05$), and $F = 2.67 > F_{crit} = 1.35$. The jerk indices calculated for task C indicated a $p < 0.05$ and $F = 2.67 > F_{crit} = 2.21$. The results showed that, for all 3 tasks, the value of jerk index was affected by the category of participants.

Another ANOVA was conducted to determine whether the value of jerk index was different in all 3 tasks for each group. These results are presented in Table 2. We found a significant difference between the jerk indices obtained for each group during the performance of the 3 tasks. Therefore, the mean value of the jerk indices changed depending on the task accomplished.

We also investigated the trend of increasing/decreasing the jerk indices quantified during the performance of each

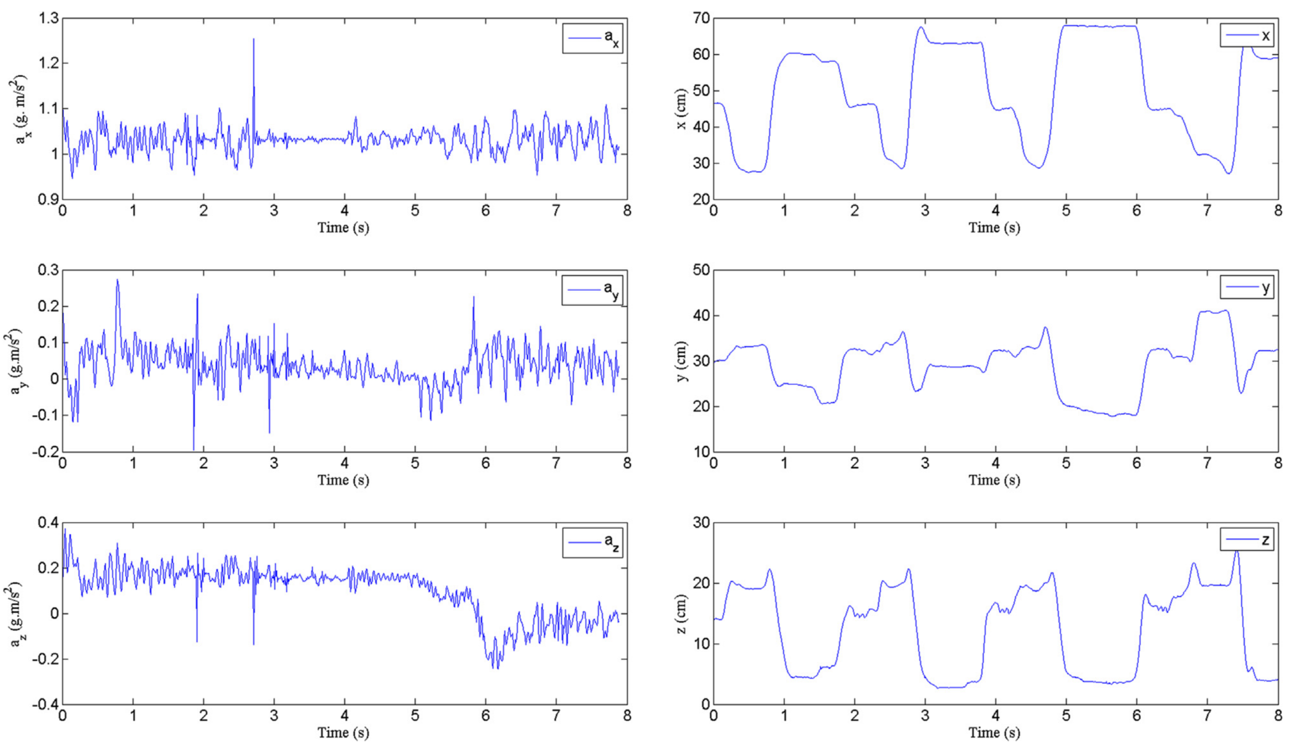


FIGURE 4. A sample acceleration signal (a_x, a_y, a_z) and a sample position signal (x, y, z) recorded by the instrumented forceps.

TABLE 1. Statistical Measures of Jerk Indices for Each Group of Participants

	$\mu \pm \sigma$			Max			Min			*k*			*s*		
Task	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
Engineers	1197 ± 90	1545 ± 77	1867 ± 67	1340	1712	1968	1020	1319	1645	0.82	1.22	2.3	0.22	0.42	1.01
Gamers	1298 ± 64	1594 ± 92	1812 ± 91	1392	1763	2020	1210	1393	1646	1.66	0.12	0.73	0.14	0.24	0.26
Residents	847 ± 107	935 ± 78	1140 ± 68	1110	1088	1320	690	775	1011	−0.66	−0.61	0.51	0.32	0.14	0.80
Surgeons	767 ± 56	860 ± 54	1043 ± 41	869	959	1151	663	755	967	0.93	−0.65	0.34	0.28	0.03	0.45

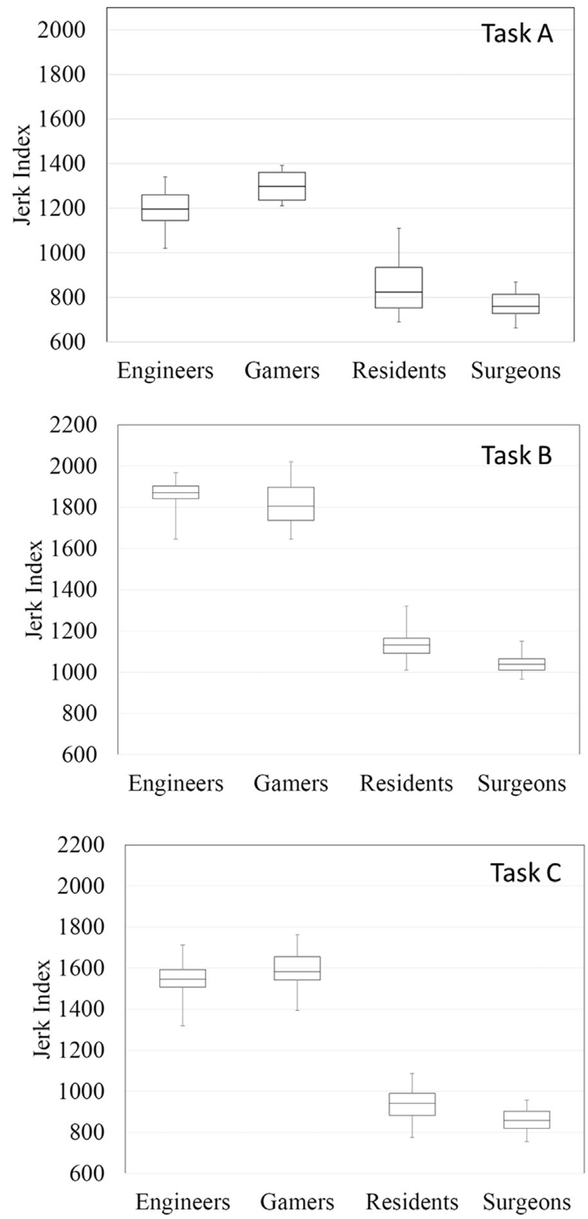


FIGURE 5. Box and whisker plots indicating the distribution of jerk indices measured for each group of participants and the 3 tasks.

task to see whether there was an improvement in participant performance as they became more familiar with experiment procedure. The study showed that the learning curve did not have a significant effect on the values of jerk index, and we could not identify any significant improvement from

TABLE 2. Results of ANOVA Test Conducted for Each Group During Performance of 3 Tasks

Group	p	F	F_{crit}
Engineers	$\ll 0.05$	2.36	3.09
Gamers	$\ll 0.05$	2.91	3.09
Residents	$\ll 0.05$	2.67	3.09
Surgeons	$\ll 0.05$	2.12	3.09

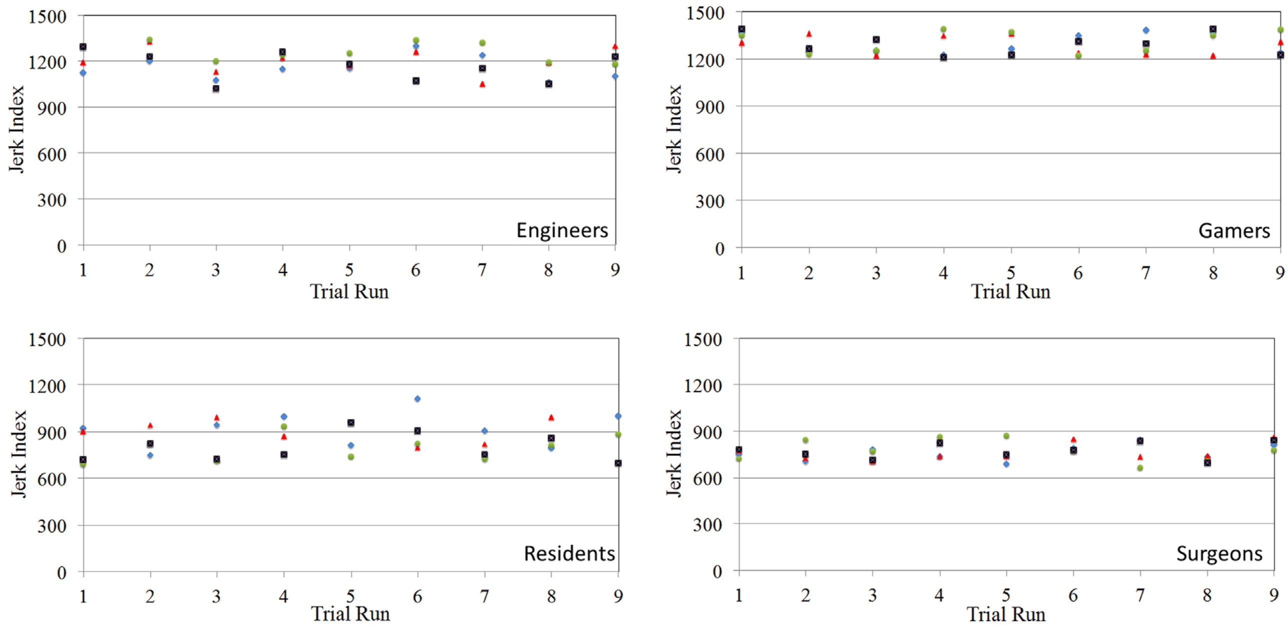


FIGURE 6. Variations of jerk index during performance of task A from trial 1 to trial 9 for all participants. Each set of marker shows participant 1 to 4 in each group.

trial 1 through trial 9. This may relate to the fact that all participants were well-trained before commencing the experiments; therefore, conducting more trial runs did not help them improve their ability. Figure 6, for example, shows the jerk index quantified for each trial of task A and shows that the jerk index does not change from trial 1 through trial 9.

The purpose of designing the 3 tasks was to examine whether the level of difficulty or type of handling the tool could affect the task performance, as well as to investigate which group of participants performed the given tasks with

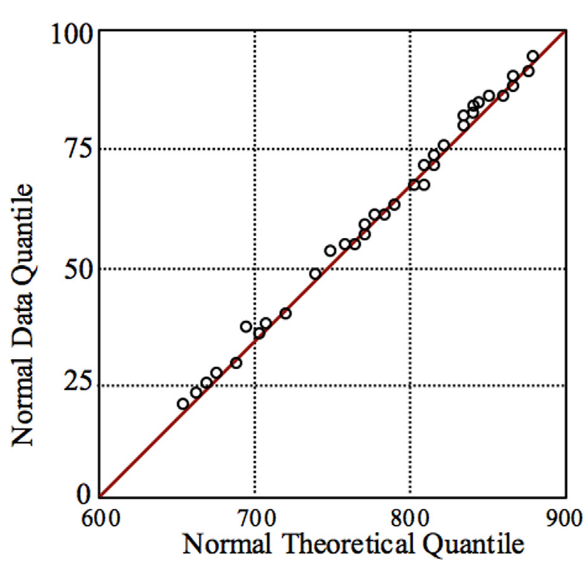


FIGURE 7. Typical Q-Q plot of jerk indices of surgeons, obtained during the performance of task A.

minimal hand vibration. It was important to ensure that the results were repeatable, that is, the same trait would be observed in future tests. The form of distribution diagram was used to predict the future behavior of results. It is well known that normally distributed data exhibit predictable traits and probabilities.[30] The random variable, under which the results were analyzed, was jerk index obtained during performance of each task.

Quantile-Quantile (Q-Q) plots were used to evaluate the type of distribution. Q-Q plots provide graphical information about statistical properties such as mean and median values of random variable. A Q-Q plot draws a theoretical line through the data points and evaluates how the actual data points adhere to the theoretical normal distribution.

In addition to Q-Q plots, kurtosis and skewness measures were used to determine the type of distribution. The kurtosis or skewness values of more than 2 indicate that data groups differ or skew to a significant degree.[31] Values of kurtosis and skewness are listed in Table 1. As observed, all measures are bounded between −2 and 2 that indicate that the distribution of data may follow a normal distribution. As observed, measures of experiments conducted by surgeons and residents were closer to zero as compared to engineers and gamers. There were 12 (4 participants × 3 tasks) sets of Q-Q plots. Figure 7 depicts a typical plot representing the graphical information of the surgeons conducting task A. As observed, the jerk indices of this set of experiments followed a pattern very close to the line, indicating that jerk indices are normally distributed. The jerk data, obtained for surgeons and residents, showed a similar linear pattern. Therefore, in future experiments repeated by surgeons or residents with the same level of expertise, there is a higher probability than engineers and gamers to observe values within the ranges that we obtained in this study. However, for gamers and engineers (with the same level of expertise), we may observe a different pattern of jerk index.

This study shows that normalized jerk index can quantify motion smoothness and can be used to assess hand motion dexterity of surgical trainees (both in micromotions and normal motion). This quantifiable normalized jerk index has the potential to assess levels of hand dexterity as a part of automated surgical training simulators as a metric to assess dexterity. Although we did not use it to improve the performance of the surgeons, by taking variables such as path length, variability in the direction of motion, and acceleration into account, it may be possible to instruct the novice toward altering/improving performance. Furthermore, the jerk index might also be used as a determinant in the selection process toward identifying skilled surgical candidates at the baseline level.

CONCLUSIONS

Technical surgical skills are acquired through years of hands-on training in the OR, supplemented by written anatomy examinations, tutorials, and laboratory-based surgical skill courses using cadavers or models. More recently, VR-based simulations have also been included in this paradigm. Although knowledge of anatomy can be assessed by written or oral examination, assessment of technical skills such as dexterity and tool handling remains more subjective than objective. Assessing surgical competence has several potential benefits, including safety of surgical training process, accreditation of specialists, and maintenance of public confidence in the surgical profession. This is particularly important as surgical training shifts to a competency-based education paradigm. Currently, surgical skills are mostly assessed subjectively whereby experienced surgeons educate and observe residents and provide feedback. Objective assessment of surgical skills is an important component for the structured learning of surgical skills, which provides feedback for training. This type of evaluation has not been widely adopted into clinical practice in part due to the high cost of simulated teaching modules, which while effective, may not be ideal given the lack of reliable metrics that accurately reflect technical skills. Motion smoothness in handling surgical tools is an essential skill that surgical residents must acquire before independently operating on patients. Motion smoothness is recognized as an indicator of skilled and coordinated hand motion. In this study, a test setup of 2 modified O'Connor Dexterity board and a Tweezer Dexterity pegboard (placed inside a phantom skull) were developed along with an instrumented bipolar forceps to quantitatively measure motion quality and smoothness of different groups (surgeons, residents, gamers, and engineers). The acquired acceleration and position signals were processed to determine the normalized jerk index as an indicator of this skill. Our results indicated that experienced surgeons completed all 3 tasks with a lower jerk index. Moreover, the residents experienced motions more similar to the surgeons compared to the engineers and gamers. It can be speculated that such a methodology might help assess the skillset of surgical residents as they, through the course of time, attain surgical competency. Such a model, in theory, could help to determine the level of competency of different trainees, allowing those with optimal dexterity to progress to a higher level of training. Furthermore, the jerk index might also be used as a determinant in the selection process toward identifying skilled surgical candidates at the baseline level.

FUNDING

This research was funded by the Ohlson Research Initiative and Project neuroArm at the University of Calgary and Canada Foundation for Innovation.

AUTHOR CONTRIBUTIONS

Ahmad Ghasemloonia and Yaser Maddahi had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Garnette Sutherland and Kourosh Zareinia.

Acquisition, analysis, or interpretation of data: Ahmad Ghasemloonia and Yaser Maddahi.

Drafting and editing of the manuscript: Ahmad Ghasemloonia, Yaser Maddahi, Kourosh Zareinia, Joseph Dort, and Sanju Lama.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Yaser Maddahi.

Obtained funding: Joseph Dort and Garnette Sutherland.

Administrative, technical, or material support: Kourosh Zareinia, Ahmad Ghasemloonia, and Sanju Lama.

Study supervision: Sanju Lama, Joseph Dort, Kourosh Zareinia, and Garnette Sutherland.

ACKNOWLEDGMENT

The authors would like to thank the Killam Trusts, whose generous support for Killam laureate Yaser Maddahi allowed his contribution to this work.

REFERENCES

1. Holmboe ES, Sherbino J, Long DM, Swing SR, Frank JR. The role of assessment in competency-based medical education. *Med Teach*. 2010;32(8):676-682.

2. Grantcharov TP, Reznick RK. Training tomorrow's surgeons: what are we looking for and how can we achieve it? *ANZ J Surg*. 2009;79(3):104-107.

3. Judkins TN, Oleynikov D, Stergiou N. Objective evaluation of expert and novice performance during robotic surgical training tasks. *Surg Endosc*. 2009;23 (3):590-597.

4. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273-278.

5. Sharma Y,Plötz T, Hammerld N, et al. Automated surgical OSATS prediction from videos. Proceedings of the 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI 2014); 29 April-2 May, Beijing, China.

6. Niitsu H, Hirabayashi N, Yoshimitsu M, et al. Using the objective structured assessment of technical skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surg Today*. 2013;43(3):271-275.

7. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg*. 2005;190(1):107-113.

8. Sroka G, Feldman LS, Vassiliou MC, Kaneva PA, Fayez R, Fried GM. Fundamentals of laparoscopic surgery simulator training to proficiency improves laparoscopic performance in the operating room—a randomized controlled trial. *Am J Surg*. 2010;199(1): 115-120.

9. Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Assessment of technical surgical skills. *Eur J Surg*. 2002;168(3):139-144.

10. Reznick RK, MacRae H. Teaching surgical skills changes in the wind. *N Engl J Med*. 2006;355(25):2664-2669.

11. Gallagher G, Ritter EM, Champion H, et al. Virtual reality simulation for the operating room proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg*. 2005;241(2):364-372.

12. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J. Objective assessment of technical surgical skills. *Br J Surg*. 2010;97(7):972-987.

13. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. *Br Med J*. 2003;327(7422):1032-1037.

14. Sinigaglia S, Megali G, Tonet O, Pietrabissa A, Dario P. Defining metrics for objective evaluation of surgical performances in laparoscopic training. *Proceedings of the 19th International Congress and Exhibition Computer Assisted Radiology and Surgery*. 2005;128: 509–514.

15. Uemura M, Tomikawa M, Kumashiro R, et al. Analysis of hand motion differentiates expert and novice surgeons. *J Surg Res*. 2014;188(1):8-13.

16. Chmarra MK, Kolkman W, Jansen FW, Grimbergen CA, Dankelman J. The influence of experience and camera holding on laparoscopic instrument movements measured with the TrEndo tracking system. *Surg Endosc*. 2007;21(11):2069-2075.

17. Hogan N, Sternad D. Sensitivity of smoothness measures to movement duration, amplitude, and arrests. *J Mot Behav*. 2009;41(6):529-534.

18. Carpinella I, Cattaneo D, Ferrarin M. Quantitative assessment of upper limb motor function in multiple sclerosis using an instrumented action research arm test. *J Neuroeng Rehabil*. 2014;11:67.

19. Choi A, Joo SB, Oh E, Mun JH. Kinematic evaluation of movement smoothness in golf: relationship between the normalized jerk cost of body joints and the clubhead. *Biomed Eng Online*. 2014;13(1):20.

20. Takada K, Yashiro K, Takagi M. Reliability and sensitivity of jerk-cost measurement for evaluating irregularity of chewing jaw movements. *Physiol Meas*. 2006;27(7):609-622.

21. Cotin S, Stylopoulos N, Ottensmeyer M, Neumann P, Rattner D, Dawson S. Metrics for laparoscopic skills trainers: the weakest link!. Dohi T, Kikinis R, editors. Medical Image Computing and Computer-Assisted Intervention. Berlin, Germany: Springer Berlin Heidelberg, 2002:35-43.

22. Chmarra MK, Klein S, de Winter JCF, Jansen FW, Dankelman J. Objective classification of residents based on their psychomotor laparoscopic skills. *Surg Endosc*. 2010;24:1031-1039.

23. Sakata K, Kogure A, Hosoda M, Isozaki K, Masuda T, Morita S. Evaluation of the age-related changes in movement smoothness in the lower extremity joints during lifting. *Gait Posture*. 2010;31(1):27-31.

24. Maddahi Y, Greene M, Gan LS, et al. Performance evaluation of a surgical telerobotic system using kinematic indices of the master hand-controller. *Lect Notes Comput Sci*. 2014;8619:167-175.

25. Maddahi Y, Ghasemloonia A, Zareinia K, Sepehri N, Sutherland GR. Quantifying force and positional frequency bands in neurosurgical tasks. *Robot Surg*. 2016;10(2):97-102.

26. Yashiro K, Nakamura T, Mizumori T, Yatani H. Clinical validity of measuring jerk-cost of jaw movement during speech: effect of mouthguard design on smoothness of jaw movements. *Proceedings of the SICE 2004 Annual Conference*. 2004;1:93–96.

27. Rosenberg LB. Virtual haptic overlays enhance performance in telepresence tasks. *Telemanipulator Telepresence Technol*. 1995;235:99-108.

28. Wuensch KL. Encyclopedia of Statistics in Behavioral Science. Chichester, UK: Wiley; 2005.

29. Maddahi Y, Sepehri N, Maddahi A, Abdolmohammadi M. Calibration of wheeled mobile robots with differential drive mechanisms: an experimental approach. *Robotica*. 2012;30(6):1029-1039.

30. Uffink J. Can the maximum entropy principle be explained as a consistency requirement? *Stud Hist Phil Mod Phys*. 1995;26(3):223-261.

31. Aggarwal YP. Statistical Methods: Concepts, Application and Computation. 3rd ed. New Delhi: Sterling Publishers; 2008.