Computer Methods in Biomechanics and Biomedical Engineering, Vol. 4, pp. 515–523
Reprints available directly from the publisher
Photocopying permitted by license only

Motion Analysis in the Assessment of Surgical Skill

VIVEK DATTA[a], SEAN MACKAY[a], ARA DARZI[a] and DUNCAN GILLIES[b,*]

[a]*Imperial College School of Medicine*, [b]*Imperial College Department of Computing, 180 Queen's Gate, London SW7 2BZ*

(Received 7 August 2000; In final form 6 February 2001)

Manual skill is now widely recognised as an important aspect of training in surgery. However, measurement of the skill of a surgeon has in the past been rather subjective in nature, relying on the judgement of experts in the analysis of videotapes. Objective measurements can be made by analysing the velocities of a surgeon's hands during a procedure. In particular, we have found that the number of movements made during a typical procedure will decrease as the surgeon's skill increases. Velocity traces display purposeful movements corrupted by uncorrelated noise from sources such as hand tremor and measurement artefacts. However, we have found that it is possible to filter the noise effectively. Furthermore, we have shown that the skill measurement obtained by counting movements is highly robust to over or under filtering.

Keywords: Motion tracking; Surgical skills; Skill assessment

1. INTRODUCTION

Technical ability is an important part of overall competence in surgery [1, 2]. In recent years, there has been an increase in peer and public scrutiny of surgical performance [3]. Although very rare, incidents where surgeons have been found to endanger patients through lack of technical competence have attracted widespread attention [4]. As a consequence there has been increased attention to skills training within surgical residency programs. Judging a surgeon's manual dexterity is difficult and comparison is notoriously difficult, due in no small part to the individuality of patients [5]. What assessment there has been is highly subjective, a good surgeon perceived as having a greater economy and precision of movement. This has highlighted the need for, amongst other things, a more quantitative and objective way of measuring a surgeon's dexterity [6, 7]. However, the only current available method to measure skill quantitatively is to equate dexterity with the time taken to complete a task. This, though, is crude and not necessarily an accurate measure of skill and efficiency.

Over the last few years computer technology is increasingly being used to help train surgeons. Parallels have been drawn between the training of

*Corresponding author. e-mail: dfg@doc.ic.ac.uk

surgeons and pilots. Computer based flight simulation is an integral part of training within the airline industry [8]. Virtual Reality simulators have been used to model the skills needed to perform laparoscopic (keyhole surgery) [9, 10, 11]. Built into these systems is the ability to measure the kinematic motion of the surgical instruments accurately and so to measure the number of movements taken, and the economy of instrument travel for a set task. These parameters have been validated as accurate, objective measures of dexterity in a virtual environment and in laparascopic surgery [12]. It would therefore seem appropriate that the same measurements, could they be made during actual surgical procedures, could form the basis of an objective measure of surgical skill. Human motion measurement has been used for some time using markers located on body articulations [13]. It is now possible, using electromagnetic tracking devices, to determine the position of a surgeon's hands while they perform a real physical task. Our research aims to show the feasibility and validity of generating measures of movement efficiency from these raw positional data.

2. METHODOLOGY

Participants were asked to perform a laboratory-based model of a surgical procedure. Subjects were recruited from different experience levels, senior house officers to consultant surgeons. The task involved applying a synthetic vein patch into synthetic artery (both made by Annexart, UK) using a continuous suturing technique. Monofilament polypropylene (Surgipro, USSC) was used for the suture material. All subjects were asked to use exactly the same technique, as taught on the Royal College of Surgeons of England Basic Surgical Skills Course [14]. The procedure was video taped and assessed for consistency of performance [15]. Hand position and movements were measured using an electromagnetic tracking system (Polhemus Inc. Isotrak II), with trackers applied to the back of the participants' hands. The trackers were attached by velcro strips. The position was centered at the mid shaft point of the third metacarpal on the dorsum of each hand. Latex gloves were worn by the subject to ensure that there was no slippage of the sensor. The tracker has a spatial resolution of better than 1mm. Samples were taken every 50 ms of the positions of each hand, in three dimensions. From these data it was possible to calculate the overall total distance moved by each hand to complete a task, the overall time, and hence the average velocity. At each sample point an instantaneous velocity was calculated. In defining purposeful movements it is the velocity that is important. Hypothetically we would expect for a given movement that the velocity would increase in magnitude as the hand moves smoothly to the given destination. The velocity would reach a maximum and then fall to near zero as the movement was completed. This pattern is clearly seen in Figure 1 which shows a typical plot of velocity data. In this case the subject was asked to make ten deliberate movements. Assuming that surgical technical performance improves with experience, the hypothesis that we would like to validate is that the number of purposeful movements is a good measure of surgical skill.

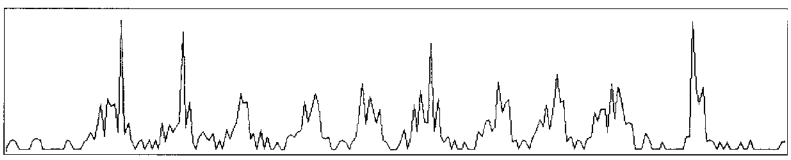

FIGURE 1 The velocity profile of the left hand performing ten deliberate movements. The vertical axis is velocity and the horizontal is time.

Figure 1 also shows that, in addition to the large clear movements there are smaller apparent movements that may or may not be considered purposeful. Moreover, at the detailed level, changes in recorded velocity are not smooth. There are two major factors that account for these effects. The first is hand tremor and the second experimental error in the tracking device. Both of these effects are hard to quantify and tend to be irregular rather than falling into any predictable pattern. Since the tracking device is electromagnetic its accuracy can be affected by the presence of other fields in the laboratory, and by the proximity of metallic objects. These are factors that cannot be controlled easily, particularly in an operating theatre. Hand tremor can similarly be affected by psychological factors and by concentration. Thus we are faced with devising a suitable way of filtering the unwanted movements to measure only what is deliberate.

Since most of the unwanted noise that we wish to eliminate from the images is high frequency, that is to say fast changing compared with the movements that we would like to measure, the most appropriate filter to use is a simple low pass filter. We tested four of these, namely the box filter, the Gaussian filter, the second order Butterworth filter and the fourth order Butterworth filter. These were all computed by taking a weighted local average at each point in the velocity trace, thus blurring its fine detail. The weights are derived from the filters, which are plotted in Figure 2.

The box filter is the simplest to compute, being just a simple local average, but does not provide an accurate or smooth frequency cut off. The Butterworth filters give a highly accurate and well controlled frequency cut off, but unfortunately they have an oscillating shape when computed. This makes them more prone to sampling errors, and thus computationally less desirable than the Gaussian filter. All four filters gave comparable results, and so the Gaussian was adopted due to its good compromise between accurate frequency cut off and computational stability. The degree to which a data set is filtered depends on the width of the filter used. If we express the width of the Gaussian filter in samples per standard deviation this gives us an indication of how local the average is in practice. Thus, a filter width of 10 samples per standard deviation means that about 1 second worth of data contributes to a local average.

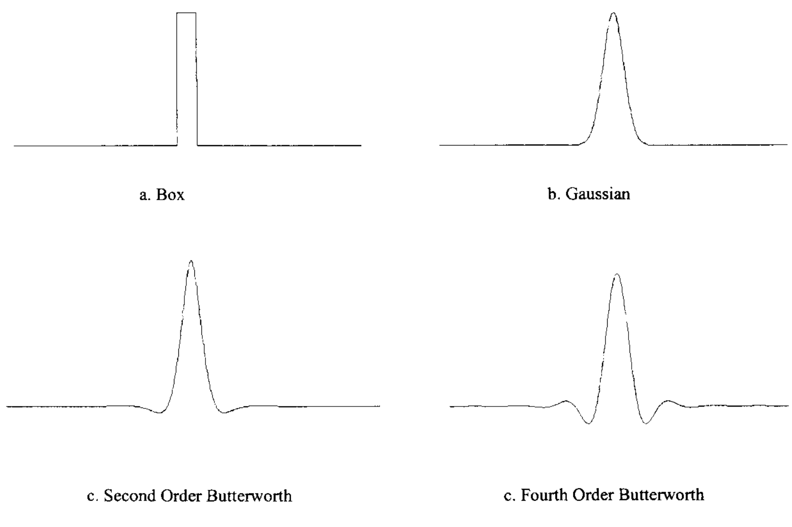

FIGURE 2 The characteristics shapes of the filters tested plotted in the time domain. Note the oscillations in the Butterworth filters which make them less suitable for discrete computation.

No low pass filters will remove entirely unwanted noise, and therefore some velocity threshold was required to eliminate entirely the small residuals remaining after the main filtering operation. To set this threshold we carried out a number of calibration experiments. These included measuring a movement trace with the trackers on a laboratory bench, and on the hands of a volunteer. Several sets of data were recorded with the subject keeping one hand still and making deliberate movements with the other. These were then filtered using a range of different velocity threshold, varying from 0.5 cm/sec to 3 cm/sec, and a range of filter widths going from 5 samples/standard deviation to 20 samples/standard deviation. The results showed a narrow operating band for the system. If the velocity threshold was set below 1.5 cm/second, then it was not possible to find a suitable size for the Gaussian filter. This was because the width required to remove the noise was such that purposeful movements were lost. Setting the velocity threshold above 1.5 cm/sec some purposeful movements were removed for all settings of the Gaussian filter. The most difficult data in this respect were the traces where the subject was asked to make ten fine movements with one hand while keeping the other still. A perfect filter could not be found, but the best compromise was to set the velocity threshold at 1.5 cm/sec and the Gaussian filter at 12 samples/standard deviation. Figure 3 shows a trace used in the calibration experiments.

The next experiments conducted were to test whether the filters were giving us statistically meaningful results. To do this we fixed the velocity threshold at 1.5 cm/sec and calculated the number of movements found as a function of the width of the filter. The behaviour of the filters for a typical experiment is shown in Figure 4. It will be noticed that the Gaussian and Butterworth filters give almost identical results. The Box filter is set to a width equivalent to one standard deviation of the Gaussian. Thus its effect is more local with less samples participating in the average, which accounts for the higher movement counts. Its performance is less smooth than the other filters. Generally, it will be seen that there is a dramatic fall off in the number of movements counted up until around a width of 15 samples per standard deviation. From a width of 15 up to 90 there is a smooth change as successively smaller movements are eliminated, and it seems almost arbitrary where we place the cut off. To help us to do this we adopted a purely statistical procedure to find where the filter would provide the best separation of our data. For each data file we calculated the

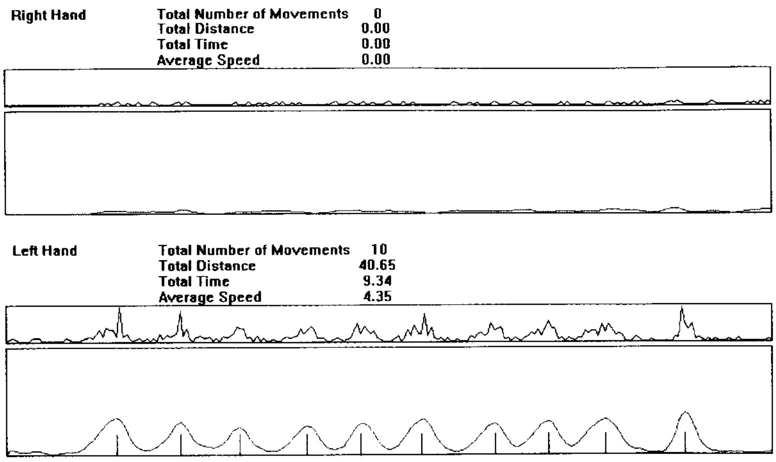

FIGURE 3 Experimental trace used for calibration. The subject was asked to keep the right hand still and make ten deliberate movements with the left hand. The upper traces show the actual measured velocities, and the lower ones the filtered data.

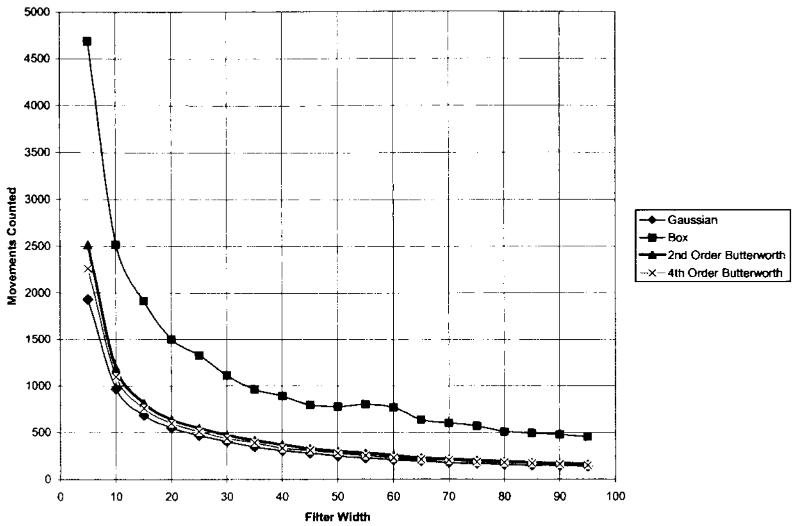

FIGURE 4 Movements counted as a function of filter size.

number of movements using a range of widths of the filter. We can write these as:

$$m_1, m_2, m_3 \cdots m_n$$

where the index represents different filter widths. Thus m_1 is the number of movements found with the filter set to 5, m_2 the number of movements found for the filter set to 10 and so on. We then normalised the results by computing:

$$n_i = \frac{m_i \times 1000}{\sum_j m_j}$$

The reason for doing this was to adjust each file to a standard number of movements eliminating the spread due to the length of the procedure and the skill of the subject. Thus the remaining variance in the system should be independent of the person carrying out the procedure, and the procedure itself. This allows us to calculate a mean and variance for the normalised results of all our data at each of the filter widths. We propose that the best setting for the filter width is the one where the variance divided by the mean is minimum. That is to say where we have the largest measurement relative to the expected variation. The results are plotted in Figure 4. It will be seen that there is a dramatic fall in this measure as the width increases, and that this levels out and then rises slightly. The minimum falls at a filter width of around 35 samples per standard deviation. This is much higher than the width used in the calibration experiments, and we know that at that level we will lose purposeful movements. Conversely, with the filter set to 12, the mean to variance ratio is around 0.3. This is equivalent to a standard deviation of around 5.5% of the mean, which will allow us to separate trainees into different categories of ability as required.

3. EXPERIMENTAL RESULTS

A set of data was collected, as described in Section 2, on a practical laboratory surgical skills exercise. The participants ranged from beginners to consultants, and consequently had widely varying abilities. For the purposes of validating the filter settings, a very crude measure of skill, namely the approximate number of years of experience, was adopted. A typical velocity trace of the whole experiment is shown in Figure 6. The total length of this particular procedure ranged from 9 to 15 mins.

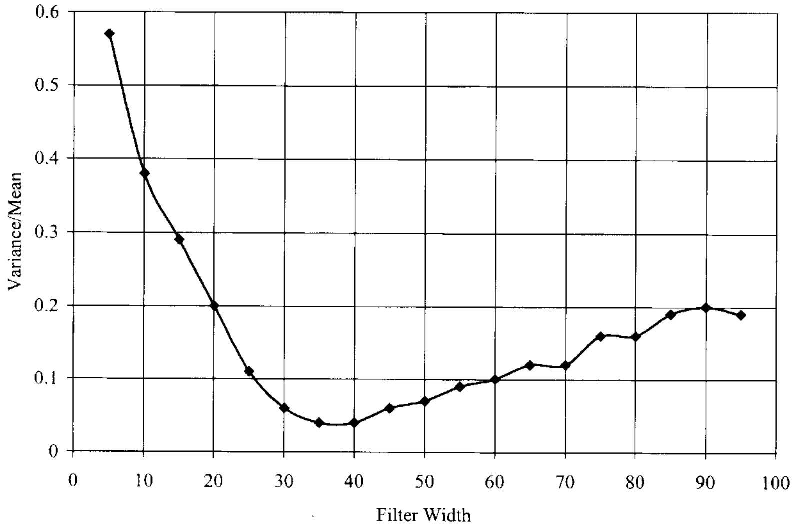

FIGURE 5 Variance divided by mean as a function of filter size. The Gaussian filter is shown.

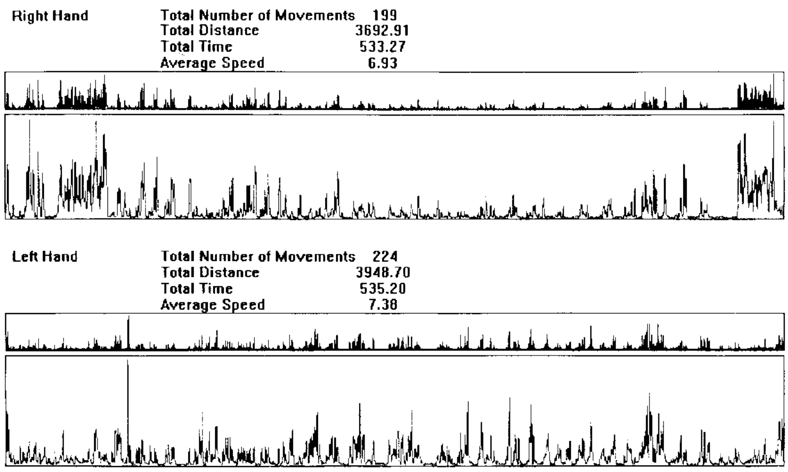

FIGURE 6 Velocity traces for a complete procedure carried out by a skilled consultant.

Figure 7 shows a scatter plot of the number of movements obtained with the filter width set at 12 samples per standard deviation, against the number of years of experience of the participants. Naturally there are outliers, some novices showing excellent natural aptitude for the procedures. However there is a clear trend in the data, and the Pearson correlation coefficient is high, around -0.7. This is a significant result with a P value of less than 0.01.

This gives us confidence that the filters are working correctly, and that we are indeed measuring what we need. In order to test the robustness of the system, the correlation coefficient was calculated for a range of filter sizes. The results are shown in Figure 8. It will be seen that for very low values of the filter width the results are unpredictable. However for values from 10 to 40 there is a stable region, indicating that the results are very robust against changes in the filter size. The actual

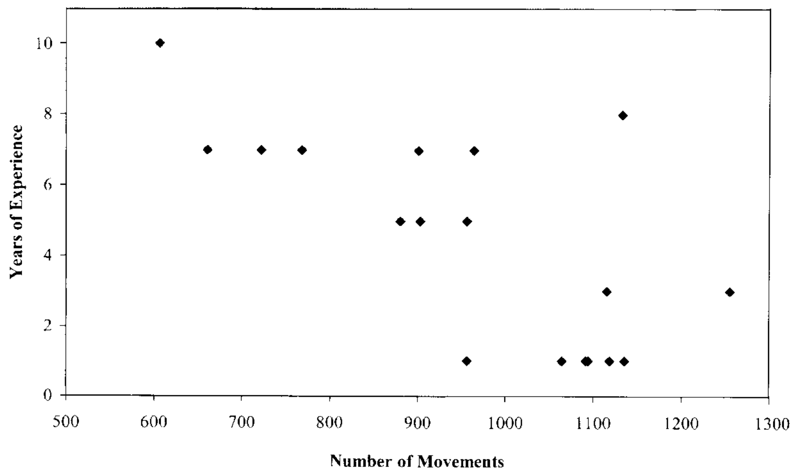

FIGURE 7 Scatter plot of the number of years of experience against the movements made in completing a surgical procedure.

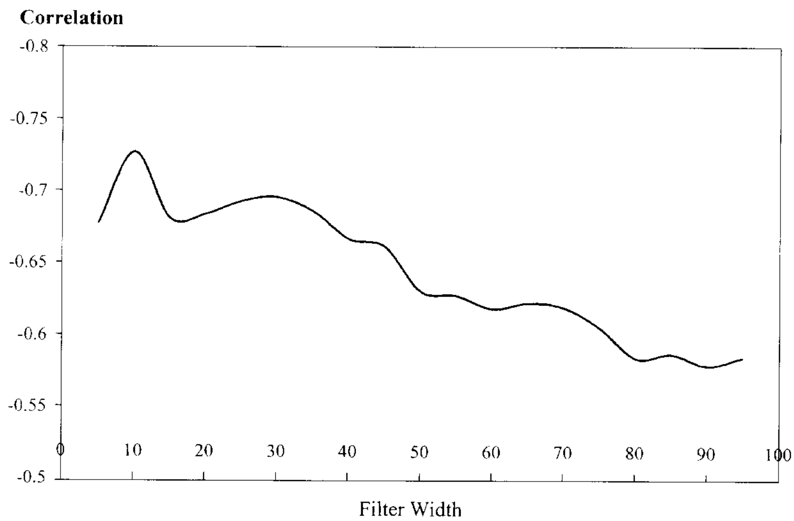

FIGURE 8 The effect on correlation of filter size.

ordering of the participants did not change over this range either.

4. CONCLUSIONS AND PROPOSALS FOR FURTHER WORK

Our intention in this paper was to explore the use of filters to that allow us to count the number of movements of a surgeon's hands during a surgical procedure. The results show that it is possible to set filters on the velocity traces that give a reasonable response on calibration data. However the margins over which we obtain a correct result in all circumstances are narrow. Statistical testing however shows that our expected results are robust to changes in filter size. When considering the variance introduced solely by the measurement process we find that the standard deviation is

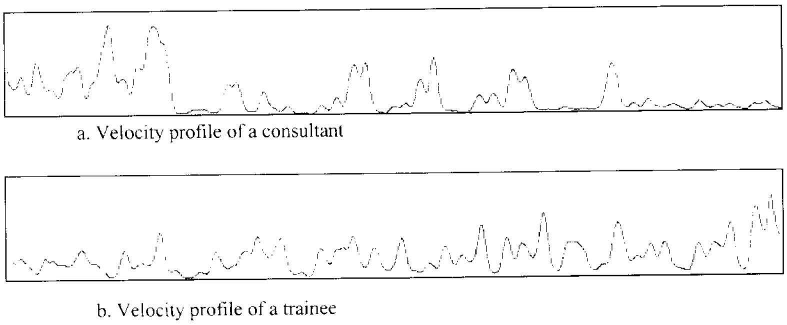

FIGURE 9 Comparison of a skilled consultant with a novice. Notice that the consultant makes fewer movements with the hands falling near to zero velocity between them. Each trace is of the right hand over a period of 100 seconds, and is filtered with a Gaussian filter set to 40 samples standard deviation.

small over a wide range of filter widths, and in actual usage the results we obtain are similarly robust.

This work is part of a major ongoing study into skills assessment in surgery. One arm of this has been video checklist scoring, a technique previously validated. In our department, this technique has been used with the scoring performed from a video recording of the procedure. Early results show an excellent correlation between motion and video analysis, lending further weight to our proposition that the former is an effective method of quantifying dexterity. A further interesting aspect that we plan to investigate is the use of other measurable features of the velocity. Figure 9 shows two traces of approximately 100 seconds taken at the beginning of a procedure. The upper trace is taken from a consultant, and it will be seen that the movements are generally shorter with long periods of rest in between. The lower trace, from a trainee shows overall more movement with less clearly defined times of inactivity.

We are, of course, a long way from turning observations of this kind into meaningful measures, and we expect that we will find a range of different styles in which skilled practitioners carry out these procedures. However, there is clearly potential for finding improved measures based on these characteristics.

References

[1] Spencer, F. (1978). "Teaching and measuring surgical techniques – the technical evaluation of competence". *Bulletin of the American College of Surgeons*, **63**(3), 9–12.

[2] Webb, W. R. and Peacock, E. E., "Success of a properly planned surgical procedure". *Ann. Surg.*, **191**(3), 388–90, 1980 Mar.

[3] Smith, R. (1998). "All changed and utterly. British medicine will be transformed by the Bristol case". *BMJ*, **316**, 1917–18.

[4] Marshall, M. N., Shekelle, P. G., Leatherman, S. and Brook, R. H. (2000). "Public disclosure of performance data: learning from the US experience". *Qual. Health Care*, **9**, 53–57.

[5] Kirby, T. J. (1979). "Dexterity testing and resident's surgical performance". *Trans. Am. Ophthalmol. Soc.*, **77**, 294–307.

[6] Graham, K. S. and Deary, I. J., "A role for aptitude training in surgery". *J. R. Coll. Surg. Edinb.*, **36**(2), 70–4, 1991 Apr.

[7] Schuenman, A. L., Pickleman, J., Hesslein, R. and Freeark, R. J., "Neuropsychologic predictors of operative skill among general surgery residents". *Surgery*, **96**(2), 288–95, 1984 Aug.

[8] Muller, M. H. (1999). "How does aviation find the ideal pilot? Suitability testing: applicability to surgery? Methods for determining basic occupational suitability". *Zentralbl. Chir.*, **124**(10), 889–94.

[9] Sutton, C., McCloy, R., Middlebrook, A., Chater, P., Wilson, M. and Stone, R. (1997). "MIST VR. A laparoscopic surgery procedures trainer and evaluator". *Stud. Health Technol. Inform.*, **39**, 598–607.

[10] Taffinder, N., Sutton, C., Fishwick, R. J., McManus, I. C. and Darzi, A., "Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery". *Proceedings of Medicine Meets Virtual Reality*, In: Westwood, J., Hoffman, H., Stredney, D. and Weghorst, S. Editors. *Technology and Informatics 50*, 1998. Jan. 28–31: San Diego, USA. Amsterdam: IOS Press 1998: pp. 124–130.

[11] Taffinder, N., McManus, I. C., Gul, Y., Russell, R. C. and Darzi, A., "The effect of sleep deprivation on surgical

dexterity on a laparoscopic simulator". *Lancet*, **352**(9135), 1191, 1998 Oct 10.

[12] Taffinder, N., Smith, S., Mair, J., Russell, R. and Darzi, A. (1999). "Can a computer measure surgical precision? Reliability, validity and feasibility of the ICSAD". *Surgical Endoscopy*, **13**(suppl. 1), 81.

[13] Lugne, P. C., Alizon, J., Collange, F. and Van Praagh, E., "Motion analysis of an articulated locomotion model by video and telemetric data". *J. Biomech.*, **32**(9), 977 81, 1999 Sep.

[14] Royal College of Surgeons of England, Basic Surgical Skills Course Instructors' Handbook.

[15] Martin, J. A., Regehr, G., Reznick, R., MacRae, H., Murnaghan, J., Hutchison, C. and Brown, M., "Objective structured assessment of technical skill (OSATS) for surgical residents". *Br. J. Surg.*, **84**(2), 273–8, 1997 Feb.