Surg Endosc (1998) 12: 1415–1418

Surgical Endoscopy
Ultrasound and
Interventional Techniques

© Springer-Verlag New York Inc. 1998

Laparoscopic visual field

Voice vs foot pedal interfaces for control of the AESOP robot

M. E. Allaf, S. V. Jackman, P. G. Schulam, J. A. Cadeddu, B. R. Lee, R. G. Moore, L. R. Kavoussi

Brady Urological Institute, Johns Hopkins Medical Institutions, 4940 Eastern Avenue, Baltimore, MD 21224, USA

Received: 24 October 1997/Accepted: 2 February 1998

Abstract
Background: In order for robotic devices to be introduced successfully into surgical practice, the development of transparent surgeon/machine interfaces is critical.
Methods: This study evaluated the standard foot pedal for the AESOP robot compared to a voice control interface. Speed, accuracy, learning curves, durability of learning at 2 weeks, and operator-interface failures were analyzed in an ex vivo model.
Results: Foot control was faster and had less operator-interface failures. Voice control was more accurate as measured by ‘‘pass points.’’ The foot control learning curve reached a plateau at the third trial, while the voice control did not fully plateau. Durability of learning favored the foot control but was not significantly different.
Conclusions: Currently, the voice control is more accurate and has the advantage of not requiring the surgeon to look away from the operative field. However, it is slower and may require more attention as an interface. As voice recognition software continues to advance, speed and transparency are anticipated to improve.

Key words: Robot — Laparoscopic surgery — Voice — Foot — Interface

Surgeons who perform laparoscopic procedures typically have to relinquish full control over their operating field of vision because the job of holding the endoscope is delegated to another individual. In addition, the need for an extra assistant may impose substantial financial and logistical burdens [4]. Furthermore, miscommunication between the surgeon and camera operator or even small inaccurate movements due to fatigue or human tremor can be frustrating and potentially dangerous.

In an attempt to address these problems, mechanical devices were developed to hold and move the endoscope. The first generation of laparoscope holders required manual adjustment; they were cumbersome to use and disrupted the flow of the procedure [1, 5]. More recently, computer-controlled robots have been developed to take on this task. One such device is the AESOP robot (Automated Endoscopic System for Optimal Positioning).

The AESOP robot was designed to provide the surgeon with direct and precise control over the visual field while both arms remain free for the delicate maneuvers required in many procedures. The surgeon-robot interface is of paramount importance since it is the means by which the surgeon communicates with and controls the robot. Thus, in order for the system to be practical, the interface should be intuitive, easy to learn, accurate, and as transparent as possible.

Currently, the AESOP robot can be controlled via a hand, foot, or voice control interface. The AESOP robot controlled with the foot pedal has been found to be more effective and accurate than a human assistant in manipulating the laparoscope [3]. The feasibility and the applicability of using foot pedal–controlled robotic arms in lieu of surgical assistants in urologic laparoscopic surgery has been confirmed [4]. In addition, the times required to learn control of the laparoscope manually and with the AESOP foot pedal were found to be equal [2].

There are certain disadvantages to the foot pedal interface. Using the foot pedal requires added concentration because the surgeon must operate using three limbs. The surgeon must often look away from the video monitor in order to reorient his or her feet with the controller. In addition, in critical situations such as excessive bleeding, the surgeon may want to concentrate fully on the situation at hand and therefore abandon the food pedal [4].

The voice-controlled system was developed to be a more intuitive and transparent interface. The voice is a more

---

This material was presented as an abstract for the 1998 annual meeting of the Society for Urology and Engineering.

Correspondence to: S. V. Jackman

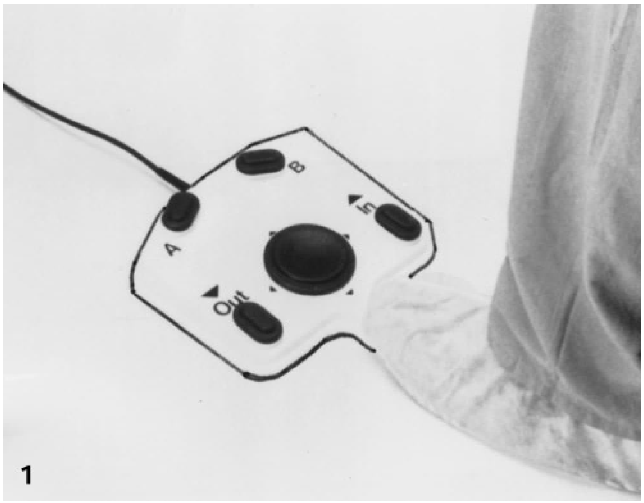

Fig. 1. Foot pedal interface. (Photo courtesy of Computer Motion)

Fig. 2. Voice control interface in use with the AESOP robot. (Photo courtesy of Computer Motion)

natural method of communication, and the use of familiar commands should not require as much concentration. It was hypothesized that the voice controller is a more transparent and functional interface than the foot pedal. We compared the two methods on the basis of speed, accuracy, learning curves, durability of learning at 2 weeks, and operator-interface failures.

Materials and methods

The AESOP 2000 robot (Computer Motion, Goleta, CA, USA) with its foot pedal and voice interfaces was used in this study. The foot pedal is shown in Fig. 1. The voice interface (software version 1.3.2) shown in use with the AESOP robot in Fig. 2 requires a voice-training stage in which the user repeats a set of 23 words six times into a microphone. A customized PCMCIA computer card with the user's speech data is then generated and utilized by the AESOP robot to recognize that user's commands. The voice trainer and recognition computers are proprietary designs of Computer Motion (version 1.07) and were those currently available at the time the study was performed.

A course resembling a multilevel baseball diamond was constructed in a pelvic trainer. It had four bases or targets of differing heights and required manipulation of the laparoscope in three dimensions. A cross hair was affixed to the end of the laparoscope used in these trials. Two attending endoscopic surgeons, one endosurgical fellow, and two surgical residents were asked to direct the laparoscope around the bases using either the foot or voice control interface. A referee timed the course and confirmed that the surgeons stopped with the entire cross hair on each base. This was repeated for a total of five consecutive trials per surgeon, first with the foot pedal and then with the voice interface. The surgeons were allowed to survey the course visually beforehand as well as manually perform the course with the laparoscope to minimize the potential confounding factor of learning the course. The two attendings and fellow had all used the foot pedal clinically approximately twice per week in the 6 months prior to the study. The two residents were familiar with the interface and had used it occasionally.

Fourteen days after the initial trials, the protocol was repeated. Additional measurements were recorded to assess the accuracy and limitations of each interface. As a measure of accuracy, the number of times the surgeons passed a base when intending to stop on it (''pass points'') was recorded for each interface. ''Operator-interface failures'' were defined as events that force the surgeon to divert attention away from the operative field and focus on the interface. They were quantified by the number of times the surgeon looked down to reorient his or her feet during the foot pedal trials and by the number of unrecognized or misrecognized commands during the voice trials.

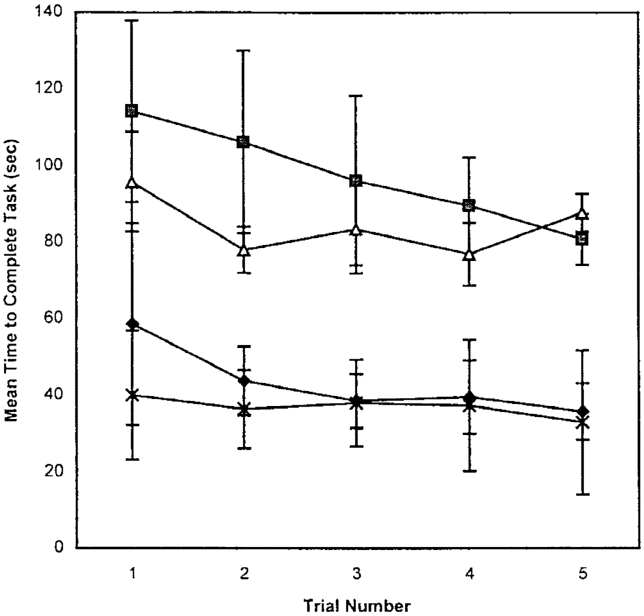

Fig. 3. Learning curves for the voice and foot pedal interfaces on days 1 and 14. Standard error bars are included. $p < 0.05$ for all comparisons between corresponding foot and voice data points. $p = 0.06$ and 0.01 for the difference between trials 1 and 5 on day 1 for foot and voice, respectively. (—◆—), Foot day 1; (—☐—), voice day 1; (━━), voice day 14; (—X—), foot day 14.

Statistical analysis between foot and voice data was performed with a two-tailed Student's *t*-test for all variables. A one-tailed comparison was performed for the decrease in time for task completion between trials 1 and 5. Values for $p < 0.05$ were considered significant.

Results

The mean times for task completion at each trial are plotted in Fig. 3. The two upper curves correspond to voice control

Table 1. Summary of results of foot pedal versus voice interfaces in AESOP robot ex vivo trials

	Foot pedal	Voice
Time for trial completion on day 1 (sec)	43 ± 15[a] sec	97 ± 21[a] sec
Operator-interface failures per trial	1.4 ± 2.0[a]	4.8 ± 4.5[a]
''Pass points'' per trial	3.2 ± 2.4[a]	1.4 ± 1.3[a]
Durability (% improvement retained at 2 wk)	82.4%[b]	54.4%[b]

[a] $p < 0.002$.
[b] not significant.

on days 1 and 14; the lower curves correspond to foot control. On day 1, the overall mean time for task completion was 43 ± 15 sec for the foot interface and 97 ± 21 sec for the voice interface. The foot trials were significantly faster than the voice trials at all points timed ($p < 0.05$). On day 1, improvement in speed due to learning between trial 1 and trial 5 was about one-third of the initial trial time for both interfaces. The mean voice time for trial 1 was 114 sec; for trial 5, it was 81 sec. The mean foot time for trial 1 was 59 sec; for trial 5, it was 36 sec. The foot control learning curve plateaued at the third trial, while the voice control curve did not level off within five trials. Our criterion for reaching a plateau was a <5% difference in mean time between consecutive trials.

The durability of the learning experience was examined by retesting the surgeons on day 14. The difference between the times recorded at the initial and final trials on day 1 was defined as the ''time learned.'' The difference between the times recorded at the initial trial on day 1 and the initial trial on day 14 was defined as the ''time retained.'' ''Durability'' was calculated as the ratio between the time retained and the time learned, with the maximum set at 100% and the minimum set at 0%. The durability of the foot interface system was found to be 80% versus 51% for the voice interface system. This difference was not significant.

The accuracy of both interfaces was measured by the number of ''pass points.'' The mean number of such pass points for the foot controlled trials was 3.2 ± 2.4 versus 1.4 ± 1.3 for the voice controlled trials (Table 1). This difference was statistically significant ($p < 0.002$).

Operator-interface failures were significantly higher in the voice trials ($p < 0.05$). The mean number of times per trial that a surgeon's voice command was either misrecognized or not recognized was 4.8 ± 4.5 times, whereas the mean number of times per trial that the surgeon looked down to reorient the foot pedal was only 1.4 ± 2.0 times. A decreasing but not statistically significant trend in the number of operator-interface failures was noted with increasing trial number. From trial 1 to trial 5, foot reorientation decreased from 2.2 to 1.0 times, and voice recognition errors declined from 6.6 to 4.8 per trial.

Average times for the three surgeons with extensive foot pedal experience were 31, 35, and 51 sec (mean, 39 sec), while those of the less experienced surgeons were 39 and 45 sec (mean, 42 sec). Average pass points per trial were 2.2, 2.6, and 2.6 (mean, 2.5) for the attendings and 4.0 and 4.6 (mean, 4.3) for the residents. These groups are too small for meaningful statistical comparison. There does not appear to be a meaningful difference in speed between the groups. However, accuracy was better for the more experienced surgeons.

Discussion

New technology is constantly being developed to solve problems and enhance surgeon performance. Physicians today are pressured to be more efficient than ever. Replacing a human operator with a robotic arm eliminates the need to schedule and pay an extra person for holding the laparoscope [4]. In addition, since the surgeon regains control of the operative field, less confusion and miscommunication should result. However, the interface between the surgeon and robot dictates whether use of a robot will result in smoother and more efficient surgery. Careful assessment of technological advancements is crucial to determine their ultimate utility. Since the foot pedal has been clinically tested and accepted, it is the standard for comparison with the next generation of interface systems.

The foot pedal interface was found to be faster to learn and operate than the voice interface system. It also had fewer operator–interface failures. The learning process for both interfaces appears durable but did not reach significance with our small sample. The voice interface, on the other hand, was more accurate. This could be due to the fact that the voice interface can be used in two ways. In one mode, the robot continues a motion until the surgeon instructs it to stop, whereas in the second mode the robot only makes a specified small movement per command spoken. The second mode seems to be more precise but time consuming, since a command must be repeated until the desired location is reached. This could partially explain why the voice-controlled trials were slower but more accurate. Further experience with the voice interface and its continuous movement mode is expected to increase speed, but it may also decrease accuracy.

In order to present a complete picture of both interfaces, failures between the operator and the interface must be analyzed. The use of a third limb is required for a foot pedal system. Not only does it necessitate added coordination and concentration, it also requires the surgeon to look down occasionally to reorient his or her foot with the pedal. This diversion of the surgeon's attention from the operative field can interrupt the flow of the procedure and may not be practical in certain circumstances. As the trials progressed, a decreasing trend in the number of times the surgeons reoriented was noticed, but never did they all complete a trial without reorienting at least once.

Although the voice interface system seems to be a more natural and transparent interface, its utility is ultimately dependent on the technological state of voice recognition systems. We found that the current state of technology allowed approximately five recognition errors in a movement-intensive 90-sec trial. These errors are considered operator-interface failures because voice command repetition forces the surgeon to shift attention away from the operative field and to the voice command interface. The high number of recognition errors may also partially explain why the voice trials were slower. Misrecognized commands can potentially be a hazard during a surgical procedure.

This study has two primary limitations. First, it is small and did not allow several comparisons to reach statistical significance. It did, however, allow statistical comparison of our primary variables: speed, accuracy, and operator-interface failures.

The second limitation is that the study was performed ex vivo. Design of the study involved creation of a laparoscopic task that was a rational compromise between one that could be repeated multiple times allowing statistical comparison and one that accurately simulated in vivo laparoscopic surgical conditions. The types of motions required during this study are very similar to those performed in surgery. The delay between periods of laparoscope movement could not be easily re-created. During surgery, this interval often results in the need for foot reorientation prior to moving the laparoscope again. Thus, one would expect more operator-interface failures with the foot control during operative use than were seen under these experimental conditions. Voice recognition failures should be unaffected, since the microphone is always instantly available. This design necessity biased the operator-interface failure portion of the study in favor of the foot pedal interface but should have had a minimal effect on speed or accuracy. This bias must be considered when interpreting the results.

Conclusions

The ideal robotic interface would be completely transparent and move the endoscopic instantly to the desired position. The voice control interface represents another step in the evolution toward this ideal. It is currently more accurate than the foot pedal and has the advantage of not requiring the surgeon to look away from the operative field. However, in our model, it is slower and may interrupt the flow of the procedure more frequently with operator-interface failures. As voice recognition software continues to advance, speed and transparency should improve and voice control will likely evolve as the primary human-robotic interface.

Acknowledgments. Drs. Kavoussi and Moore are noncompensated consultants for Computer Motion.

References

1. Berci G, Sackier JM, Paz-Partlow M (1991) New ideas and improved instrumentation for laparoscopic cholecystectomy. Surg Endosc 5: 1–3
2. Jacobs LK, Shayani V, Sackier JM (1997) Determination of the learning curve of the AESOP robot. Surg Endosc 11: 54–55
3. Kavoussi LR, Moore RG, Adams JB, Partin AW (1995) Comparison of robotic versus human laparoscopic camera control. J Urol 154: 2134–2136
4. Partin AW, Adams JB, Moore RG, Kavoussi LR (1995) Complete robot-assisted laparoscopic urologic surgery: a preliminary report. J Am Coll Surg 181: 552–557
5. Sackier JM, Wang Y (1994) Robotically assisted laparoscopic surgery. From concept to development. Surg Endosc 8: 63–66.