Omurtag, A., Roy, R.N., Dehais, F., Chatty, L., and Garbey, M. (2018). Tracking Team Mental Workload by Multimodal Measurements in The Operating Room. In Neuroergonomics (Ayaz, H & Dehais F. Eds.)

Neuroergonomics

The Brain at Work and in Everyday Life

Editors: Hasan Ayaz Frederic Dehais
Paperback ISBN: 9780128119266
Imprint: Academic Press
Published Date: 1st November 2018
https://www.elsevier.com/books/neuroergonomics/ayaz/978-0-12-811926-6

CHAPTER 16

Tracking Team Mental Workload by Multimodal Measurements in The Operating Room

Ahmet Omurtag
ahmet.omurtag@ntu.ac.uk
Nottingham Trent University, United Kingdom

Raphaëlle N. Roy
raphaelle.roy@isae.fr
ISAE-SUPAERO, Université de Toulouse, Toulouse, France

Frédéric Dehais
frederic.dehais@isae.fr
ISAE-SUPAERO, Université de Toulouse, Toulouse, France

Luc Chatty
luc.chatty@phelma.grenoble-inp.fr
Houston Methodist Hospital and University of Houston
Houston, TX, United States

Marc Garbey
garbeymarc@gmail.com
Houston Methodist Hospital, TX, United States
and LaSIE UMR 7356 CNRS, University of La Rochelle, La Rochelle, France

Abstract

Mental workload and its effects on surgical performance are underexplored topics, despite their importance for operating room (OR) efficiency and patient safety. We developed a multimodal platform that can simultaneously collect data from EEG, heart rate and breathing rate, tool handle pressure, and eye tracker from mobile subjects. We performed experiments using the Fundamentals of Laparoscopic Surgery model, with 22 subjects of varying skill levels ranging from nonsurgeon to expert. The results indicated significant modulations of the measurements depending on pupil size, heart rate variability, P300 response, tool pressure, task difficulty, time-on-task, and skill level. These provide evidence that physiology based metrics can be used in automated classification of fine gradations of skill, the assessment and certification of surgery trainees, developing real-time flags and warnings for the OR, and validating new OR technology.

Keywords

Surgery training, workload, physiological measurements, neuroergonomics

Introduction

About 15 million operating room procedures are performed annually in the U.S. (Weiss & Elixhauser, 2006). A "hotspot" for medical errors, inpatient surgery is associated with 0.4-0.8% rate of death and 3-17% rate of major complications (Haynes et al., 2009). Studies suggest that about half of surgical complications are avoidable (Gawande, Thomas, Zinner, & Brennan, 1999; Kable, Gibberd, & Spigelman, 2002) and high-functioning teams have significant reductions in the number of adverse events (Mazzocco et al., 2009). New techniques that are being introduced potentially improve patient safety but impose dramatic new demands on surgeons' abilities and workload. More than one million laparoscopic surgeries are performed annually in the U.S. where a surgeon operates with an indirect, narrow visual access and minimal tactile feedback. Such conditions require new skills with different learning curves and new training methods beyond the traditional master-apprentice format (Van Hove, Tuijthof, Verdaasdonk, Stassen, & Dankelman, 2010). In fact as healthcare patterns shift toward prevention and quality, previously unexamined aspects of the operating room come into sharper focus and surgeons and trainees are scrutinized for their performance (Kao & Thomas, 2008; Kohn, Corrigan, & Donaldson, 2000; Pavlidis et al., 2012; Risucci, Geiss, Gellman, Pinard, & Rosser, 2001).

Surgeons use sophisticated instruments for extended periods often under time pressure, communicate with nurses and anesthesiologists, and interact with the complex interfaces of monitors. They possess technical skills acquired through long training. They also deploy an array of non-technical skills (Yule et al., 2008). These include situation awareness (gathering and understanding information and anticipating future states) and task management (responding to change). A strategic action may be, for example, deciding whether to convert a laparoscopic to an open-incision procedure. If the primary tasks (e.g. suturing) present unusual difficulty, this may impair the detection of an important alarm (Frédéric Dehais et al., 2014) or undermine proper planning. Even nearly automated mental processes, such as correcting for camera angle (Klein, Riley, Warm, & Matthews, 2005) or mismatches between the endoscope's optical axis and the instruments' plan on the monitor (Patil, Hanna, & Cuschieri, 2004) may be taking resources away from the surgeon's overall functions. Changes in mental workload due to training or new instrument design will have far reaching implications not only for efficiency but also for patient outcomes.

Behavioral and physiological measurements can help improve surgeons' workload monitoring. In developing measures of surgeon workload, hybrid or multi-modal approaches are preferable to unimodal ones, since they are able to deliver greater sets of information that illuminates the operator's functioning from multiple perspectives. Distinct measurement methods often have different strengths and shortcomings and may compensate for each other's artifacts. Furthermore, as hardware becomes increasingly miniaturized and sensor design improves, the cost and effort related to including additional modalities decreases (Gramann et al., 2011).

Yurko et al. (2010) utilized NASA-TLX to analyze the laparoscopic performance of novice trainees and to explain the extent of the transfer of their simulator-acquired skill to the operating room (OR). They found that the mental and physical demand ratings obtained at the beginning of training predicted part of the subsequent animal operating room performance scores (inadvertent injuries and suturing quality). Subjective methods such as NASA-TLX may be disruptive and only provide intermittent information. However the usefulness of the information highlights the need for unobtrusive, continuous means for tracking surgeon mental load.

Despite the apparent need, mental workload tracking in the OR using physiological measurements is under-explored. Although some studies have used physiological measurements to compare standard versus robotic assisted surgery (Hubert et al., 2013) or monitored surgeons using electroencephalography (Zander et al., 2016), we are not aware of any study that uses multimodal techniques in this area. We present a system of measurements for the operating room whose immediate purpose was to generate a large, multimodal dataset suitable for quantifying mental workload. The acquired datasets were recorded from electroencephalography (EEG), eye-tracking, electrocardiography, plethysmography and instruments with pressure sensors. We thus computed the pupil diameter (Peysakhovich, Causse, Scannella, & Dehais, 2015) and heart rate variability (Durantin, Gagnon, Tremblay, & Dehais, 2014) that are commonly used to assess mental effort in response to task demand. We also measured auditory evoked potentials (AEP), generated by low probability auditory stimuli, that are useful for investigating the effects of workload on perceptual processing (Roy et al., 2015) and have proven to be an efficient indirect means to derive mental workload in multitasking scenarios (Roy & Frey, 2016) . Expected benefits from this ongoing research effort include improved training programs and certification, more effective development of new technology, real-time safety alerts, and models capable of assisting OR management. Beyond this, the system is intended as a source of data that can be mined in order to quantify team dynamics and efficiency.

Methods

We performed experiments with 22 healthy volunteer subjects (4 females) on the Fundamentals of Laparoscopic Surgery (FLS) model of assessment (Vassiliou, Dunkin, Marks, & Fried, 2010). The subjects varied in level of experience (4 experts or board certified surgeons; 8 surgical residents from Post Graduate Years (PGY) 1-4; 10 nonsurgeons or beginners with no experience with FLS). They performed standard manual tasks peg transfer, string pass, and circle cut, listed in order of increasing difficulty (Peters et al., 2004). During performance their brain activity was recorded at 500 Hz using a 20 channel dry-sensor EEG at the International 10-20 electrode sites with ear lobes as the reference and ground (Quick-20, Cognionics Inc.). Pupil size and point-of-regard data were collected by an infrared-based eye tracking system at 60 Hz (The EyeTribe Tracker). Hexoskin wearable vest (Carré Technologies Inc., Montreal, Que., Canada) was used to monitor heart rate (HR) and breathing rate (BR).

In addition, an in-house designed pressure sensor was mounted at both left and right tool handles to monitor the force exerted by the subjects' thumbs while operating the instruments. We formulated several metrics likely to reveal differences of skill and training. Active time segments were defined as those when the pressure was above a fixed threshold. We defined Right-Left Overlap as those time segments when the right and left pressures were both Active. Finally we defined Right-Left Asymmetry as $\left|R-L\right|/\left(R+L\right)$ where R (and L) was the time average of the amplitude of the right (and left) tool pressure. We verified that the derived measures were not significantly affected by changes in the threshold within a wide range of values.

Figure **1** illustrates a participant using the experimental setup and the types of data that were collected. The figure includes an additional device for monitoring tool trajectories (Smart Trocar (Toti, Garbey, Sherman, Bass, & Dunkin, 2015)) whose data were not used in this paper.

Regarding the experimental protocol, sound probes were displayed during the surgical training task (100 ms duration, every 3 s on average, selected randomly from 6 different frequencies (750-2000 Hz)). The subjects were instructed to ignore the sound probes. All modalities were centrally controlled from a graphical user interface (GUI) capable of configuring the type of experiment to be performed, stimulus type (only sound probe was implemented), the data modalities to be included in the recording, as well as displaying video feedback from the FLS camera. Implemented in Matlab (The MathWorks, Inc., Natick, Massachusetts, United States), the GUI is part of a software platform that collects real-time, synchronized data from all modalities and stores them for off-line analysis. To study the effects of expertise and task difficulty, each subject performed Peg Transfer, String Pass, and Circle Cut while multimodal data were collected. In a subsequent experiment designed to measure the effects of time-on-task, groups with different skill levels (expert and surgical resident) performed peg transfer three times without a break. Each session was preceded by a one minute resting state recording.

To calculate the heart rate variability (HRV) from the ECG data, a widely used marker of autonomic activity (Task Force of the European Society of Cardiology, 1996), the time series of the normal-to-normal intervals in the electrocardiogram was resampled onto a regular time grid with cubic spline interpolation, and the spectral power in its low frequency (LF) (0.05-0.15 Hz) and high-frequency (HF) (0.15-0.5 Hz) bands were extracted. The HRV was defined as HF/(HF+LF).

As regards the EEG data, although (Zander et al., 2016) have shown the interest of studying frequency measures extracted from the EEG to perform mental workload monitoring during laparoscopic tasks, in this study we investigated whether event-related potentials (ERPs) extracted from ignored auditory probes (Roy et al., 2015; Roy & Frey, 2016) could also be used to monitor training during FLS. Indeed, tasks that require discrimination between classes of stimuli evoke in particular a large positive voltage deflection during about 300-500 ms after stimulus onset, known as the P300 component. The amplitude of the P300 is larger when the subject's attention is more focused on the task, and it is modulated by numerous factors including cognitive ability and mental workload (Gevins & Smith, 2000). We therefore chose to focus on P300 in this study and extracted its amplitude from electrode Cz.

Several statistical comparisons were performed on the extracted markers to determine the impact of expertise, training time and surgery sub-task. The significance of inter-group differences was determined by using the one-way Anova test.

Results

Figure 2 shows the results derived from measurements from subjects of varying skill levels during FLS tasks of varying difficulty. (A) The circle cut exercise resulted in subject averaged pupil size significantly different from that during String Pass, an easier task. The average pupil size in peg transfer, the easiest task, although similar to that of String Pass, did not significantly differ from the other tasks. (B) Heart rate variability varied significantly between operators of different skill levels (Expert and Nonsurgeon) performing the same set of FLS exercises. (C) The subject averaged P300 amplitude of response at Cz to sound probe decreased (regardless of skill level) as subjects spent more time on a repeated Peg Transfer task.

Figure 3A shows a segment of the tool pressure time series from two participants. The Expert time series was shifted up for clarity. As the pegs were picked up one by one by the left tool and passed to the right tool, the Expert's left and right pressures indicated a phase-locked wave pattern, absent from those of the Resident.

Figure 3B shows that pressure derived metrics (the fraction of Active time, Right-Left Overlap, and Right-Left Asymmetry) were significantly different between the Expert and Resident groups.

Discussion

Here we examined the feasibility of multimodal physiological measurements for tracking mental workload during surgery. We developed a platform that can simultaneously collect data from EEG, heart rate and breathing rate, tool handle pressure, and eye tracker from mobile subjects. The FLS assessment model was used as the experimental setting. The results indicated that the FLS task difficulty correlated significantly with pupil size and that HRV was related to operator skill level, indicating that untrained operators experienced a task to be harder (Durantin et al., 2014).

The P300 response to the ignored sound probes decreased significantly with training and time-on-task during a repeated FLS task. The effects of time-on-task and learning are always difficult to disentangle. Fatigue and workload effects often interact (Roy, Bonnet, Charbonnier, & Campagne, 2013) and monitoring systems should be designed to take these phenomena into account. In any case this amplitude decrement could be expected from the mental fatigue monitoring literature (Frederic Dehais et al., 2018). In addition our setup showed that the applied pressure on the tool handles contained patterns capable of robustly discriminating between Experts and Residents. Such metrics can be developed further to provide mechanisms for the automated classification of finer gradations of skill, the assessment and certification of surgery trainees, real-time flags and warnings for the OR, and validation of new OR technology.

An advantage of multimodal quantification of operators' activities is that such measurements can reveal varying degrees of effort that may go into similar levels of overt performance. Under some conditions, for example in testing a new instrument, a surgeon may make extra efforts to increase her primary task performance at the expense of additional mental load which may go undetected. If secondary tasks are introduced they may influence the primary task, or fail to provide accurate estimates because the subject did not reach capacity (Byrne, Tweed, & Halligan, 2014). Behavioral metrics may decouple from the mental load also when trainees attain a performance plateau, where the *only* effect of additional practice is to decrease the mental load (Wickens, Hollands, Banbury, & Parasuraman, 2015). If the trainee stops practicing at this stage, they may be left unprepared for the stressful situations that may arise later (Carswell, Clarke, & Seales, 2005; Johnston & Cannon-Bowers, 1996).

Conclusion

Despite mounting evidence, the field of surgery has not received sufficient attention from researchers developing physiology based methods to continuously track operators' mental workloads. Results presented here suggest that quantifying mental workload and other previously unexplored aspects of surgery through multimodal measurements can improve surgery training, and ultimately impact efficiency and safety in OR.

References

Byrne, A., Tweed, N., & Halligan, C. (2014). A pilot study of the mental workload of objective structured clinical examination examiners. *Medical Education*, *48*(3), 262–267.

Carswell, C. M., Clarke, D., & Seales, W. B. (2005). Assessing mental workload during laparoscopic surgery. *Surgical Innovation*, *12*(1), 80–90.

Dehais, Frédéric, Causse, M., Vachon, F., Régis, N., Menant, E., & Tremblay, S. (2014). Failure to detect critical auditory alerts in the cockpit evidence for inattentional deafness. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *56*(4), 631–644.

Dehais, Frederic, Di Flumeri, G., Verdiere, K., Borghini, G., Babiloni, F., & Roy, R. N. (2018). Monitoring pilot’s cognitive fatigue with engagement features in simulated and actual flight conditions using an hybrid fNIRS-EEG passive BCI. In *2018 IEEE Int’l Conf. on Systems Man and Cybernetics*.

Durantin, G., Gagnon, J.-F., Tremblay, S., & Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural Brain Research*, *259*, 16–23.

Gawande, A. A., Thomas, E. J., Zinner, M. J., & Brennan, T. A. (1999). The incidence and nature of surgical adverse events in Colorado and Utah in 1992. *Surgery*, *126*(1), 66–75.

Gevins, A., & Smith, M. E. (2000). Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cerebral Cortex*, *10*(9), 829–839.

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T.-P., Lin, C.-T., … Makeig, S. (2011). Cognition in action: imaging brain/body dynamics in mobile humans. *Reviews in the Neurosciences*, *22*(6), 593–608.

Haynes, A. B., Weiser, T. G., Berry, W. R., Lipsitz, S. R., Breizat, A.-H. S., Dellinger, E. P., … Gawande, A. A. (2009). A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *New England Journal of Medicine*, *360*(5), 491–499. https://doi.org/10.1056/NEJMsa0810119

Johnston, J. H., & Cannon-Bowers, J. A. (1996). Training for stress exposure. *Stress and Human Performance*, 223–256.

Kable, A. K., Gibberd, R. W., & Spigelman, A. D. (2002). Adverse events in surgical patients in Australia. *International Journal for Quality in Health Care*, *14*(4), 269–276.

Kao, L. S., & Thomas, E. J. (2008). Navigating towards improved surgical safety using aviation-based strategies. *Journal of Surgical Research*, *145*(2), 327–335.

Klein, M. I., Riley, M. A., Warm, J. S., & Matthews, G. (2005). Perceived mental workload in an endocopic surgery simulator. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, pp. 1014–1018). SAGE Publications.

Kohn, L. ., Corrigan, J. M., & Donaldson, M. S. (2000). *To Err is Human: Building a Safer Health System*. Washington DC: Committee on Quality of Health Care in America, Institute of Medicine.

Mazzocco, K., Petitti, D. B., Fong, K. T., Bonacum, D., Brookey, J., Graham, S., … Thomas, E. J. (2009). Surgical team behaviors and patient outcomes. *The American Journal of Surgery*, *197*(5), 678–685.

Patil, P. V., Hanna, G. B., & Cuschieri, A. (2004). Effect of the angle between the optical axis of the endoscope and the instruments' plane on monitor image and surgical performance. *Surgical Endoscopy And Other Interventional Techniques*, *18*(1), 111–114.

Pavlidis, I., Tsiamyrtzis, P., Shastri, D., Wesley, A., Zhou, Y., Lindner, P., … Dunkin, B. (2012). Fast by nature-how stress patterns define human experience and performance in dexterous tasks. *Scientific Reports*, *2*, 305.

Peters, J. H., Fried, G. M., Swanstrom, L. L., Soper, N. J., Sillin, L. F., Schirmer, B., … others. (2004). Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery*, *135*(1), 21–27.

Peysakhovich, V., Causse, M., Scannella, S., & Dehais, F. (2015). Frequency analysis of a task-evoked pupillary response: Luminance-independent measure of mental effort. *International Journal of Psychophysiology*, *97*(1), 30–37.

Risucci, D., Geiss, A., Gellman, L., Pinard, B., & Rosser, J. (2001). Surgeon-specific factors in the acquisition of laparoscopic surgical skills. *The American Journal of Surgery*, *181*(4), 289–293.

Roy, R. N., Bonnet, S., Charbonnier, S., & Campagne, A. (2013). Mental fatigue and working memory load estimation: interaction and implications for EEG-based passive BCI. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 6607–6610). IEEE. Retrieved from http://ieeexplore.ieee.org.ezproxy.lib.uh.edu/xpls/abs_all.jsp?arnumber=6611070

Roy, R. N., Breust, A., Bonnet, S., Porcherot, J., Charbonnier, S., Godin, C., & Campagne, A. (2015). Influence of workload on auditory evoked potentials in a single-stimulus paradigm. In *2nd international confrenrence on physiological computing, PhyCS 2015*. Retrieved from https://www.researchgate.net/profile/Aurelie_Campagne/publication/271470655_Influ ence_of_Workload_on_Auditory_Evoked_Potentials_in_a_Single-Stimulus_Paradigm/links/54c8b3dc0cf289f0ced0b4e9.pdf

Roy, R. N., & Frey, J. (2016). Neurophysiological Markers for Passive Brain–Computer Interfaces. In L. Bougrain & M. Clerc (Eds.), *Brain–Computer Interfaces: Methods, Applications, and Perspectives* (Vol. 1, pp. 85–100). UK: ISTE-Wiley. Retrieved from http://onlinelibrary.wiley.com.ezproxy.lib.uh.edu/doi/10.1002/9781119144977.ch5/summary

Task Force of the European Society of Cardiology. (1996). Heart rate variability standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*, *17*, 354–381.

Toti, G., Garbey, M., Sherman, V., Bass, B. L., & Dunkin, B. J. (2015). A Smart Trocar for Automatic Tool Recognition in Laparoscopic Surgery. *Surgical Innovation*, *22*(1), 77–82.

Van Hove, P. D., Tuijthof, G. J. M., Verdaasdonk, E. G. G., Stassen, L. P. S., & Dankelman, J. (2010). Objective assessment of technical surgical skills. *British Journal of Surgery*, *97*(7), 972–987.

Vassiliou, M. C., Dunkin, B. J., Marks, J. M., & Fried, G. M. (2010). FLS and FES: comprehensive models of training and assessment. *The Surgical Clinics of North America*, *90*(3), 535–558. https://doi.org/10.1016/j.suc.2010.02.012

Weiss, A. J., & Elixhauser, A. (2006). Trends in Operating Room Procedures in U.S. Hospitals, 2001–2011: Statistical Brief #171. In *Healthcare Cost and Utilization Project (HCUP) Statistical Briefs*. Rockville (MD): Agency for Health Care Policy and Research (US). Retrieved from http://www.ncbi.nlm.nih.gov/books/NBK201926/

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology & human performance*. Psychology Press. Retrieved from https://books-google-com.ezproxy.lib.uh.edu/books?hl=en&lr=&id=_rFmCgAAQBAJ&oi=fnd&pg=PP1&dq=wickens+engineering+psychology+and+human+performance&ots=wWuketjFKL&sig=k5NGnbaMBeWReuBhQP6qIdSmd3c

Yule, S., Flin, R., Maran, N., Rowley, D., Youngson, G., & Paterson-Brown, S. (2008). Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behavior rating system. *World Journal of Surgery*, *32*(4), 548–556.

Yurko, Y. Y., Scerbo, M. W., Prabhu, A. S., Acker, C. E., & Stefanidis, D. (2010). Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. *Simulation in Healthcare*, *5*(5), 267–271.

Zander, T. O., Shetty, K., Lorenz, R., Leff, D. R., Krol, L. R., Darzi, A. W., … Yang, G.-Z. (2016). Automated Task Load Detection with Electroencephalography: Towards Passive Brain–Computer Interfacing in Robotic Surgery. *Journal of Medical Robotics Research*, 1–10.

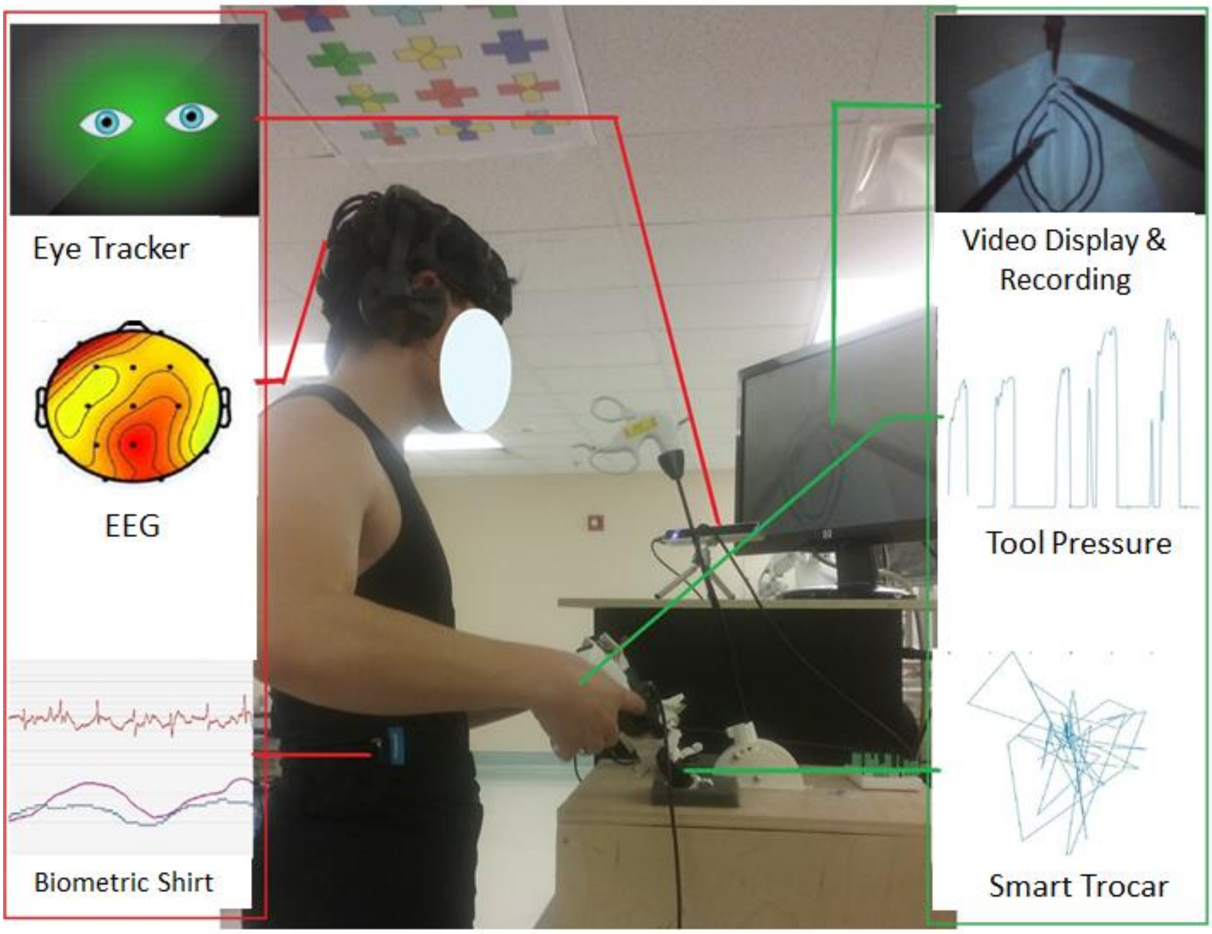

Figure 1 Types of data being collected while a participant uses the experimental setup.

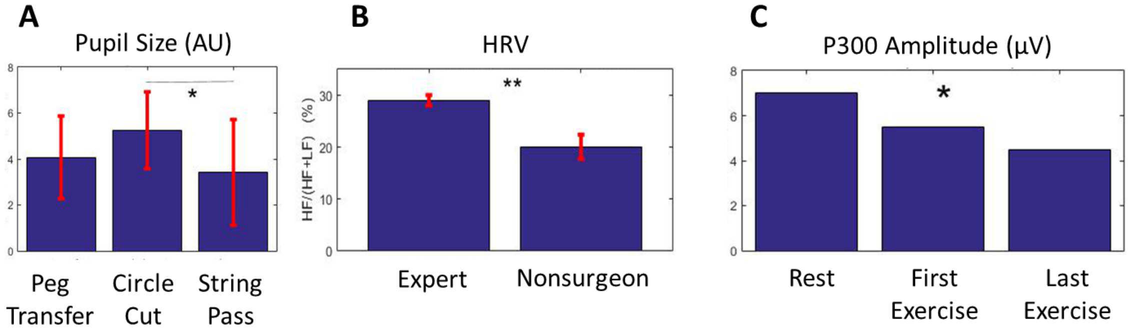

Figure 2 Results from (A-B) three types of standard FLS tasks and (C) three repetitions of the peg transfer task. (A) Pupil size as a function of type of task. (B) HRV as a function of skill level. (C) Amplitude of the P300 response to sound probe as a function of time-on-task. (*p<0.05; **p<0.01.)

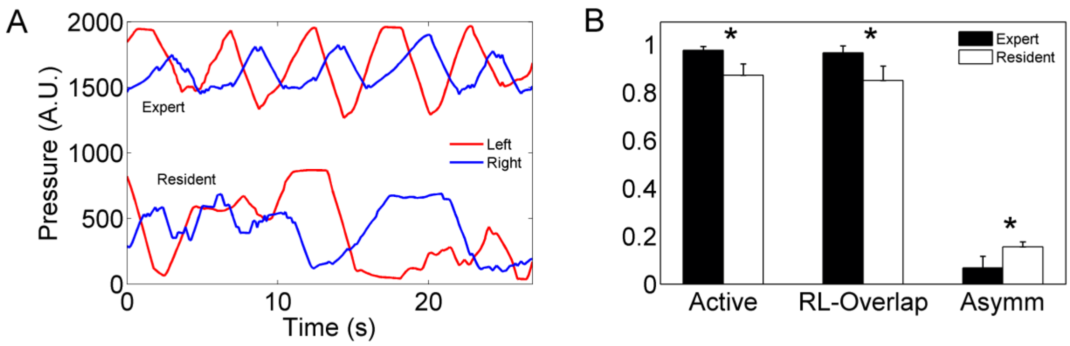

Figure 3 Results derived from tool pressure measured continuously as subjects (Expert and Residents) performed the peg transfer task. Pressure time series often differed visibly between Experts and Residents (A). The fraction of time during which the pressure was above a fixed threshold (Active), the fraction of time during which Right-Left pressures were both Active (RL-Overlap), and the Right-Left pressure asymmetry (Asymm) were significantly different between two groups (B). (*p<0.05.)