Surg Endosc
DOI 10.1007/s00464-016-4851-8

Gaze entropy reflects surgical task load

Leandro L. Di Stasi[1,2,3] · Carolina Diaz-Piedra[1,2,3] · Héctor Rieiro[1,2,4] · José M. Sánchez Carrión[5] · Mercedes Martin Berrido[5] · Gonzalo Olivares[5] · Andrés Catena[2]

Received: 13 July 2015 / Accepted: 2 March 2016

Abstract

Background Task (over-)load imposed on surgeons is a main contributing factor to surgical errors. Recent research has shown that gaze metrics represent a valid and objective index to asses operator task load in non-surgical scenarios. Thus, gaze metrics have the potential to improve workplace safety by providing accurate measurements of task load variations. However, the direct relationship between gaze metrics and surgical task load has not been investigated yet. We studied the effects of surgical task complexity on the gaze metrics of surgical trainees.

Methods We recorded the eye movements of 18 surgical residents, using a mobile eye tracker system, during the performance of three high-fidelity virtual simulations of laparoscopic exercises of increasing complexity level: *Clip Applying* exercise, *Cutting Big* exercise, and *Translocation of Objects* exercise. We also measured performance accuracy and subjective rating of complexity.

Results Gaze entropy and velocity linearly increased with increased task complexity: Visual exploration pattern became less stereotyped (i.e., more random) and faster during the more complex exercises. Residents performed better the *Clip Applying* exercise and the *Cutting Big* exercise than the *Translocation of Objects* exercise and their perceived task complexity differed accordingly.

Conclusions Our data show that gaze metrics are a valid and reliable surgical task load index. These findings have potential impacts to improve patient safety by providing accurate measurements of surgeon task (over-)load and might provide future indices to assess residents' learning curves, independently of expensive virtual simulators or time-consuming expert evaluation.

Keywords Eye metrics · Neuroergonomics · Surgical skills assessment · Patient safety · Saccades

Leandro L. Di Stasi and Carolina Diaz-Piedra have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00464-016-4851-8) contains supplementary material, which is available to authorized users.

✉ Leandro L. Di Stasi
distasi@ugr.es

✉ Carolina Diaz-Piedra
dipie@ugr.es

1 Neuroergonomics and Operator Performance Laboratory, University of Granada, Campus de Cartuja s/n, 18071 Granada, Spain

2 Mind, Brain, and Behavior Research Center (CIMCYC), University of Granada, Campus de Cartuja s/n, 18071 Granada, Spain

3 College of Nursing and Health Innovation, Arizona State University, Phoenix, AZ, USA

4 Department of Signal Theory and Communications, University of Vigo, Vigo, Spain

5 Line of Activity of the Andalusian Public Foundation for Progress and Health, Ministry of Equality, Health and Social Policy of the Regional Government of Andalusia, IAVANTE, Granada, Spain

Although the healthcare system heavily invests in the creation of systematic and structured medical training assessment protocols in order to improve patient's safety [1], medical errors continue to occur [2]. Task (over-)load imposed on surgeons is a main contributing factor leading to errors [3, 4]. For example, while new surgical techniques present several advantages to the patient's safety and quality of life—such as reduced postoperative pain and earlier return to activities of daily living [5]—surgeons have started to face unexpected higher levels of work load [6, 7]. Therefore, new assessment metrics are needed to inform about the costs related to these task load variations [8]. Quantifying surgeons' task load provides an opportunity to understand surgeons' response to the challenges posed by surgical procedures, and consequently, it represents a first step to ensure patient's safety and to improve surgical training programs [9].

Effective surgical training represents a valid tool to countermeasure (over-)load situations in medical practice [10]. However, surgical training and its assessment is particularly challenging due to the fundamental ethical tension between medical education and patients' safety [11]. Current assessment tools are based on validated observer-generated scoring systems [12], on self-reported tests and questionnaires [13], and on computer-based simulations metrics [14]. Although sometimes sophisticated, such methods have methodological and setting restrictions, however. The main limitations for subjective-based methods rise in their off-line nature or the response biases related to personal and motivational factors [15]. Computer-based methods have very limited applicability outside expensive simulation laboratories [16, 17]. Thus, the development of objective and feasible systems to assess surgical task load remains a major challenge in the healthcare field [18, 19].

Vision is the dominant sensory system supporting human function, especially in psychomotor tasks, as it is vital to identify the action's target and provides feedback to enable corrections [20]. In fact, gaze metrics represent one of the most valid, reliable, and objective indices to assess operator performance [21–23], including physicians [24, 25]. Specifically, in surgical scenarios, gaze metrics have been used to quantify surgical skills [25–27], to study differences between novice and expert surgeons [21, 28, 29], and to analyze surgeon' scanning behavior [30, 31]. Previous studies have also tried to relate surgical task load to gaze metrics using indirect relationships, as those mediated by surgeon's expertise [21] or time-on-task [19]. To date, however, no studies have considered the direct relationship between gaze-based metrics and surgical task load, by manipulating the objective complexity of the task (see Ref. [18] for the recent literature review on eye-based metrics in surgery).

In non-surgical scenarios, gaze entropy—a measure of the uncertainty over the gaze position at any point in time [32] (see "Gaze dynamics analyses: Gaze entropy [and velocity]" section for more details)—has been proved to be a sensitive and valid task load index [33–35]. Briefly, gaze entropy increases with task load [34]: that is, visual exploration patterns become less stereotyped (i.e., more random) during more complex tasks. Furthermore, in medical applications, gaze entropy has been successfully used to recognize the rising and falling interest levels of expert radiologists during medical imaging assessment and, then, track their learning curves [25]. When compared with other eye metrics, such as pupil response [21, 27, 36], gaze entropy offers the advantage that the measured value does not depend on external factors such as environmental light or emotional state [37, 38]. Furthermore, unlike other sensitive indices, as saccadic (fastest movement of the eye) velocity [19], gaze entropy does not rely on intrusive and bulky eye tracking systems for the acquisition of its signals [34, 39]. Thus, gaze entropy could be measured with mobile and wearable eye trackers (e.g., small-framed glasses eye trackers) which are robust to head movements and changes in ambient illumination, and, most importantly, are easy and comfortable to use. These features make possible to assess surgeon's performance in real medical settings [21, 40]. For these reasons, a task load index based on gaze entropy has the potential to provide an accurate and unbiased measure of the task demands imposed on the surgeon and therefore be a valid index to use in real and simulated surgical scenarios.

Here, we aimed to investigate the effect of task load on the gaze entropy of surgical residents while performing high-fidelity simulated laparoscopic exercises. Residents performed three standardized exercises with different levels of task complexity (low, medium, and high). We expected that task complexity would increase gaze entropy, as well as the perceived task complexity, while, at the same time, would reduce laparoscopic performance.

Materials and methods

Ethical approval

We conducted the study in conformity with the Code of Ethics of the World Medical Association [41]. The experiments were carried out under the guidelines of the University of Granada's Institutional Review Board (IRB approval #899). Written informed consent was obtained from each resident prior to the study.

Experimental design

Each experimental session included three simulated laparoscopic exercises corresponding to three levels of task complexity (low, medium, and high) [42]. We used a Latin square design to minimize the potential effect of confounding factors, including practice and learning effects and task-switching costs (i.e., the costs associated with going from a complex task to an easy one). The dependent variables for the analysis were the perceived level of task complexity, the accuracy in performing the tasks, and the gaze-based metrics (gaze entropy and velocity, see "Gaze dynamics analyses: Gaze entropy [and velocity]" section). Sleepiness level before the experimental session and fatigue between tasks were also assessed. Each experimental session lasted ~45 min.

Sample size

A sample of 18 residents was set to obtain a full balanced design (i.e., equal numbers of residents in each of six possible simulated laparoscopic exercise sequences). Owing to lack of applicable pilot data—being this study first of its nature—no power calculation was undertaken. Such a number was considered appropriate based on a previous cohort, where statistically significant differences in the gaze metrics were found [21].

Participants

Residents attended IAVANTE (Andalusian Public Foundation for Progress and Health), in Granada (Spain), for laparoscopic training. Eighteen residents (7 males and 11 females) volunteered to participate in the study (mean age, height, and weight: 27 ± 2.6 years, 168 ± 8.1 cm, and 67 ± 14.0 kg). Participants had received postgraduate training for an average of ~2 years (min 1 year, max 4 years). To reduce possible bias due to previous experience (i.e., knowledge-based shortcuts), we involved only laparoscopically naive residents [17]. Most of the residents were junior surgical trainees (only two of them were middle surgical trainees). Residency specialities included: pediatrics (2), urology (4), general surgery (5), and obstetrics–gynecology (7). All participants had normal or corrected-to-normal vision. Residents were non-smokers and mostly right-hand dominant (only one resident was left-hand dominant). They reported an average of 7 h of sleep (range 6–8 h) during the night previous to the study. Before the experiment, each resident filled in the Stanford Sleepiness Scale [43] (SSS) for screening purposes. No participants scored >3 (had they done so, they would have been excluded from further testing [44, 45]). All participants were naive to the aim of the experiment.

Apparatus and tasks

Residents performed three simulated laparoscopic exercises from the Basic Laparoscopic Skills Acquisition Module (Lap Mentor™, Simbionix®): *the Clip Applying* exercise, the *Cutting Big* exercise, and the *Translocation of Objects* exercise, using the Lap Mentor™ high-fidelity virtual reality laparoscopic surgery simulator (Fig. 1). In *the Clip Applying* exercise, residents used two clip appliers to clip nine leaking ducts in a specific segment. In the *Cutting Big* exercise, residents used one duck bill to retract and one pair of straight endoscopic scissors to safely cut and separate a gel circular form attached to a metal ring. In the *Translocation of Objects* exercise, residents used two duck bills to manipulate six solid objects and place them into the orientation of matching transparent ones. The lookzone area, defined as the region of interest around the relevant task targets (i.e., leaking ducts, gel circular form, and solid/transparent objects) and respective tools [46], did not differ across the three simulated laparoscopic exercises (*Clip Applying* = ~343 cm^2; *Cutting Big* = ~353 cm^2; and *Translocation of Objects*: ~344 cm^2).

Procedure

The advanced Multi-purpose Simulation and Technological Innovation Complex in IAVANTE houses several laparoscopic and advanced human robotic simulators, which provide health professionals with general clinical skills and specific surgical training. The day before the experimental session, all residents underwent a 5-min simulator familiarization phase. Previous to the beginning of the experimental session, we recorded the date of birth, sex, hand dominance, years of postgraduate training, and residency speciality for each resident. Sleepiness was assessed in that moment. Then, we set up the eye tracking system and performed the calibration procedure. An instruction screen indicating the type of task to be performed preceded each simulated laparoscopic exercise. No time limits were imposed on the residents to perform the three exercises. Residents did not rest between exercises, except to answer the questionnaires described below.

Questionnaires

We used the NASA-Task Load Index [NASA-TLX] questionnaire [47] as an indicator of the degree of task complexity that residents experienced while performing the simulated laparoscopic exercises. The NASA-TLX is a scale with six bipolar dimensions: mental demand (MD); physical demand (PD); temporal demand (TD); own performance; effort; and frustration. The first three dimensions (MD, PD, and TD) reflect task-related factors such as task complexity [19]. Furthermore, because surgeon fatigue might be a confounding factor in affecting performance [48, 49], we asked residents to fill in the Borg rating of perceived exertion [50] [Borg Scale] after each simulated laparoscopic exercise (i.e., three different measuring times). This scale indicates the levels of perceived fatigue. The score ranges from 6 to 20, where higher scores reflect higher levels of fatigue.

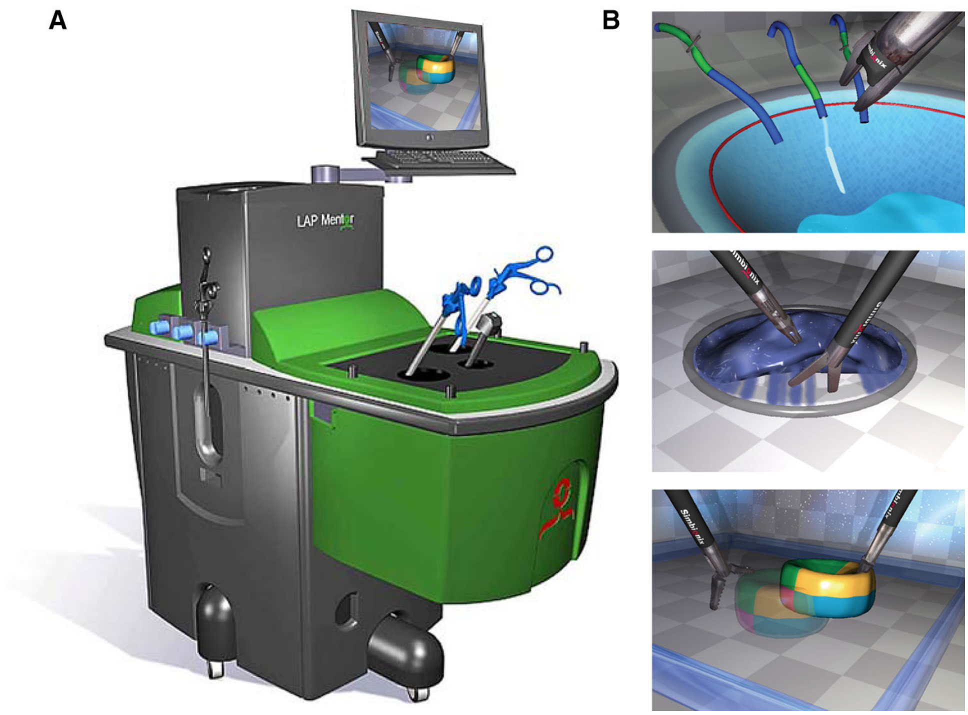

Fig. 1 Experimental setup. **A** Residents used the Lap Mentor™ surgery simulator (Simbionix®) for the laparoscopic tasks. The position of the simulator tower was fixed. The laparoscopic tasks were displayed on a 17″ computer monitor facing the subject. The monitor was mounted on the simulator tower and positioned about 160 cm above the floor level and about 60 cm from the participants' eyes. **B** Screenshots showing the three laparoscopic exercises (extracted from the Basic Task Module of the LAP Mentor™): *Clip Applying* exercises (*top*), *Cutting Big* exercises (*middle*), and *Translocation of Objects* exercises (*bottom*). Figure adapted from (http://simbionix.com/)

Performance data

We used the subjects' accuracy performing the exercises as an indicator of task complexity [51]. For the *Clip Applying* exercise, we considered the number of clipped ducts divided by the total number of clipping attempts. For the *Cutting Big* exercise, we considered the number of retraction operations without overstretching injuries to the tissue divided by the total number of retraction operations. Finally, for the *Transposition of Objects* exercise, we considered the average ideal number of translocations per object divided by the average number of translocations per object. To compare residents' performance across the three simulated exercises, accuracy was always expressed in percentage.

Eye movement recordings and analyses

We sampled eye movements binocularly at 30 Hz, using a wearable eye tracking system (Tobii Glasses 2.0, Tobii Sweden). The device consists of a small recoding unit—attached to a belt worn by the residents—and an eye tracking unit mounted on an eyeglasses frame. The eye tracker is connected to the recoding unit via a HDMI cable. Recordings are stored in a SD memory card. See the recent works by Morales and colleagues [39] for a detailed description of the system.

We recorded eye movements only during the execution of the three simulated laparoscopic exercises: i.e., collected data reflect surgeon's gaze behavior during his/her performance. We identified blink periods as portions of the raw data where eye information was missing for 100 ms or more and removed these segments from the analysis. We further removed the 200 ms before and after each blink or semi-blink to eliminate the initial and final parts during which the pupil was still partially occluded. Gaze data from three residents were lost due to a failure of the recording system.

Gaze dynamics analyses: Gaze entropy (and velocity)

While operating, surgeon's eyes are constantly making voluntary rapid movements (i.e., saccades) to sample ambient regions and extract relevant information (i.e., fixation periods). During each fixation, in addition to extracting information, the surgeon has to make a decision as to where to look next [52]. The sequential periods of rapid saccades and steady fixations define surgeon's gaze behavior. Here we focused on two features of surgeon's gaze behavior: gaze entropy and velocity.

To measure (gaze) entropy, we used Shannon's entropy formula [53], defined as:

$$H_g(X) = -\sum p(x,y) \cdot \log_2 p(x,y)$$

where $p(x,y)$ is the probability of the subject's gaze falling in the (x,y) position of the visual field for a given sample (see Fig. 2), estimated from the full recording. This gives a

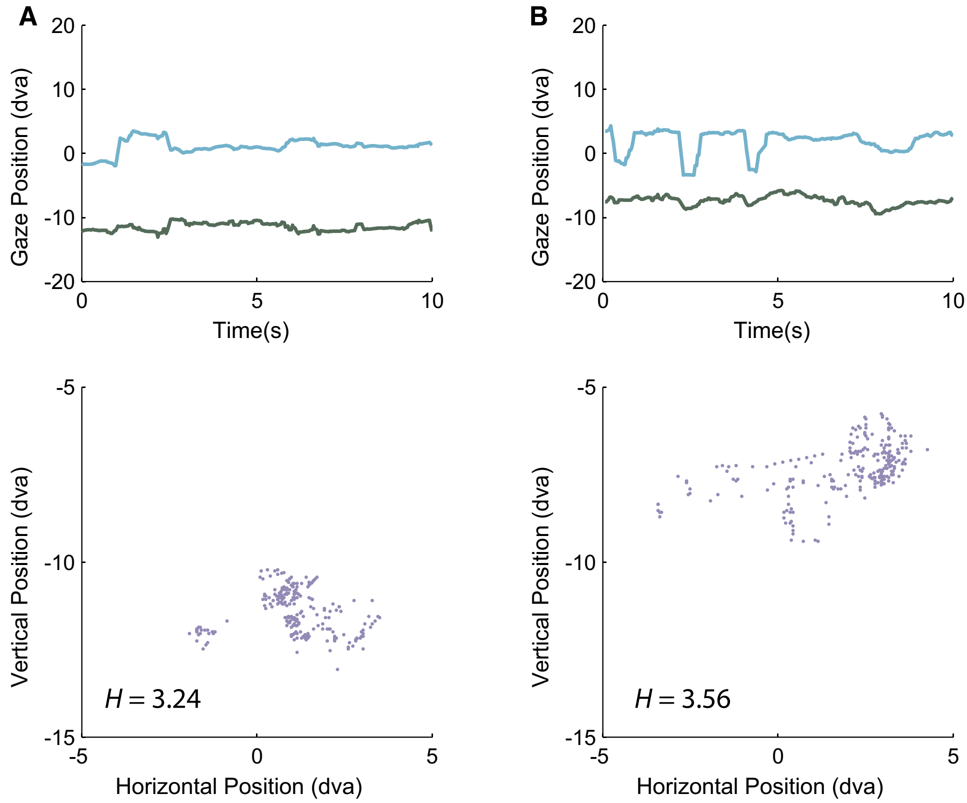

Fig. 2 Examples of gaze traces and entropy values obtained in the experiment. The *top row* shows *vertical* (*teal color*) and horizontal (*pine color*) gaze position traces across time. A rapid change in the trace on the Y-axis indicates a saccadic movement, while a still trace implies a fixation period. The bottom row shows the gaze position of each sample in the horizontal and vertical axes. The exact estimated entropies for each sequence are shown and labeled as H in the corresponding plots. The left column **A** shows data for a low-entropy period, while the right column **B** shows data for a higher entropy ($\sim$9 % increment) period. Each period has duration of 10 s (300 samples with our recording equipment) (Color figure online)

measure of the average uncertainty over the position of gaze on an instant in time during the simulated exercise, or, equivalently, the information provided by a single observation, measured in bits. Therefore, it is a measure of gaze dispersion. In order to calculate the gaze entropy, we divided the visual field [81 $\times$ 48 degrees of visual angle (dva); i.e., the total visual field allowed by the eye tracker device] in 3888 bins of 1 $\times$ 1 dva, and calculated the probabilities of the gaze falling in each of these bins on any given time sample. Note that 2 $\times$ 2 dva and 4 $\times$ 4 dva sized bins produced equivalent results. Gaze entropy, as defined here, provides a good indication of the dispersion of gaze over the visual field. However, it is still possible that this dispersion obeys to the spatial properties of the task at hand and not necessarily its complexity. To exclude the possibility that we are conflating task complexity with its spatial structure, we use an independent, complimentary measure: the gaze velocity, which has been previously used as a proxy to study gaze behavior [54]. Thus, we defined gaze velocity averaged over the simulated exercise, using the harmonic mean, as:

$$\bar{v}_{\mathrm{g}} = \frac{N \cdot f_{\mathrm{s}}}{\sum_{n=1}^{N-1} \frac{1}{r_{\mathrm{g}}(n) - r_{\mathrm{g}}(n-1)}}$$

where f_s is the sampling rate of the recording device (in this case, 30 Hz), N is the number of samples recorded during the simulated exercise, and $r_g(n)$ is the *(x,y)* position of the gaze (in units of degrees of visual angle) on the sample *n*. Gaze velocity provides information about how fast the gaze is continuously moving through the exercises, independently of the spatial configuration of its elements. Therefore, a high average velocity will indicate that the gaze is moving faster, which explains at least a fraction of the entropy, independently of the spatial properties of the task.

To make our conclusions more likely to be generalized to real medical environments (i.e., robust), we computed gaze entropy and velocity values, considering all the recorded raw gaze points instead of fixations coordinates or saccade epochs, respectively. We took this decision also considering the technical constraints from the eye tracker, namely high levels of noise system and low sampling rate. Notwithstanding the above, we obtained similar results (see *Supplementary material* section) when applied classical gaze entropy calculations (e.g., [25]).

Statistical analyses

To analyze the effects of task complexity, we calculated separate repeated-measures ANOVA (one for each dependent variable: gaze entropy and gaze velocity, laparoscopic accuracy, and NASA-TLX scores) with the three simulated exercises (*Clip Applying* exercise, *Cutting*

Big exercise*, and Translocation of Objects* exercise) as the within-subject factor. Thus, we compared each resident to him/herself across exercises, and variability between residents was part of the error terms. We also performed separate trend analysis (one for each dependent variable) to identify the existence of specific trends in our data. Furthermore, to estimate possible effects of perceived level of fatigue on our main dependent variables, we analyzed the Borg Scale scores across the exercises, independently from the complexity of the performed task. In all analyses, we used the Bonferroni-Holm correction for multiple comparisons [55]. Significance levels were always set at $\alpha < .05$.

Results

We aimed to determine the effect of task complexity on surgeons' gaze entropy, surgical performance, and perceived task complexity. Residents performed three standardized simulated laparoscopic exercises while we recorded their eye movements.

Effectiveness of surgical task complexity manipulation: subjective ratings and accuracy

Task complexity (as expressed by the NASA-TLX scores) linearly increased from the easy to more complex exercises, $F(1, 17) = 95.38$, $p < .05$. Accordingly, task complexity differed across the exercises, $F(2, 34) = 37.15$, $p < .05$, residents experienced lower levels of complexity for the *Clip Applying* exercise than for the *Cutting Big* exercise, and perceived lower levels of complexity for the *Cutting Big* exercise than for the *Translocation of Objects* exercise (corrected p values <.05, see Table 1).

We found an opposite linear trend for the laparoscopic accuracy, $F(1, 17) = 17.51$, $p < .05$. Accuracy differed across the exercises, $F(2, 34) = 6.22$, $p < .05$, residents performed better the *Clip Applying* exercise and the *Cutting Big* exercise than the *Translocation of Objects* exercise (corrected p values <.05, see Table 1).

Finally, surgeon fatigue (as expressed by the Borg Scale) was stable across the experimental session, $F(2, 34) = 2.54$, $p > .05$; mean ± SD: 1st measuring time = 9 ± 2.3; 2nd measuring time = 10 ± 2.6; and 3rd measuring time = 10 ± 2.8.

All together, these results indicate that the task complexity manipulation was successful.

Effect of surgical task complexity on residents' gaze dynamics indices

Gaze entropy linearly increased with increasing task complexity, $F(1, 14) = 117.97$, $p < .05$, indicating that exploration pattern become less stereotyped (i.e., more random) during the more complex exercises. Gaze entropy was lower in the *Clip Applying* exercise and the *Cutting Big* exercise than the *Translocation of Objects* exercise, $F(2, 28) = 28.46$, $p < .05$; corrected p values <.05 (see Fig. 3; Table 1). In the same way, gaze velocity linearly increased with increased task complexity, $F(1, 14) = 20.39$, $p < .05$, indicating that exploration pattern becomes more rapid during the more complex exercises. Gaze velocity was slower in the *Clip Applying* exercise and the *Cutting Big* exercise than in the *Translocation of Objects* exercise, $F(2, 28) = 6.96$, $p < .05$; corrected p values <.05 (see Fig. 3; Table 1).

Discussion

We studied the effects of surgical task complexity on the gaze metrics of surgical trainees. Our results show that task complexity modulates gaze dynamics during standardized

Table 1 Effect of task complexity on gaze dynamics (entropy and velocity), laparoscopic accuracy, execution time, and subjective ratings

	Clip Applying	Cutting Big	Translocation of Objects
Gaze entropy* (bits), M ± SD	6.798 ± 0.7	6.939 ± 0.9	7.654 ± 0.7
Gaze velocity* (dva/s), M ± SD	6.754 ± 0.1	7.472 ± 0.1	8.797 ± 0.1
Accuracy*, † (%), M ± SD	58.233 ± 0.1	58.081 ± 0.3	38.110 ± 0.2
NASA-TLX*, M ± SD	22.863 ± 15.1	37.569 ± 17.4	54.605 ± 18.7

The mean and standard deviation of the gaze entropy and velocity were calculated from fifteen residents ($n = 15$). NASA-TLX values range between 0 and 100, with higher values indicating higher task load. Accuracy values are reported in percentage, with higher values indicating better performance. The mean and standard deviation of the NASA-TLX scales, accuracy score, and execution time were calculated from all residents ($n = 18$)

dva/s Degree of visual angle/second, *M* Mean, *NASA-TLX* NASA-Task Load Index, *SD* Standard deviation

*Bonferroni-Holm corrected p values < .05

† Execution time (sec), M ± SD: Clip Applying: 97 ± 22; Cutting Big: 192 ± 22; and Translocation of Objects: 577 ± 275

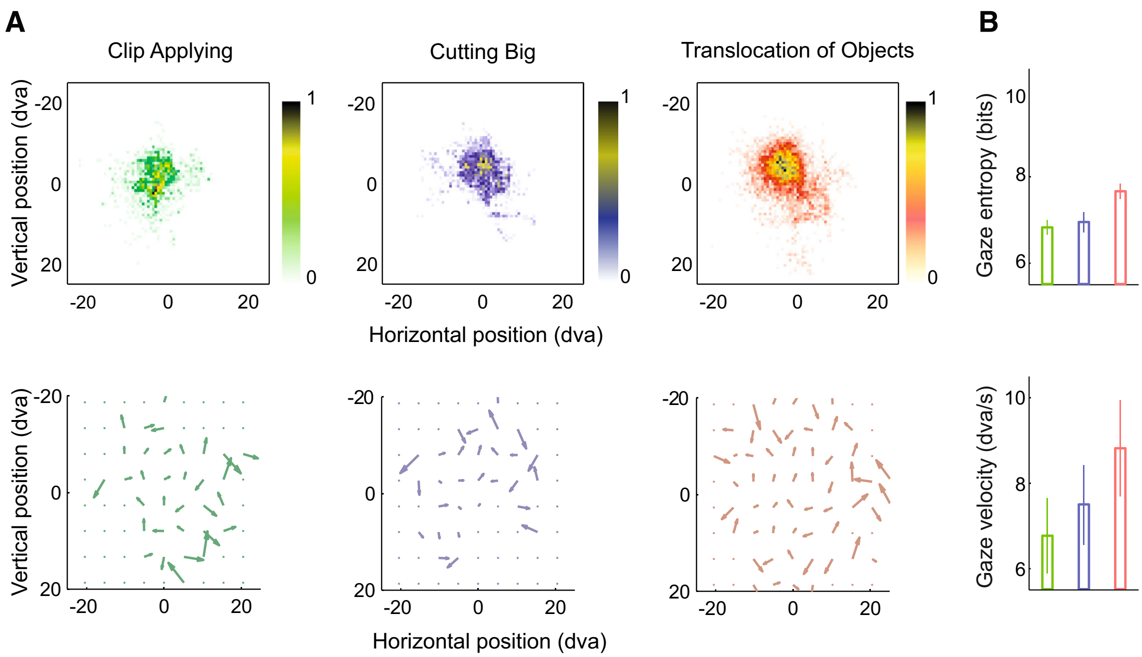

Fig. 3 Analysis of gaze entropy and velocity across the three laparoscopic exercises. **A** Example of heat maps (*top*) and velocity (*bottom*) maps for resident *#2* across the different exercises. Heat maps are normalized to the maximum value of each map. *Yellowish/darker* regions of the visual field indicate areas where residents directed their gaze more often. Velocity maps show the harmonic average of the gaze velocity. Gaze velocity moduli are normalized over the three exercises and represented by the arrow length using a logarithmic scale. The *arrow direction* represents the dominant velocity direction. *Colored dots* represent regions of the visual field where gaze measurements were not registered. Briefly, resident *#2* shows higher gaze dispersion in the *Translocation of Objects* exercises, which also elicit higher gaze velocities over the majority of gaze positions. **B** Average entropy and velocity across all residents for each exercise. Exercises are indicated in *green* (*Clip Applying* exercises), *blue* (*Cutting Big* exercises), and *red* (*Translocation of Objects* exercises). *Error bars* represent the standard error from the mean ($n = 15$). *Note*. The *Supplementary material* section includes complementary analyses on entropy and gaze velocity metrics, as well as additional analyses on classical eye movement metrics (fixation rate and duration) (Color figure online)

simulated laparoscopic exercises: visual exploration becomes less stereotyped (i.e., more random) and faster during more complex exercises. Perceived levels of task complexity as well as surgical performance provide an independent validation of these results. This is well illustrated by Fig. 3, which shows that gaze entropy and velocity linearly increase from the easiest to the hardest exercises. These results are consistent with previous reports of the effects of task complexity on gaze entropy [33–35], now applied to the healthcare field. Gaze entropy behaved similarly to other entropy-based metrics (i.e., pupil and hand movements), recently studied in surgical scenarios [21, 56]. Finally, our previous studies have shown (saccadic) gaze velocity to be sensitive to arousal variations (see Ref. [22] for a recent review). Since task complexity modulates arousal [57], laparoscopic complexity may have modulated the residents' arousal levels [58], and consequently their gaze velocity as well.

In order to accurately perform laparoscopic procedures, surgeons need to learn where to look to sample the necessary information. They also need to use their eye movements to coordinate their hands to reach the task-relevant locations [59]. Here, residents changed their visual exploration to compensate for the increase in task complexity. To understand this phenomenon, we must consider the information-seeking process behind psychomotor tasks. When performing easy tasks, visual exploration strategies orient surgeons' gaze to highly informative elements of the simulation [60, 61]. However, the optimal visual exploration strategies, i.e., those that will direct the surgeons' gaze only to the most relevant elements, are not known in advance [62]. It is reasonable to assume that residents might follow a far-from-optimal approach [63]. While this suboptimal approach might serve the resident well in the less complex tasks, in the most complex tasks, it causes their gaze to move constantly over different places of the visual field, not all of them useful, while searching for the points containing the information needed to perform the task.

We must note that residents prevalently performed short unfamiliar psychomotor tasks. Even though the *Translocation of Objects* exercise has also a cognitive component, as the residents had to perform several mental representations and rotations to correctly accomplish the task, future research should investigate the sensitivity of gaze metrics to detect surgeons' load variations while performing more analytical exercises (e.g., during clinical decision-making procedures).

Finally, one potential caveat of this study is that our analyses mixed together data falling inside and outside the *lookzone* areas. However, since the size of the *lookzone* areas did not differ across the exercises, there should not be any bias across conditions related to the *lookzone* factor, and therefore, we can conclude that the different spatial patterns, including the gaze falling outside the area of interest, are caused by differences in task complexity. Future analyses will benefit from efficient analysis tools able to handle dynamically changing areas of interest [64] in head-unrestrained conditions (i.e., without the use of chin/forehead support) during free viewing tasks.

Conclusions

Our data support gaze indices as a valid, sensitive, and easy way to objectively assess the task load imposed by surgical procedures and to know the surgeon's capacity to operate without risk to patients' safety or their own. Furthermore, because new low-cost eye tracker solutions offer the possibility of continuously and unobtrusively monitoring gaze behavior in real assessment protocols, our findings have potential impacts for the development of new entropy-based measures to evaluate residents' learning curves, independently of expensive virtual simulators or time-consuming expert evaluation.

Acknowledgments We thank the Tobii Group for the Eye-TrackAwards Stipend awarded to LLDS, and the IAVANTE staff (Andalusian Public Foundation for Progress and Health) for their help during the data collection. This study was funded by the Campus of International Excellence (BioTic Granada) Research Programme (Research Project V7-2015 to CDP). The project was approved by the Talentia Postdoc Program launched by the Andalusian Knowledge Agency, co-funded by the European Commission's 7th Framework Programme for Research and Technological Development—Marie Skłodowska-Curie actions—and the Andalusian Department of Economy, Innovation, Science and Employment (COFUND—Grant Agreement No. 267226 to LLDS). Research by LLDS is funded by the BBVA Foundation Program for Research, Innovation, and Cultural Creation (Grant No. 2015-2). CDP is supported by a UGR Postdoctoral Fellowship (2013 University of Granada Research Plan). Research by AC is funded by a Spanish Ministry of Economy and Competitiveness grant (PSI2012-39292 to AC). Research by LLDS, CDP, and AC is funded by a Spanish Department of Transportation grant (SPIP2014-1426 to LLDS).

Compliance with ethical standards

Disclosures Leandro L. Di Stasi, Carolina Diaz-Piedra, Héctor Rieiro, José M. Sánchez Carrión, Mercedes Martin Berrido, Gonzalo Olivares, and Andrés Catena have no conflicts of interest or financial ties to disclose.

References

1. Khuri SF, Henderson WG, Daley J, Jonasson O, Jones RS, Campbell DA Jr, Fink AS, Mentzer RM Jr, Steeger JE (2007) The patient safety in surgery study: background, study design, and patient populations. J Am Coll Surg 204:1089–1102. doi:10.1016/j.jamcollsurg.2007.03.028
2. Pham JC, Aswani MS, Rosen M, Lee H, Huddle M, Weeks K, Pronovost PJ (2012) Reducing medical errors and adverse events. Annu Rev Med 63:447–463. doi:10.1146/annurev-med-061410-121352
3. Graafland M, Schraagen JMC, Boermeester MA, Bemelman WA, Schijven MP (2015) Training situational awareness to reduce surgical errors in the operating room. Br J Surg 102:16–23. doi:10.1002/bjs.9643
4. Yurko YY, Scerbo MW, Prabhu AS, Acker CE, Stefanidis D (2010) Higher mental workload is associated with poorer laparoscopic performance as measured by the NASA-TLX tool. Simul Healthc J Soc Simul Healthc 5:267–271. doi:10.1097/SIH.0b013e3181e3f329
5. Marks J, Tacchino R, Roberts K, Onders R, Denoto G, Paraskeva P, Rivas H, Soper N, Rosemurgy A, Shah S (2011) Prospective randomized controlled trial of traditional laparoscopic cholecystectomy versus single-incision laparoscopic cholecystectomy: report of preliminary data. Am J Surg 201:369–373. doi:10.1016/j.amjsurg.2010.09.012
6. Islam A, Castellvi AO, Tesfay ST, Castellvi AD, Wright AS, Scott DJ (2011) Early surgeon impressions and technical difficulty associated with laparoendoscopic single-site surgery: a Society of American Gastrointestinal and Endoscopic Surgeons learning center study. Surg Endosc 25:2597–2603. doi:10.1007/s00464-011-1594-4
7. Canes D, Desai MM, Aron M, Haber G-P, Goel RK, Stein RJ, Kaouk JH, Gill IS (2008) Transumbilical single-port surgery: evolution and current status. Eur Urol 54:1020–1030. doi:10.1016/j.eururo.2008.07.009
8. Park A, Lee G, Seagull FJ, Meenaghan N, Dexter D (2010) Patients benefit while surgeons suffer: an impending epidemic. J Am Coll Surg 210:306–313. doi:10.1016/j.jamcollsurg.2009.10.017
9. Zheng B, Tien G, Atkins SM, Swindells C, Tanin H, Meneghetti A, Qayumi KA, Neely O, Panton M (2011) Surgeon's vigilance in the operating room. Am J Surg 201:673–677. doi:10.1016/j.amjsurg.2011.01.016
10. Amalberti R, Vincent C, Auroy Y, de Maurice GS (2006) Violations and migrations in health care: a framework for understanding and management. Q Saf Health Care 15:i66–i71. doi:10.1136/qshc.2005.015982
11. Ziv A, Wolpe PR, Small SD, Glick S (2006) Simulation-based medical education: an ethical imperative. Simul Healthc J Soc Simul Healthc 1:252–256. doi:10.1097/01.SIH.0000242724.08501.63
12. Martin JA, Regehr G, Reznick R, Macrae H, Murnaghan J, Hutchison C, Brown M (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 84:273–278. doi:10.1046/j.1365-2168.1997.02502.x
13. Hart SG (2006) NASA-task load index (NASA-TLX); 20 years later. In: Proceedings of the human factors and ergonomics society annual meeting, pp 904–908
14. Yiannakopoulou E, Nikiteas N, Perrea D, Tsigris C (2015) Virtual reality simulators and training in laparoscopic surgery. Int J Surg 13:60–64. doi:10.1016/j.ijsu.2014.11.014
15. Di Stasi LL, Renner R, Catena A, Cañas JJ, Velichkovsky BM, Pannasch S (2012) Towards a driver fatigue test based on the saccadic main sequence: a partial validation by subjective report

data. Transp Res C Emerg Technol 21:122–133. doi:10.1016/j.trc.2011.07.002
16. D'Angelo A-LD, Rutherford DN, Ray RD, Laufer S, Kwan C, Cohen ER, Mason A, Pugh CM (2015) Idle time: an underdeveloped performance metric for assessing surgical skill. Am J Surg 209:645–651. doi:10.1016/j.amjsurg.2014.12.013
17. Sodergren M, McGregor C, Farne HA, Cao J, Lv Z, Purkayastha S, Athanasiou T, Darzi A, Paraskeva P (2013) A randomised comparative study evaluating learning curves of novices in a basic single-incision laparoscopic surgery task. J Gastrointest Surg Off J Soc Surg Aliment Tract 17:569–575. doi:10.1007/s11605-012-2113-1
18. Tien T, Pucher PH, Sodergren MH, Sriskandarajah K, Yang G-Z, Darzi A (2014) Eye tracking for skills assessment and training: a systematic review. J Surg Res 191:169–178. doi:10.1016/j.jss.2014.04.032
19. Di Stasi LL, McCamy MB, Macknik SL, Mankin JA, Hooft N, Catena A, Martinez-Conde S (2014) Saccadic eye movement metrics reflect surgical residents'fatigue. Ann Surg 259:824–829. doi:10.1097/SLA.0000000000000260
20. Causer J, McCormick SA, Holmes PS (2013) Congruency of gaze metrics in action, imagery and action observation. Front Hum Neurosci. doi:10.3389/fnhum.2013.00604
21. Tien T, Pucher PH, Sodergren MH, Sriskandarajah K, Yang G-Z, Darzi A (2015) Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair. Surg Endosc Interv Tech 29:405–413. doi:10.1007/s00464-014-3683-7
22. Di Stasi LL, Catena A, Cañas JJ, Macknik SL, Martinez-Conde S (2013) Saccadic velocity as an arousal index in naturalistic tasks. Neurosci Biobehav Rev 37:968–975. doi:10.1016/j.neubiorev.2013.03.011
23. Di Stasi LL, Marchitto M, Antolí A, Cañas JJ (2013) Saccadic peak velocity as an alternative index of operator attention: a short review. Rev Eur Psychol Appliquée European Rev Appl Psychol 63:335–343. doi:10.1016/j.erap.2013.09.001
24. Krupinski EA, Tillack AA, Richter L, Henderson JT, Bhattacharyya AK, Scott KM, Graham AR, Descour MR, Davis JR, Weinstein RS (2006) Eye-movement study and human performance using telepathology virtual slides. Implications for medical education and differences with experience. Hum Pathol 37:1543–1556. doi:10.1016/j.humpath.2006.08.024
25. Alzubaidi M, Patel A, Panchanathan S, Black J, John A (2010) Reading a radiologist's mind: monitoring rising and falling interest levels while scanning chest X-rays. In: Proceedings of the SPIE international society for optical engineering. San Diego, CA, USA, pp 0F1–0F10
26. Richstone L, Schwartz MJ, Seideman C, Cadeddu J, Marshall S, Kavoussi LR (2010) Eye metrics as an objective assessment of surgical skill. Ann Surg 252:177–182. doi:10.1097/SLA.0b013e3181e464fb
27. Cao Y, Miura S, Kobayashi Y, Kawamura K, Sugano S, Fujie M (2016) Pupil variation applied to the eye tracking control of an endoscopic manipulator. IEEE Robot Autom Lett 1–1. doi:10.1109/LRA.2016.2521894
28. Tien G, Zheng B, Atkins MS (2011) Quantifying surgeons' vigilance during laparoscopic operations using eyegaze tracking. Stud Health Technol Inform 163:658–662
29. Eivazi S, Bednarik R, Tukiainen M, von und zu Fraunberg M, Leinonen V, Jääskeläinen JE (2012) Gaze behaviour of expert and novice microneurosurgeons differs during observations of tumor removal recordings. In: Proceedings of the symposium on eye tracking research and applications. New York, NY, USA, pp 377–380
30. Atkins MS, Tien G, Khan R, Meneghetti A, Zheng B (2012) What do surgeons see: capturing and synchronizing eye gaze for surgery applications. Surg Innov. doi:10.1177/1553350612449075
31. Giovinco NA, Sutton SM, Miller JD, Rankin TM, Gonzalez GW, Najafi B, Armstrong DG (2015) A passing glance? Differences in eye tracking and gaze patterns between trainees and experts reading plain film bunion radiographs. J Foot Ankle Surg 54:382–391. doi:10.1053/j.jfas.2014.08.013
32. Allsop J, Gray R (2014) Flying under pressure: effects of anxiety on attention and gaze behavior in aviation. J Appl Res Mem Cogn 3:63–71. doi:10.1016/j.jarmac.2014.04.010
33. Tole JR, Stephens AT, Harris RL, Ephrath AR (1982) Visual scanning behavior and mental workload in aircraft pilots. Aviat Space Environ Med 53:54–61
34. Di Nocera F, Camilli M, Terenzi M (2007) A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. J Cogn Eng Decis Mak 1:271–285
35. Camilli M, Nacchia R, Terenzi M, Di Nocera F (2008) ASTEF: a simple tool for examining fixations. Behav Res Methods 40:373–382
36. Zheng B, Jiang X, Atkins MS (2015) Detection of changes in surgical difficulty: evidence from pupil responses. Surg Innov. doi:10.1177/1553350615573582
37. Beatty J, Lucero-Wagoner B (2000) The pupillary system. In: Cacioppo JT, Tassinary LG, Berntson GG (eds) Handbook of psychophysiology, 2nd edn. Cambridge University Press, New York, pp 142–162
38. Bradley MM, Miccoli L, Escrig MA, Lang PJ (2008) The pupil as a measure of emotional arousal and autonomic activation. Psychophysiology 45:602–607. doi:10.1111/j.1469-8986.2008.00654.x
39. Morales JM, Di Stasi LL, Diaz-Piedra C, Morillas C, Romero S (2015) Real-time monitoring of biomedical signals to improve road safety. Springer, Palma de Mallorca, pp 89–97
40. Szulewski A, Howes D (2014) Combining first-person video and gaze-tracking in medical simulation: a technical feasibility study. Sci World J 2014:e975752. doi:10.1155/2014/975752
41. WMA (1964) Human experimentation: code of ethics of WMA. Br Med J 2:177
42. Kössi J, Luostarinen M (2009) Virtual reality laparoscopic simulator as an aid in surgical resident education: two years' experience. Scand J Surg 98:48–54. doi:10.1177/145749690909800109
43. Hoddes E, Zarcone V, Smythe H, Phillips R, Dement WC (1973) Quantification of sleepiness: a new approach. Psychophysiology 10:431–436
44. Connor J, Norton R, Ameratunga S, Robinson E, Civil I, Dunn R, Bailey J, Jackson R (2002) Driver sleepiness and risk of serious injury to car occupants: population based case control study. BMJ 324:1125
45. Di Stasi LL, Diaz-Piedra C, Suárez J, McCamy MB, Martinez-Conde S, Roca-Dorda J, Catena A (2015) Task complexity modulates pilot electroencephalographic activity during real flights. Psychophysiology. doi:10.1111/psyp.12419
46. Wilson MR, McGrath JS, Vine SJ, Brewer J, Defriend D, Masters RSW (2011) Perceptual impairment and psychomotor control in virtual laparoscopic surgery. Surg Endosc 25:2268–2274. doi:10.1007/s00464-010-1546-4
47. Hart SG, Staveland LE (1988) Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Hum Ment Workload 1:139–183
48. Corker HP, Singh P, Sodergren MH, Balaji S, Kwasnicki RM, Darzi AW, Paraskeva P (2015) A randomized controlled study to establish the effect of articulating instruments on performance in single-incision laparoscopic surgery. J Surg Educ 72:1–7. doi:10.1016/j.jsurg.2014.08.004
49. Fransen SAF, Mertens LS, Botden SMBI, Stassen LPS, Bouvy ND (2011) Performance curve of basic skills in single-incision

laparoscopy versus conventional laparoscopy: Is it really more difficult for the novice? Surg Endosc 26:1231–1237. doi:10.1007/s00464-011-2041-2
50. Borg G (1998) Borg's perceived exertion and pain scales. Human Kinetics Publishers, Champaign
51. Wickens CD, McCarley JS (2007) Applied attention theory. CRC Press, Boca Raton
52. Schiller PH, Tehovnik EJ (2005) Neural mechanisms underlying target selection with saccadic eye movements. Prog Brain Res 149:157–171. doi:10.1016/S0079-6123(05)49012-3
53. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423. doi:10.1002/j.1538-7305.1948.tb01338.x
54. Leigh RJ, Thurston SE, Tomsak RL, Grossman GE, Lanska DJ (1989) Effect of monocular visual loss upon stability of gaze. Invest Ophthalmol Vis Sci 30:288–292
55. Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70
56. Watson RA (2013) Quantification of surgical technique using an inertial measurement unit. Simul Healthc J Soc Simul Healthc 8:162–165. doi:10.1097/SIH.0b013e318277803a
57. Yerkes RM, Dodson JD (1908) The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol 18:459–482. doi:10.1002/cne.920180503
58. Di Stasi LL, McCamy MB, Pannasch S, Renner R, Catena A, Cañas JJ, Velichkovsky BM, Martinez-Conde S (2015) Effects of driving time on microsaccadic dynamics. Exp Brain Res 233:599–605. doi:10.1007/s00221-014-4139-y
59. Tramper JJ, Lamont A, Flanders M, Gielen S (2013) Gaze is driven by an internal goal trajectory in a visuomotor task. Eur J Neurosci 37:1112–1119. doi:10.1111/ejn.12107
60. Najemnik J, Geisler WS (2005) Optimal eye movement strategies in visual search. Nature 434:387–391. doi:10.1038/nature03390
61. McCamy MB, Otero-Millan J, Di Stasi LL, Macknik SL, Martinez-Conde S (2014) Highly informative natural scene regions increase microsaccade production during visual scanning. J Neurosci 34:2956–2966. doi:10.1523/JNEUROSCI.4448-13.2014
62. Gottlieb J, Oudeyer P-Y, Lopes M, Baranes A (2013) Information-seeking, curiosity, and attention: computational and neural mechanisms. Trends Cogn Sci 17:585–593. doi:10.1016/j.tics.2013.09.001
63. Jarodzka H, Balslev T, Holmqvist K, Nyström M, Scheiter K, Gerjets P, Eika B (2012) Conveying clinical reasoning based on visual observation via eye-movement modelling examples. Instr Sci 40:813–827. doi:10.1007/s11251-012-9218-5
64. Papenmeier F, Huff M (2010) DynAOI: a tool for matching eye-movement data with dynamic areas of interest in animations and movies. Behav Res Methods 42:179–187. doi:10.3758/BRM.42.1.179