Using Multimodal Data to Predict Surgeon Situation Awareness

Aurelien Lechappe, Mathieu Chollet, Jerome Rigaud, Caroline Cao

Using Multimodal Data to Predict Surgeon Situation Awareness

Aurelien Lechappe[1,2,4], Mathieu Chollet[1,2], Jerome Rigaud[3] and Caroline G.L. Cao[1,2,5]

[1] IMT Atlantique Bretagne-Pays de la Loire
[2] Laboratoire des Sciences du Numérique de Nantes (LS2N)
[3] Centre Hospitalier Universitaire de Nantes
[4] Université de Bretagne Occidentale (UBO)
[5] Wright State University

Abstract. The use of robotic surgical systems creates new team dynamics in operating rooms and constitutes a major challenge for the development of crucial non-technical skills such as situation awareness (SA). Techniques for assessing SA mostly rely on subjective assessments, observation or interviews; few utilize multimodal measures that combine physiological, behavioural, and subjective indicators. We proposed a conceptual model relating situation awareness with mental workload (MW), stress and communication. To validate this model, we collected subjective feedback, measurable behaviours and physiological signals from surgeons performing a robot-assisted radical prostatectomy procedure. Preliminary results suggest that subjective MW is a better indicator of SA than subjective stress. Physiological measures did not correlate with subjective measures of stress and MW. Results also suggest that some indicators of communication quality associated with various levels of SA tend to be linked with surgical complexity.

Keywords: Mental Workload, Stress, Communication, Physiological Signals, Robot-Assisted Surgery

1 Introduction

Many laparoscopic surgical procedures are increasingly performed with robotic assistance. Robot-assisted surgery (RAS) has many advantages for both patients and surgeons. However, the use of such devices creates new challenges inside the operating room. Unlike in traditional open surgery or laparoscopic surgery, the robot creates a physical barrier between the surgeon and the rest of the surgical team and the patient, leading to altered communication and team dynamics in the operating room (OR). In surgery, a lack of non-technical skills has been linked to a higher risk of surgical complications [1]. Non-technical skills such as communication, teamwork and decision-making are dependent on the operators' situation awareness (SA) [1]. In this complex environment, where distributed information is needed to perform the task, developing and maintaining SA can be difficult [2].

2 Situation Awareness: Definition and Assessment

2.1 Definition of situation awareness

Endsley [3] defines SA as "The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (p.36). Yule et al. [4] propose a similar definition applied to surgery: "developing and maintaining a dynamic awareness of the situation in the operating room, based on assembling data from the environment (patient, team, time, displays, equipment), understanding what they mean, and thinking ahead about what may happen next". Smith and Hancock [5] propose a definition considering the influence of other factors on SA and describes this concept as an "adaptive, externally directed consciousness directly related to stress, mental workload, and other energetic constructs that are facets of consciousness" (p.138). Therefore, the influence of other cognitive factors have to be considered for assessing SA.

2.2 Assessment of situation awareness

Several techniques for assessing SA have been developed [6]. One of these, called "Freeze-probe technique" requires the operator to take a pause during the execution of a task to answer questions about the situation (e.g., Endsley's SAGAT survey [7]). This technique has not been applied to robotic surgery but has been applied in the medical area [2]. The "Real-time probe technique" is similar except that the questions are administered while the action is being performed without pause. Situation awareness can also be assessed using self-rating techniques such as the SART questionnaire [8]. Less intrusive techniques for assessing SA include observations by experts [6]. Some observation scales focus on assessing an individual operator (e.g., NOTSS [4], ICARS [9]), while others assess each team member and then derive an overall team score from individual assessments (e.g., OTAS [10], SPLINTS [11]). Performance measures can also be used to assess SA [12], as can psychophysiological indicators such as eye-tracking measures [13], EEG [14] and cardiac activity [15].

SA is a complex cognitive process functioning at a higher level, with underlying factors such as stress and MW [16]. Thus, an alternative technique for SA assessment would be to measure some of the factors underlying this development using psychophysiological and behavioural measures. Stress, MW and communication are measurable factors that we consider in our SA model since they are described as having an influence on SA [5, 17].

3 Model and Hypothesis

We proposed a novel framework for understanding SA in RAS contexts, and modeled SA and its relationship with measurable multimodal variables, such as communication, mental workload (MW), stress, heart rate and skin conductance [18]. In this model (Fig. 1), stress and MW influence each other [19], presumably varying in a similar way. Stress influences SA [20, 21]. In particular, stressed operators may still believe they have good SA, while their mental model of the situation actually diverges from reality, increasing operational risks. Mental workload influences SA, which may vary depending on the context [22]. Additionally, MW and stress have an influence on team communication quality [21]. Finally, communication quality influences SA [17, 23]. For example, low SA may be associated with repetitions, lack of action verbalization and the emergence of off-topic discussion in surgery, while high SA may be associated with numerous action verbalization, and with proximity to the operating table [24].

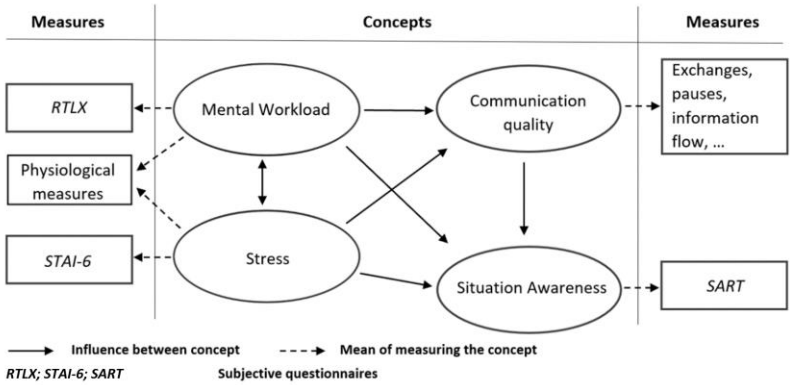

Fig. 1. Graphical representation of our SA model [18]

The main objective of this research is to validate the proposed model for assessing SA in RAS with multimodal data. We hypothesized the following:

- **H1** Physiological changes should be observed with variations in mental workload and stress levels.
- **H2** Situation awareness can be influenced by the mental workload level on the one hand, and stress on the other.
- **H3** Situation awareness is influenced by communication quality.

4 Method

To validate the model, we collected data from three surgeons of varying levels of experience, who collectively performed a total of seven radical prostatectomy procedures using a da Vinci Xi Robotic Surgical System.

4.1 Materials

Self-rating questionnaires. STAI-6, RTLX and SART, which respectively measures stress, MW and SA, were used to gather subjective feedback from the surgeons at predefined interruption points during the surgery. The STAI-6 [25] is an effective scale for assessing perceived stress [6]. The RTLX [26] is a short version of the NASA-TLX, sensitive to variations in MW in laparoscopic surgery. We removed the "physical demand" subscale as it is less relevant to our context. Finally, the SART [8] measures an operator's subjective SA.

Physiological measures. An Empatica E4 sensor was used to collect surgeons' physiological responses such as skin conductance (EDA), heart rate (HR), and heart rate variability (HRV) throughout the procedure [27]. HR is collected at a frequency of 1Hz and EDA at a frequency of 4Hz.

Communication indicators. Transcriptions of the surgeons' verbalizations were recorded and annotated with a behaviour coding scheme. In this scheme, three types of exchanges related to the procedure: Requests for task execution, action verbalizations and teaching. Other exchanges can be related to surgical or robotic materials or irrelevant to the surgery. Total ratios for each exchange type were calculated for all phases of surgery to normalize the data since phases do not have the same duration. Communication indicators of SA, such as silences, interactions related to surgical task execution, action verbalization, teaching and irrelevant discussions, were noted [24].

4.2 Procedure

At the start of each surgical procedure, the surgical team was informed of the necessary interruptions to collect data. Four interruption points were selected based on consultation with expert surgeons to ensure minimal risk of distraction or surgical complications due to the interruption. During these breaks, the surgeon completed the three self-rating questionnaires on stress, MW and SA. Throughout the surgical procedure, audio-visual and Empatica recordings were continuous.

5 Results

Scores of each questionnaire for all interruption points were analysed using repeated measures ANOVA followed by a linear regression analysis. A similar analysis was performed on the physiological data, and the RTLX and STAI-6 questionnaires for correlations with MW or stress. In addition, each surgeon was asked to rank order the four surgical phases as a function of perceived difficulty (1: easiest; 2: easy; 3: complex; 4: most complex phase). Friedman-test and Wilcoxon test were performed to analyse differences in communication patterns across these surgical phases. In these preliminary analyses, we chose to set the p value at 0.1 [32].

5.1 Self-rating questionnaires

There were no significant differences across surgical phases for the three questionnaires: SART ($F(3, 18) = 1.885$; $p = .17$); RTLX ($F(3, 18) = 1,887$; $p = .17$); STAI-6 ($F(3, 18) = 0.623$; $p = .61$)**.** The analyses also showed no correlation between SA and stress level ($t(26) = -0.761$, $p = .45$), but a significant correlation between SA and MW level ($t(26) = 6.981$, $p < .001$)**.** Mental workload and SA scores evolved similarly for each phase of the procedure.

5.2 Physiological measures

Due to artefacts on the HR data, only three surgeries with valid HR and HRV data and seven with valid EDA data were analysed. While analysis showed significant differences between phases of the procedure for HR ($F(3, 6) = 7.224$; $p = .02$) and HRV measurements ($F(3, 6) = 8.014$; $p = .016$), pairwise comparisons showed no significant difference between each phase of the procedure for these physiological measurements. Skin conductance levels were not significant different between phases of the procedure ($F(3, 18) = 2.207$; $p = .12$). Pairwise comparisons showed similar results**.** Furthermore, analyses indicated no correlation between all physiological indicators and subjective measures of stress and MW.

5.3 Communication indicators

We analysed the patterns of all communication indicators as a function of the complexity of the four surgical phases. The ratio of silent periods was not different between phases (X2(3) = 0.429, $p = .93$), neither was the ratio of verbal exchanges relative to the task execution, (X2(3) = 5.914, $p = .12$). However, one-tailed Wilcoxon tests showed significantly more interactions relative to the task execution in the most complex (4) and complex (3) phases than the easiest (1) phase (4-1: Z = 25, $p = .04$; 3-1: Z = 25, $p = .04$). There was significantly more interaction relative to task execution in the easy (2) phase than the easiest (1) phase (2-1: Z = 22, $p = .11$). Other tests showed no differences.

Analysis of action verbalization showed a significant difference between the phases (X2(3) = 6.6, $p = .09$). One-tailed Wilcoxon tests showed significantly more action verbalization exchanges in the most complex (4) phase than the easiest (1) (Z = 24, $p = .05$). They also showed more action verbalization exchanges in the most complex phase (4) than the complex phase (3) (Z = 27, $p = .016$).

No differences in teaching exchanges were found between the different phases (X2(3) = 5.435, $p = .14$). However, one-tailed Wilcoxon tests showed significantly more teaching exchanges in the easiest phase (1) than all other three phases (1-4: Z = 2, $p = .047$; 1-3: Z = 25, $p = .039$; 1-2: Z = 27, $p = .016$). Finally, analysis of irrelevant discussions showed no differences between the phases (X2(3) = 1.258, $p = .74$).

6 Discussion

6.1 Physiological measures, mental workload and stress

The results obtained do not support our first hypothesis (H1) that we should observe physiological changes related to changes in stress and MW. A first explanation of these results could be the limitations of the Empatica e4 sensor. Indeed, recent studies have shown that the data collected by this sensor could be strongly degraded due to its low recording frequency and the presence of artefacts related to the wrist-device's high sensitivity to movement [28]. These authors also discussed the validity of EDA and HRV measurements and their derived parameters measured by this sensor due to the limitations presented above. Other explanations include the small number of surgical procedures studied, as well as the large individual differences in physiological data.

We chose the Empatica e4 sensor for its ease of application and placement in the operating room, since environmental constraints severely limit the type of device that can be used. However, considering the difficulties to collect valid physiological data with this wrist-worn sensor, it would be prudent to consider the use of different measuring devices such as jackets with embedded physiological sensors instead. These alternate devices could be worn by all members of the surgical team without disrupting the sterile field in the operating environment.

6.2 Situation awareness, mental workload and stress

With regard to our second hypothesis (H2), results did not confirm the role of stress measured by questionnaire as an indicator of SA, but MW was. Although the STAI-6 scale has already been successfully used to assess stress associated with the execution of surgical tasks [6], it would appear that it is not suitable for repeated measurement of the stress felt by operators in a real surgical situation. Regarding MW and SA, our results showed that these two constructs evolved similarly for each phase of the procedure. Several hypotheses can explain these results. Endsley [29] questions the design of the SART questionnaire, which would evaluate cognitive processes close to MW. Another hypothesis supported by some authors [22, 30, 31] is that a high but reasonable MW would be necessary to achieve a high level of SA.

Future research could attempt to test our model using more appropriate measurement methods for assessment in a real surgical context. Research could be performed to find a more relevant stress measurement for this context. Future studies could also work on the design of an evaluation tool based on the SAGAT questionnaire [7], which would provide a more reliable measure of the operator's SA. The assessment tool should focus on items directly related to the situation.

6.3 Communication quality, task complexity and situation awareness

To test H3, we analysed surgeons’ communication patterns according to the complexity of the surgical phases in order to assess if some indicators associated with various SA levels are related to task complexity.

Silences and exchanges related to the task execution were analysed with the expectation that surgeons’ communications decrease, and task execution exchanges increase, with task complexity, presumably due to increased MW and stress. Results showed that surgeons performed more task execution exchanges in the two complex (3,4) phases compared to the easiest (1) and tended to perform more task execution exchanges in the easy (2) phase than the easiest (1) phase. However, no differences in silence patterns were found between the phases of different complexity. According to the literature, these results about surgeons’ verbalization and teaching exchanges would suggest better SA in more complex compared to easier phases. Indeed, high SA has been linked to increased verbalizations, and a lack of verbalizations with low SA [24]. Here, surgeons performed more verbalizations in the most complex (4) phase as compared to the complex (3) and the easiest (1) phases. Similarly, the emergence of off-topic exchanges such as teaching and irrelevant discussions has been linked to low SA [21]. Our results showed that surgeons performed more teaching exchanges in the easiest (1) phase than in other phases. Irrelevant discussions did not differ significantly across phases.

In general, most of the significant results regarding communication patterns were between the easiest and most complex phases. A potential explanation could be that the different phases of the prostatectomy procedure that we selected do not have large enough differences in terms of surgical complexity. This could also be directly due to the radical prostatectomy procedure, which could be an overly standardized procedure with small differences in complexity between each step of the procedure. This point could also explain why no significant differences were found between the different phases for the RTLX, and SART scores.

Other factors potentially influencing communication during surgery could include the number of operators in the operating room, their familiarity, or the level's experience of each operator in the considered surgery. While our results did not fully validate the hypothesis that surgeons’ communication patterns are influenced by task complexity, and that communication quality influences SA, we observed some interesting results about action verbalization and teaching exchanges that can be linked with surgeons’ SA. However, it should be noted that the interpretation of the data is to be moderated given the presence of some uncontrolled factors discussed above, and the small sample of surgeons studied. Nevertheless, these results may provide insights for preparing future studies in this field.

7 Conclusion

We presented an exploratory study where we collected preliminary data in order to validate a model of operators’ situation awareness and its relationship with mental workload, stress and team communication. This study attests to the difficulty of developing a method to assess SA using multimodal parameters such as subjective, physiological and behavioural indicators. Contrary to our conceptual model, preliminary data did not support the role of stress as an indicator of SA, while MW was supported. The physiological data were noisy and thus inconclusive. Finally, the analysis of surgeons’ communication patterns according to phase complexity raised potential trends of interest, but the presence of uncontrolled factors prevailed.

Future research to validate our model should involve a larger sample of surgeons, more reliable physiological sensors, different measures of stress and SA (e.g., SAGAT questionnaire), and perhaps with a surgical procedure with phases presenting larger differences in complexity. Finally, given the collaborative nature of the surgical team, it might be valuable to assess SA for the entire surgical team and not just the surgeon.

References

1. Pradarelli, J.C., Yule, S., Smink, D.S.: Evaluating non-technical skills in surgery. In: Dimick, J.B. and Lubitz, C.C. (eds.) Health Services Research. pp. 125–135. Springer International Publishing, Cham (2020). https://doi.org/10.1007/978-3-030-28357-5_12.
2. Gardner, A.K., Kosemund, M., Martinez, J.: Examining the feasibility and predictive validity of the SAGAT tool to assess situation awareness among medical trainees. Simul Healthc. 12, 17–21 (2017). https://doi.org/10.1097/SIH.0000000000000181.
3. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum Factors. 37, 32–64 (1995). https://doi.org/10.1518/001872095779049543.
4. Yule, S., Flin, R., Paterson-Brown, S., Maran, N., Rowley, D.: Development of a rating system for surgeons' non-technical skills. Med Educ. 40, 1098–1104 (2006) https://doi.org/10.1111/j.1365-2929.2006.02610.x.
5. Smith, K., Hancock, P.: Situation awareness is adaptive, externally directed consciousness. Human Factors: The Journal of the Human Factors and Ergonomics Society. 37, 137–148 (1995). https://doi.org/10.1518/001872095779049444.
6. Anderson-Montoya, B.L., Scerbo, M.W.: Human Factors Psychology in Surgery. In: Stefanidis, D., Korndorffer, J.R., and Sweet, R. (eds.) Comprehensive Healthcare Simulation: Surgery and Surgical Subspecialties. pp. 153–167. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-319-98276-2_14.
7. Endsley, M.R.: Situation awareness global assessment technique (SAGAT). In: Proceedings of the IEEE 1988 National Aerospace and Electronics Conference. pp. 789–795. IEEE, Dayton, OH, USA (1988). https://doi.org/10.1109/NAECON.1988.195097.
8. Selcon, S.J., Taylor, R.M.: Evaluation of the situational awareness rating technique(SART) as a tool for aircrew systems design. AGARD, Situational Awareness in Aerospace Operations 8 p(SEE N 90-28972 23-53). (1990).
9. Raison, N., Wood, T., Brunckhorst, O., Abe, T., Ross, T., Challacombe, B., Khan, M.S., Novara, G., Buffi, N., Van Der Poel, H., McIlhenny, C., Dasgupta, P., Ahmed, K.: Development and validation of a tool for non-technical skills evaluation in robotic surgery—the ICARS system. Surgical Endoscopy. 31, 5403–5410 (2017). https://doi.org/10.1007/s00464-017-5622-x.
10. Healey, A.N., Undre, S., Sevdalis, N., Koutantji, M., Vincent, C.A.: The complexity of measuring interprofessional teamwork in the operating theatre. Journal of interprofessional care. 20, 485–495 (2006).
11. Mitchell, L., Flin, R., Yule, S., Mitchell, J., Coutts, K., Youngson, G.: Development of a behavioural marker system for scrub practitioners' non-technical skills (SPLINTS system): SPLINTS behavioural marker system. Journal of Evaluation in Clinical Practice. 19, 317–323 (2013). https://doi.org/10.1111/j.1365-2753.2012.01825.x.
12. Endsley, M.: Measurement of situation awareness in dynamic systems. Human Factors. 37, 65 (1995). https://doi.org/10.1518/001872095779049499.
13. Tien, G., Atkins, M.S., Zheng, B., Swindells, C.: Measuring situation awareness of surgeons in laparoscopic training. In: Proceedings of the 2010 symposium on eye-tracking research & applications. pp. 149–152 (2010).
14. Fernandez Rojas, R., Debie, E., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S.G., Garratt, M., Abbass, H.: Encephalographic assessment of situation awareness in teleoperation of human-swarm teaming. In: Gedeon, T., Wong, K.W., and Lee, M. (eds.) Neural Information Processing. pp. 530–539. Springer International Publishing, Cham (2019). https://doi.org/10.1007/978-3-030-36808-1_58.
15. Mehta, R.K., Peres, S.C., Shortz, A.E., Hoyle, W., Lee, M., Saini, G., Chan, H.-C., Pryor, M.W.: Operator situation awareness and physiological states during offshore well control scenarios. Journal of Loss Prevention in the Process Industries. 55, 332–337 (2018). https://doi.org/10.1016/j.jlp.2018.07.010.

16. Koester, T.: Situation awareness and situation dependent behaviour adjustment in the maritime work domain. In: Hum.-Centered Comput. Cogn. Soc. Ergon. Aspects. 3, 255 (2019).
17. Nofi, A.A.: Defining and Measuring Shared Situational Awareness. Center For Naval Analyses AlexandriA VA (2000).
18. Lechappe, A., Chollet, M., Rigaud, J., Cao, C.G.L.: Assessment of situation awareness during robotic surgery using multimodal data. In: Companion Publication of the 2020 International Conference on Multimodal Interaction. pp. 412–416. Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3395035.3425205.
19. Karim, R.U.: The effect of stress on task capacity and situational awareness. Doctoral Dissertation, North Dakota State University of Agriculture and Applied Science (2012).
20. Price, T., Tenan, M., Head, J., Maslin, W., LaFiandra, M.: Acute stress causes over confidence in situation awareness. In: 2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA). pp. 1–6 (2016). https://doi.org/10.1109/COGSIMA.2016.7497778.
21. Wheelock, A., Suliman, A., Wharton, R., Babu, E.D., Hull, L., Vincent, C., Sevdalis, N., Arora, S.: The impact of operating room distractions on stress, workload, and teamwork. Annals of Surgery. 261, 1079–1084 (2015). https://doi.org/10.1097/SLA.0000000000001051.
22. Vidulich, M.A.: The relationship between mental workload and situation awareness. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 44, 3-460-3–463 (2000). https://doi.org/10.1177/154193120004402122.
23. Parush, A., Kramer, C., Foster-Hunt, T., Momtahan, K., Hunter, A., Sohmer, B.: Communication and team situation awareness in the OR: Implications for augmentative information display. Journal of Biomedical Informatics. 44, 477–485 (2011). https://doi.org/10.1016/j.jbi.2010.04.002.
24. Randell, R., Honey, S., Hindmarsh, J., Alvarado, N., Greenhalgh, J., Pearman, A., Long, A., Cope, A., Gill, A., Gardner, P., Kotze, A., Wilkinson, D., Jayne, D., Croft, J., Dowding, D.: A realist process evaluation of robot-assisted surgery: integration into routine practice and impacts on communication, collaboration and decision-making. Health Serv Deliv Res. 5, 1–140 (2017). https://doi.org/10.3310/hsdr05200.
25. Marteau, T.M., Bekker, H.: The development of a six-item short-form of the state scale of the Spielberger State—Trait Anxiety Inventory (STAI). British Journal of Clinical Psychology. 31, 301–306 (1992). https://doi.org/10.1111/j.2044-8260.1992.tb00997.x.
26. Byers, J.C.: Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? Advances in Industrial Ergonomics and Safety l: Taylor and Francis. (1989).
27. Echeverria, V., Martinez-Maldonado, R., Buckingham Shum, S.: Towards collaboration translucence: Giving meaning to multimodal group data. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–16 (2019).
28. Milstein, N., Gordon, I.: Validating measures of electrodermal activity and heart rate variability derived drom the Empatica E4 utilized in research settings that involve interactive dyadic states. Front Behav Neurosci. 14, 148 (2020). https://doi.org/10.3389/fnbeh.2020.00148.
29. Endsley, M.R.: The divergence of objective and subjective situation awareness: A meta-analysis. Journal of Cognitive Engineering and Decision Making. 14, 34–53 (2020). https://doi.org/10.1177/1555343419874248.
30. Hendy, K.C.: Situation awareness and workload: Birds of a feather? Defence and Civil Institute of Environmental Medicine (1995).
31. Lee, Y.H., Jeon, J.-D., Choi, Y.-C.: Air traffic controllers' situation awareness and workload under dynamic air traffic situations. Transportation Journal. 51, 338–352 (2012). https://doi.org/10.5325/transportationj.51.3.0338.
32. Wasserstein, R.L., Schirm, A.L., Lazar, N.A.: Moving to a world beyond 'p < 0.05'. The American Statistician (2020). DOI: 10.1080/00031305.2019.1583913