Neuroergonomics and Cognitive Engineering, Vol. 102, 2023, 103–112
https://doi.org/10.54941/ahfe1003010

Teamwork Objective Assessment Through Neurophysiological Data Analysis: A Preliminary Multimodal Data Validation

Gianluca Borghini[1,2], Vincenzo Ronca[3,2], Pietro Aricò[3,2], Gianluca Di Flumeri[1,2], Andrea Giorgi[4,2], Stefano Bonelli[5], Laura Moens[5], Lidia Castagneto Gissey[6], Maria Irene Bellini[6], Giovanni Casella[6], and Fabio Babiloni[1,2,7]

[1]Department of Molecular Medicine, Sapienza University of Rome, Rome, 00161, Italy
[2]Brainsigns SRL, Rome, 00198, Italy
[3]Department of Computer, Control, and Management Engineering "Antonio Ruberti", University of Rome "Sapienza", Rome, 00185, Italy
[4]Department of Anatomical, Histological, Forensic and Orthopedic Sciences, Sapienza University of Rome, 00161 Rome, Italy
[5]Deep Blue SRL, Rome, Italy
[6]Department of Surgery, Sapienza University of Rome, Rome, 00161, Italy
[7]School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China

ABSTRACT

Teamwork efficiency and safety are inextricably linked. The capability of having online insights and access to objective information regarding cognitive and emotional aspects of the team members using neurophysiological measures (brain activity, skin conductance, heart rate) will endow a tool which can support Instructors during the assessment and management of teams. Such neurophysiological measures can be seen as the physical interface that will enable for gathering insights about all the aspects relating to Human Factors (HFs) of the operators. The study aimed at developing and validating a methodology able to objectively measure the teamwork dynamics and efficiency. This objective has been performed in a real surgery-related context. A data-driven approach based on machine - learning (ML) and multivariate autoregressive (MVAR) models has been employed to develop the Neurometrics - based teamwork model. Such a model considered the co-variations both within each HF (e.g., Low vs High Stress) and between different HFs (e.g., Attention vs Workload) to consider their simultaneous coexistence. The results of this preliminary study demonstrated that it is possible to quantify the teamwork of operators while dealing with real tasks and endow additional information for a more accurate teams assessment and management.

Keywords: Teamwork assessment, Neurophysiological data processing, Human factors, Neuroergonomics, Multimodal data analysis, Mutual information, Surgery, Machine – learning, MVAR

INTRODUCTION

As many as two-thirds of all performance degradations were shown to arise from dysfunctional team dynamics like faulty communication, and errant decision-making rather than incompetency (Wilson et al. 1996; Saposnik et al. 2016; Sexton, Thomas, and Helmreich 2000). Although Expert's supervisions or tools like Crew\Team Resource Management (CRM\TRM) training are adopted to facilitate open communication, teamwork and safety protocols (Foushee and Helmreich 1988; Helmreich, Merritt, and Wilhelm 1999; Ashcroft, Wilkinson, and Khan 2021), they do not allow for assessing the online cognitive and emotional states of operators while they execute tasks to manage and eventually intervene on team dynamics. In fact, most state-of-the-art solutions for assessing team dynamics are still based on surveys or performance-related data (Gorman et al. 2013; 2016). The capability of having online insights and access to objective information regarding cognitive and emotional aspects of the team members (i.e., Human Factors - HF) using neurophysiological measures (brain activity, skin conductance, heart rate) will therefore provide a tool which can support Instructors during the assessment and management of teams. For instance, it can support Instructors to better train or allocate members to teams on the basis of their "cognitive" interactions and engagement (Kurmann et al. 2012) (Gogalniceanu et al. 2021). Such neurophysiological measures can be seen as the physical interface that will enable for gathering insights about all the aspects relating to HFs of the operators.

The study aimed to develop and validate a methodology able to objectively measure the teamwork dynamics and efficiency. This has been performed in a surgery-related context by recruiting professional personnel. In particular, the results described in this paper have been derived from the first experimental campaign, therefore they have to be considered as preliminary. For a comprehensive and accurate evaluation of the teamwork, a multimodal approach is necessary to consider the operators' HFs and their co-variations. "Multimodal approach" given that we employed neurophysiological (Electroencephalogram - EEG, Photoplethysmography - PPG, Electrodermal Activity - EDA), behavioural (reaction time, performance), and subjective data (self-reports) to characterise the team members accurately and from different perspectives. A data-driven approach was employed to develop the Neurometrics - based Teamwork model. In particular, such a multivariate autoregressive (MVAR) model is able to consider co-variations both within each HF (e.g., Low vs High Stress) and between different HFs (e.g., Attention vs Workload) to consider their simultaneous coexistence (Sciaraffa et al. 2021). Validation of the evidence derived from the neurophysiological data happened via the behavioural and subjective measures provided by the team members and the Subject Matter of Expert (SME) who supervised the team and their interactions.

MATERIAL AND METHODS

Participants and Experimental Protocol

The experimental group consisted of surgical teams from the Polyclinic Hospital "Umberto I" of Rome (Italy). Each team consisted of four members:

Surgeon (S), First Assistant (A1), Second Assistant (A2) and Scrub Nurse (N). In particular, two categories of team have been defined based on the Surgeon's experience: Experts and Novices. The teams were evaluated on the operative performance of an inguinal hernia repair according to Lichtenstein. In particular, based on the ideal surgery plan, four sub-teams have been identified and considered for the overall teamwork assessment: $S - A1$, $S - N$, $A1 - N$, $A2 - N$. The surgery consisted of three phases, that is, (Phase 1) the isolation of the spermatic cord, (Phase 2) the preparation of the herniated sac, and (Phase 3) the repair of the hernia with the mesh insertion. During the entire surgery, the surgeons' behavioural, subjective and neurophysiological data were collected. The Expert group included surgeons who completed their surgical training program and who are recognized in the National Specialist Register. Expert surgeons must have attained at least 70 inguinal hernia repairs as the chief operating surgeon. Novice surgeons included residents actively training in a general surgery program. The criteria for recruitment in the Novice group were set at a minimum of 3 years residency and experience in hernia surgery. Surgeons from both groups were recruited from the surgical department of Sapienza University Hospital, Department of Surgical Sciences.

Behavioural Data

An expert surgeon (Subject Matter of Expert - SME) who did not actively participate in the surgical procedures had evaluated the operative performance of the teams during the surgery. Behavioural data and team cooperation were assessed by the SME according to the following parameters:

- Overall operative time (min)
- Operative time from skin incision to spermatic cord identification and isolation (min)
- Operative time for complete hernia sac dissection (min)
- Number of stitches for mesh fixation (n)
- Number of retractors used (n)
- Number of forceps/haemostats/clamps used during the procedure (n)
- Number of gauzes used during the procedure (n).

Each parameter, which was identified as indicators of how well the operation was going, was evaluated every five minutes during the whole surgical procedure.

Self – Reports

In order to have a subjective measure of the quality of collaboration and performance of the surgical team, questionnaires were filled out during and right after the end of the surgery. In particular, the SME filled out a questionnaire every five minutes and at the end of each phase of the surgery. The questionnaire contained a question about the observed quality of collaboration between the surgical team members, which was rated on a 5-point Likert scale. Right after the end of the surgery, all four members filled out a survey to evaluate the quality of the overall collaboration and performance of the team. In addition to the overall collaboration of the team, they also rated the quality of collaboration experienced with each of the other team members.

Neurophysiological Data

The team members' brain activity (Electroencephalogram - EEG), heart activity (Photoplethysmography - PPG) and skin sweating (Electrodermal Activity - EDA) were recorded while dealing with the surgery.

EEG Data Collection and Analysis

The EEG was collected simultaneously and synchronously from the four team members by the digital monitoring system Mindtooth Touch (Mindtooth project G.A. 950998) with a sample rate of 125 Hz, referenced to the left mastoid and grounded on the right mastoid, and the brain scalp positions were AFz, AF3, AF4, AF7, AF8, Pz, P3, and P4. Water-based electrode contact was considered good when impedance values were below 50 KOhm (Kappenman and Luck 2010). The EEG was firstly band-pass filtered with a fifth-order Butterworth filter in the interval 2–30 Hz. The blink artefacts were detected by means of the Reblinca method (Gianluca Di Flumeri et al. 2016) and corrected by leveraging the ocular component estimated through a multi-channel Wiener Filter (MWF) (Somers, Francart, and Bertrand 2018). EEG signals were segmented into epochs of 1 s, and if the EEG signal amplitude exceeded ± 80 (μV), it was marked as an artefact (threshold criterion). From the artifact-free EEG, the Global Field Power (GFP) was calculated for the EEG frequency bands of interest (theta, alpha and beta). The bands were defined according to each member's Individual Alpha Frequency (IAF) value (Klimesch 1999). Since the alpha peak is mainly prominent during rest conditions, the operators were asked to keep their eyes closed for a minute before starting the surgery. Such a condition was therefore used to estimate the IAF value for each participant.

PPG and EDA Data Collection

The PPG and EDA were collected by the Shimmer3 GSR3+ Unit (Shimmer sensing, Ireland) with a sample rate of 64 Hz. Due to the hygienic restrictions, the device was fixed on the left ankle with the corresponding electrodes placed around the toes. In particular, the EDA electrodes were located around the index and middle toes, while the PPG sensor was around the hallux.

PPG Data Analysis

Raw PPG data were digitally filtered by using a 5th-order Butterworth band-pass filter (1–5 Hz) to exclude the continuous component, as well as slow signal drifting, and emphasise the PPG signal patterns related to the pulse. At this point, the Pan–Tompkins algorithm (Pan and Tompkins 1985) was employed to detect the pulse-related peaks so as to calculate the Inter-Beat Intervals (IBI signal). The so-obtained IBI signals were processed to remove any type of artefacts by means of the HRVAS Matlab suite (Ramshur 2010). At this point, clean IBI signals were processed to estimate the Heart Rate (HR) as "Beats per minute". Then, the HR values of each operator were normalised by subtracting the individual mean baseline HR value and dividing the result by the HR individual standard deviation. The IBI signal was also analysed to estimate the Heart Rate Variability (HRV). In particular, the HRV was analysed in the frequency domain by computing the Lomb–Scargle periodogram (Ruf 1999) of the IBI signal. Analysis has shown that the Lomb–Scargle periodogram can produce a more accurate estimate of the Power Spectrum Density (PSD) than Fast Fourier Transform methods for typical HR data. Since the HR data are unevenly sampled data, another advantage of the Lomb–Scargle method is that in contrast to Fast Fourier Transform–based methods, it is able to be used without the need to resample and detrend the RR data (Clifford and Tarassenko 2005). According to the scientific literature, the PSD of the HRV signal was computed over Low (LF: 0.04–0.15 Hz) and High Frequencies (HF: 0.15–0.4 Hz), and then the LF/HF ratio was computed as a relevant indicator of HRV (Peabody et al. 2023). The LF/HF values of each operator were normalised by subtracting the individual mean baseline LF/HF value and dividing the result by the individual HR standard deviation.

EDA Data Analysis

The EDA was first low-pass-filtered with a cut-off frequency of 1 Hz, and then an artefact correction Matlab tool was applied in order to remove discontinuities and spurious peaks from the signals. Lastly, the signals were processed by using the Ledalab suite (Bach 2014). A continuous decomposition analysis (Cohen 2013) was applied in order to estimate the tonic (SCL) and phasic (SCR) components. The SCL is the slow-changing component of the EDA signal, mostly related to the global arousal of the participant. On the contrary, the SCR is the fast-changing component of the EDA signal, usually related to single stimuli reactions (Borghini et al. 2020). The SCL and SCR values of each operator were normalised by subtracting the individual mean baseline values and dividing the result by the corresponding standard deviation.

Neurometrics - Based Teamwork Model

The cooperation can be considered as the output of a MVAR system composed of the interaction between behavioural, affective and cognitive mechanisms belonging to two or more than two people that are cooperating. The considered measurements corresponded to the neurophysiological synchronous time-series describing the affective and cognitive state of each team member (Sciaraffa et al. 2021). In particular, the cognitive state was associated with the Mental Workload index, computed as the GFP Theta on the parietal EEG channels (Borghini et al. 2014) (Sciaraffa et al. 2022; Rooseleer et al. 2022). The affective state was associated with the Approach-Withdrawal index, which corresponds to the unbalance between the right and left frontal brain activity (G. Di Flumeri et al. 2017; Giorgi et al. 2021) (Simonetti et al. 2023). If any sample of the time series thus obtained was missing (i.e., was NaN) due to the artefact rejection methods application, the missing values were substituted by the spline interpolation of the nearest epochs. Finally, each time-series was normalised according to the z-score. To explain the interactions between the different components of the system, the Mutual Information was computed. Mutual Information allows to discover the maximum information shared between two random variables even multivariate (Lewis, Weekes, and Wang 2007). In this case, each variable included the two different time-series describing the affective and cognitive state of team members cooperating. It was already proved that there is a statistical relationship between cooperation effectiveness and the exchange of information between variables: higher mutual information values are associated with enhanced cooperation (Parasuraman 2000). Therefore, the Teamwork index was obtained by computing the Mutual Information on 90-seconds buffers shifted by one-second. In particular, the Teamwork model was defined by evaluating the cooperation between the four surgeons' sub - teams.

RESULTS

Self - Reports

Figure 1 represents the results derived from the analysis of the subjective data gathered from each team member and related to the perceived collaboration between the above-mentioned sub - teams. The results indicated that the best cooperative interactions were perceived between the S - N, S - A1 and A1 - N. The average collaboration was 4.8 on a 5-point Likert scale by the team members.

Neurometrics – Based Teamwork Assessment

Figure 2 shows the Neurometrics – based Teamwork index derived from the team members' neurophysiological measurements and computed as the

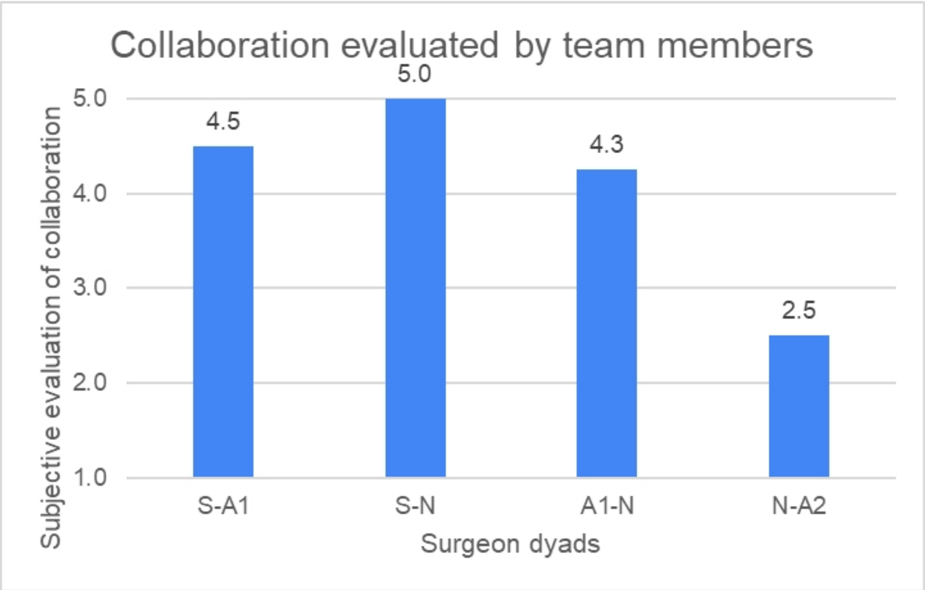

Figure 1: Subjective evaluation of the collaboration between the sub-teams provided by each team member at the end of the surgery.

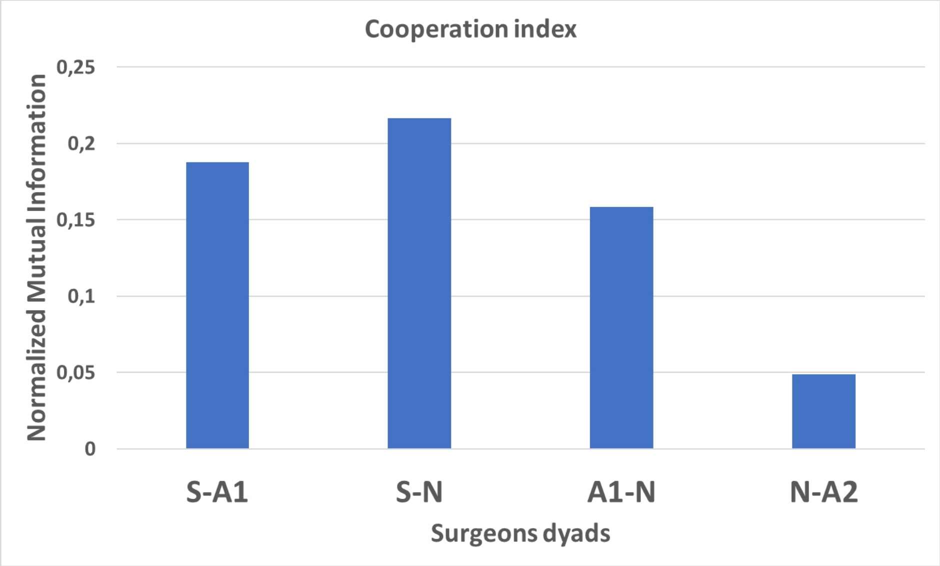

Figure 2: The neurophysiological cooperation index estimated for each surgeons dyad along the entire surgery procedure.

Mutual Information for each sub -team. It can be observed how the most relevant cooperation was found between S - A1, S - N and A1 - N. It has to be noted the similarity between the neurophysiological teamwork assessment and the subjective evaluation (Figure 1).

CONCLUSION

The presented study aimed to develop and validate a methodology able to objectively measure the teamwork dynamics and efficiency. This was performed in a surgery-related context by evaluating the teamwork efficiency through subjective, behavioural and neurophysiological measurements. The results were related to the first experimental campaign. Although they have been derived by only a team, they look very promising. In fact, the subjective (Figure 1) and neurophysiological (Figure 2) teamwork assessment exhibited the same trends. The capability of obtaining online information on the cooperation among the team members without interrupting them and interfering with tasks execution, demonstrated by the neurophysiological measures, is undoubtedly and enormous advantage and would allow the instructors to eventually intervene accordingly and promptly in order to guarantee the proper safety and outcome of the intervention. The data analysis was conducted by considering surgeons sub - teams identified according to the standard procedure for this kind of surgery intervention. As confirmation of this assumption, the results derived from the neurophysiological measurements revealed that the most relevant cooperation was evaluated between the S-A1 and S-N dyads. The next step of the study is the development of the methodology to combine the different sub – teams neurophysiological cooperation indexes. In this regards, ML and MVAR models will be considered to consider the role and rank of each member within the team. The final aim is to provide a real -time overall Neurometrics – based Teamwork index to quantify the whole team while dealing with surgery activities.

ACKNOWLEDGMENT

This work was co-funded by the grants "CHALLENGES: CompreHensive frAmework to aLLEge and aNalyse surGEons' Stress" (Bando Ateneo Medio 2021) and "BRAINORCHESTRA: Multimodal teamwork assessment through hyperscanning technique" (Bando Sapienza Medio 2022) provided by the Sapienza University of Rome, respectively, to Giovanni Casella and Gianluca Borghini. The funds provided by the European Commission through the Horizon 2020 projects "FITDRIVE: Monitoring devices for overall FITness of Drivers" (GA n. 953432) and "MINDTOOTH: Wearable device to decode human mind by neurometrics for a new concept of smart interaction with the surrounding environment" (GA n. 950998), and H2020-SESAR-2019-2 projects: Transparent artificial intelligence and automation to air traffic management systems, "ARTIMATION" (GA n. 894238).

REFERENCES

Ashcroft, James, Aimee Wilkinson, and Mansoor Khan. 2021. "A Systematic Review of Trauma Crew Resource Management Training: What Can the United States and the United Kingdom Learn From Each Other?" *Journal of Surgical Education* 78 (1): 245–64.

Bach, Dominik R. 2014. "A Head-to-Head Comparison of SCRalyze and Ledalab, Two Model-Based Methods for Skin Conductance Analysis." *Biological Psychology* 103 (December): 63–68.

Borghini, Gianluca, Laura Astolfi, Giovanni Vecchiato, Donatella Mattia, and Fabio Babiloni. 2014. "Measuring Neurophysiological Signals in Aircraft Pilots and Car Drivers for the Assessment of Mental Workload, Fatigue and Drowsiness." *Neuroscience and Biobehavioral Reviews* 44 (July): 58–75.

Borghini, Gianluca, Gianluca Di Flumeri, Pietro Aricò, Nicolina Sciaraffa, Stefano Bonelli, Martina Ragosta, Paola Tomasello, et al. 2020. "A Multimodal and Signals Fusion Approach for Assessing the Impact of Stressful Events on Air Traffic Controllers." *Scientific Reports* 10 (1): 1–18.

Clifford, Gari D., and Lionel Tarassenko. 2005. "Quantifying Errors in Spectral Estimates of HRV Due to Beat Replacement and Resampling." *IEEE Transactions on Bio-Medical Engineering* 52 (4): 630–38.

Cohen, Jacob. 2013. "Statistical Power Analysis for the Behavioral Sciences." *Statistical Power Analysis for the Behavioral Sciences*, May.

Flumeri, G. Di, P. Arico, G. Borghini, N. Sciaraffa, A. G. Maglione, D. Rossi, E. Modica, et al. 2017. "EEG-Based Approach-Withdrawal Index for the Pleasantness Evaluation during Taste Experience in Realistic Settings." In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*.

Flumeri, Gianluca Di, Pietro Aricò, Gianluca Borghini, Alfredo Colosimo, and Fabio Babiloni. 2016. "A New Regression-Based Method for the Eye Blinks Artifacts Correction in the EEG Signal, without Using Any EOG Channel." *Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*.

Foushee, H. Clayton, and Robert L. Helmreich. 1988. "Group Interaction and Flight Crew Performance." In *Human Factors in Aviation*, 189–227. Elsevier.

Giorgi, Andrea, Vincenzo Ronca, Alessia Vozzi, Nicolina Sciaraffa, Antonello di Florio, Luca Tamborra, Ilaria Simonetti, et al. 2021. "Wearable Technologies for Mental Workload, Stress, and Emotional State Assessment during Working-Like Tasks: A Comparison with Laboratory Technologies." *Sensors 2021, Vol. 21, Page 2332* 21 (7): 2332.

Gogalniceanu, Petrut, Francis Calder, Chris Callaghan, Nick Sevdalis, and Nizam Mamode. 2021. "Surgeons Are Not Pilots: Is the Aviation Safety Paradigm Relevant to Modern Surgical Practice?" *Journal of Surgical Education* 78 (5): 1393–99.

Gorman, Jamie C., Melanie J. Martin, Terri A. Dunbar, Ronald H. Stevens, and Trysha Galloway. 2013. "Analysis of Semantic Content and Its Relation to Team Neurophysiology during Submarine Crew Training." In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8027 LNAI:143–52. Springer, Berlin, Heidelberg.

Gorman, Jamie C., Melanie J. Martin, Terri A. Dunbar, Ronald H. Stevens, Trysha L. Galloway, Polemnia G. Amazeen, and Aaron D. Likens. 2016. "Cross-Level Effects between Neurophysiology and Communication during Team Training." *Human Factors* 58 (1): 181–99.

Helmreich, Robert L., Ashleigh C. Merritt, and John A. Wilhelm. 1999. "The Evolution of Crew Resource Management Training in Commercial Aviation." *International Journal of Aviation Psychology* 9 (1):19–32.

Kappenman, Emily S., and Steven J. Luck. 2010. "The Effects of Electrode Impedance on Data Quality and Statistical Significance in ERP Recordings." *Psychophysiology* 47 (5): 888..

Klimesch, W. 1999. "EEG Alpha and Theta Oscillations Reflect Cognitiveand Memory Performance: A Review and Analysis." *Brain Research. Brain Research Reviews* 29 (2–3): 169–95.

Kurmann, Anita, Franziska Tschan, Norbert K. Semmer, Julia Seelandt, Daniel Candinas, and Guido Beldi. 2012. "Human Factors in the Operating Room - The Surgeon's View." *Trends in Anaesthesia and Critical Care* 2 (5): 224–27.

Lewis, Richard S., Nicole Y. Weekes, and Tracy H. Wang. 2007. "The Effect of a Naturalistic Stressor on Frontal EEG Asymmetry, Stress, and Health." *Biological Psychology* 75 (3): 239–47.

Pan, J., and W. J. Tompkins. 1985. "A Real-Time QRS Detection Algorithm." *IEEE Transactions on Bio-Medical Engineering* 32 (3):230–36.

Parasuraman, Raja. 2000. *The Attentive Brain*.

Peabody, Jeremy E., Rebecca Ryznar, Markus T. Ziesmann, Lawrence Gillman, Jeremy E. Peabody, Rebecca J. Ryznar, Markus T. Ziesmann, and Lawrence M. Gillman. 2023. "A Systematic Review of Heart Rate Variability as a Measure of Stress in Medical Professionals." *Cureus* 15 (1).

Ramshur, John T. 2010. *Design, Evaluation, and Application of Heart Rate Variability Analysis Software (HRVAS)*. University of Memphis.

Rooseleer, Frédéric, Barry Kirwan, Elizabeth Humm, and Diana Paola Moreno_Alarcon. 2022. "The Application of Human Factors in Wake Vortex Encounter Flight Simulations for the Reduction of Flight Upset Risk and Startle Response." *Human Error, Reliability, Resilience, and Performance* 33.

Ruf, T. 1999. "The Lomb-Scargle Periodogram in Biological Rhythm Research: Analysis of Incomplete and Unequally Spaced Time-Series." *Biological Rhythm Research* 30 (2): 178–201.

Saposnik, Gustavo, Donald Redelmeier, Christian C. Ruff, and Philippe N. Tobler. 2016. "Cognitive Biases Associated with Medical Decisions: A Systematic Review." *BMC Medical Informatics and Decision Making* 16 (1): 1–14.

Sciaraffa, Nicolina, Gianluca Di Flumeri, Daniele Germano, Andrea Giorgi, Antonio Di Florio, Gianluca Borghini, Alessia Vozzi, et al. 2022. "Validation of a Light EEG-Based Measure for Real-Time Stress Monitoring during Realistic Driving." *Brain Sciences 2022, Vol. 12, Page 304* 12 (3): 304.

Sciaraffa, Nicolina, Jieqiong Liu, Pietro Aricò, Gianluca Di Flumeri, Bianca M. S. Inguscio, Gianluca Borghini, and Fabio Babiloni. 2021. "Multivariate Model for Cooperation: Bridging Social Physiological Compliance and Hyperscanning." *Social Cognitive and Affective Neuroscience* 16 (1–2): 193–209.

Sexton, Bryan, Eric Thomas, and Robert L. Helmreich. 2000. "Error, Stress, and Teamwork in Medicine and Aviation: Cross Sectional Surveys." *Ugeskrift for Laeger* 162 (19): 2725.

Simonetti, Ilaria, Luca Tamborra, Andrea Giorgi, Vincenzo Ronca, Alessia Vozzi, Pietro Aricò, Gianluca Borghini, et al. 2023. "Neurophysiological Evaluation of Students' Experience during Remote and Face-to-Face Lessons: A Case Study at Driving School." *Brain Sciences 2023, Vol. 13, Page 95* 13 (1): 95.

Somers, Ben, Tom Francart, and Alexander Bertrand. 2018. "A Generic EEG Artifact Removal Algorithm Based on the Multi-Channel Wiener Filter." *Journal of Neural Engineering* 15 (3): 36007.

Wilson, Ross McL, William B Runciman, Robert W Gibberd, Bernadette T Harrison, and John D Hamilton. 1996. "Quality in Australian Health Care Study." *Medical Journal of Australia* 164 (12): 754–754.