(Article begins on next page)

Promoting operator's wellbeing in Industry 5.0: detecting mental and physical fatigue

Valeria Villani, Marta Gabbi and Lorenzo Sabattini

Abstract— **Building on the benefits of Industry 4.0, Industry 5.0 promotes human-centricity of factories, placing operator's wellbeing at the center of production. Production processes are expected to be designed around operator's needs, for a more sustainable contribution of industry to society. In this broad context, in this paper we consider the problem of monitoring operator's condition and, specifically, detecting any mental or physical fatigue they might be experiencing in workplaces. Indeed, if the onset of any source of fatigue is monitored, it is possible to introduce assistive strategies that preserve productivity, on the one side, and operator's wellbeing, on the other side. To achieve this goal, we detect mental and physical fatigue conditions via physiological monitoring, by means of a wearable device that measures cardiac activity. We design an experimental protocol such that test participants are exposed to mental and physical fatigue. Their heart rate variability is then extracted and analysed to discriminate among rest, mental fatigue, physical fatigue and joint mental and physical fatigue. The achieved results show that statistically significant differences can be found in time-domain metrics. Moreover, the analysis of the empirical distribution functions shows, for each metrics, the conditions that exhibit the greatest differences and, hence, that can be distinguished more accurately. However, results show also that, in the presence of physical fatigue, it is difficult to detect the presence of additional mental fatigue.**

I. INTRODUCTION

In the very recent years, Industry 5.0 has become the vision paradigm for the factories of the future. In general terms, it complements and extends Industry 4.0, leveraging the technological pillars introduced by the previous paradigm to provide a vision of industry that aims beyond efficiency and productivity as the sole goals. In particular, this novel vision places worker's wellbeing at the centre of the production process and technology is used to provide prosperity beyond jobs and growth [1].

Among the technological pillars introduced with Industry 4.0, collaborative robotics plays an important role. Collaborative robots allow for combining advantages of automation, such as accuracy and repeatability, with the flexibility and cognitive soft skills of humans. Human-robot collaboration, hence, results in a team with such advanced capabilities that cannot find otherwise solely in robots or human operators. As a result, the market for collaborative robots has experienced robust growth in the last years and it is predicted to increase further in the coming years[1]. The main assumption of collab-

Authors are with the Department of Sciences and Methods for Engineering (DISMI), University of Modena and Reggio Emilia, 42100 Reggio Emilia, Italy. Emails: {valeria.villani, lorenzo.sabattini}@unimore.it

[1]https://www.interactanalysis.com/the-collaborative-robot-market-2021-28-grounds-for-optimism-after-a-turbulent-two-years/

orative robots is that robots are in charge of those tasks that are more demanding for users, while not requiring advanced reasoning capabilities, which are left to humans. This is the case of physically demanding and tedious repetitive tasks, like assembly, material dispensing, material transportation, load lifting and packaging, just to cite few examples [2]. On the other side, human operators are in charge of supervision tasks and other cognitively demanding activities.

Notwithstanding, in actual industrial workplaces, and in particular in small and medium-sized enterprises (SMEs), it is quite common that many operations are still executed manually. This is mostly due to the fact that roles might be allocated in a loose manner, so that operators are in charge of both supervision and operational tasks, over different workcells. Additionally, small-batch and flexible production, typical of SMEs, require generalization capabilities and frequent adjustment to the production parameters and process, which are unattainable for robots. Moreover, collaborative robots commonly used in SMEs have low payload; hence, most of load lifting is still up to operators. As a result, human operators are often overwhelmed being responsible for physically and cognitively demanding tasks, with a consequent decrease in production performance and personal wellbeing.

To overcome this issue, a possible solution can be found in making robots aware of operator's fatigue condition and adapting collaboration accordingly. This would reproduce successful collaboration schemes typical in human-human interactions: when two (non conflicting) human agents collaborate at the same task and one of them is not able to cope with the task temporarily, the other adapts their behavior to keep overall team performance constant and allow the fatigued teammate some rest. Reproducing this collaboration paradigm when teaming up with a robot implies that, on the one side, the robot is able to recognize fatigue onset, and, on the other side, proper beahvior adaptation strategies are implemented. Regarding the latter issues, examples of adaptation strategies resort to dynamically allocating tasks [3], [4] or simplifying the task in charge to the human [5], [6]. Conversely, this work focuses on the former issue, that is making the robot aware of any mental or physical fatigue the human operator is currently experiencing.

Some studies in the literature have already reported that variations in cardiac activity can be observed during either moderate and intense physical exercise [7], [8] or cognitive load [9]–[12]. While in previous studies the two kinds of fatigue have been considered separately, in this paper we consider the case that factory operators might be exposed

to different sources of fatigue at the same time. Hence, we propose an approach based on physiological monitoring to detect the occurrence of joint mental and physical fatigue. Specifically, we use a wearable device to measure cardiac activity in an unobtrusive manner, such that measurement does not limit subject's freedom to move and the approach could be used in real industrial working scenarios, perspectively. A commercial chest band is used to this end and heart rate variability (HRV) is analysed from measured heart activity. We set up an experimental protocol that consists in inducing mental and physical fatigue in test participants and we perform thorough statistical analysis to assess whether cardiac activity can be successfully used to discriminate among rest and mental and physical fatigue conditions.

The achieved results show that statistically significant differences exist when considering time-domain analysis of HRV. However, a full characterization of different combinations of fatigue sources cannot be achieved considering cardiac activity, only. To this end, including information referred to the collaborative task the human subject is in charge of and metrics extracted from other physiological signals, such as electromyographic or electrodermal activity, could be beneficial.

The rest of the paper is organized as follows. Section II presents the study we carried out and the experimental methodology implemented to answer the research question. Then, Sec. III presents and discusses the achieved results, while Sec. IV follows with some concluding remarks.

II. METHODS

The aim of our experiments was to assess how accurately cognitive and physical fatigue can be detected and distinguished by using commercial wearable devices. To this end, we considered a chest band, worn by test participants. Cognitive and physical effort were elicited and subject's cardiac activity was analyzed. In this section, we describe the experiments carried out to answer the considered research question.

A. Experimental protocol

The experiment consisted of four parts, with a total duration of approximately 27 minutes. Before the beginning of the experiment, participants wore the sensor for cardiac activity monitoring and were introduced to the test by the experimenter. Figure 1 depicts the organization of the experimental protocol.

It started with keeping the test participant at rest to record their baseline activity. In this phase, which lasted 3 minutes, the subject was seated and was not exposed to any external stimuli. The following two phases aimed to induce mental or physical fatigue.

As regards mental fatigue, we used the same test in [12]. Specifically, subjects were exposed to cognitive stressors consisting in a combination of memory, mathematics and visual tasks. An HTML-based application was developed and shown to test participants on a laptop. The aim of the application was to present the test participant with some cognitive tests, meant as stressors. Cognitive stressors are external stimuli demanding significant mental engagement and attention. They are used to elicit fatigue and assess cognitive effort [13], [14]. In our experiment, we considered four different stressors, presented for 1 minute each. They are shown in Fig. 1(b). In particular, the application started with the Stroop test with five different colors [15]. Then, a fast counting test followed, consisting in presenting randomly a group of dots (from four to seven) and asking the subject to enter the number of dots on the keyboard, under a strict time pressure (each group was shown for 1.5 seconds). The following test consisted in math calculation of multiplications with operands from 1 to 13. After that, the 2-back test was shown [16], consisting in a sequence of different figures shown to test subject who was asked to detect whether the current image had appeared two images back in the sequence. Next, the Stroop test was shown again. The test finished with a last round of math calculations with a higher difficulty level, since operands were from 9 to 20. Further details can be found in [12]. The total duration of the mental fatigue inducing test is 6 minutes.
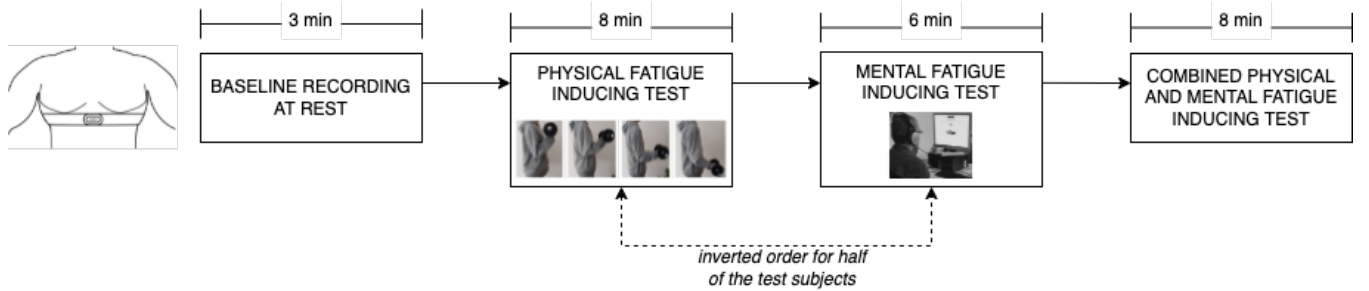
The physical fatigue inducing test consisted in isometric exercises for biceps and triceps contraction. Specifically, the test subject was asked to hold a dumbell with different angles of elbow flexion for 1 minute per angle. The angles we considered are $0°$, $45°$, $90°$ and $135°$ and are shown in Fig. 1(c). The dumbell was hold with the dominant arm. The isometric task was repeated twice, with the subject in a standing (as in the case of Fig. 1(c)) and sitting position. As a result, this part of the experiment lasted 8 minutes.

Finally, the last part of the experiment aimed to both induce mental and physical fatigue at the same time. To this end, we combined the two above mentioned tests: test participants were asked to perform the mental fatigue induction test, while holding the dumbell as in the physical fatigue induction test in the sitting position. In this part of the test, the dumbell was hold with the non dominant arm, while the dominant arm was used to enter inputs required in the mental fatigue induction test. This part of the experiment lasted 8 minutes; the fast counting test and math calculations with higher difficulty level were administered twice, to extend the duration of the mental fatigue induction test to that of the physical fatigue induction test.

A pause of 30 seconds was introduced between two consecutive tests, to let the participant have some rest and position themselves for the next test. Moreover, after the initial resting condition, half of the test participants were administered the physical fatigue inducing test before the mental one, whereas the other half performed the mental test before the physical one. The combined test was administered at the end of the experiment for all the participants.

B. Experimental setup

Cardiac activity was recorded using the Polar H10 (Polar Electro Inc., Bethpage, NY, USA) chest strap. It records cardiac electrical surface activity by means of electrodes and has been reported as trustworthy device to record heart

(a) Organization of the tests.
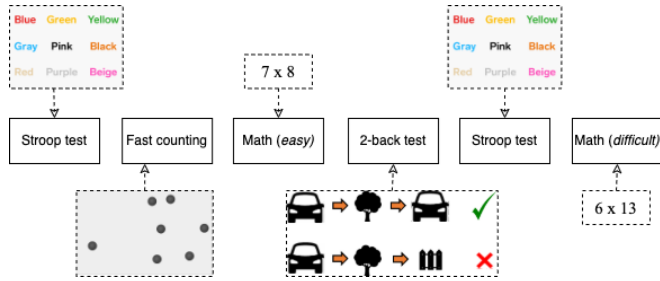
(b) Mental fatigue inducing test. All the tests were integrated in a HTML-based application, run on a laptop, as shown in Fig. 1(a). Each single test lasted 1 minute.

(c) Physical fatigue inducing test. Each isometric position was hold for 1 minute. The test was repetaed in sitting position.

Fig. 1. Experimental protocol, and organization of mental and physical inducing tests.

rate [8], [11], [12], [17]. In particular, in [8], [11], [18] this device was successfully compared to ECG recordings. The device was connected through Bluetooth to a laptop and data were accessed using the GATT (Generic Attribute Profile) protocol. As regards statistical analysis, MATLAB R2021a and its Statistics and Machine Learning Toolbox were used.

C. Test participants

A total of 30 (15 females, 15 males, age rage: 18-38 years, mean age: 23.5 years) volunteer students were enrolled in the experiment. All of them were completely naïve to the experimental task and goals and they were all new to the test for mental fatigue induction. In the physical fatigue induction test, 16 subjects lifted a 3 kg dumbbell, while the remaining 14 lifted a 5 kg dumbbell.

The study protocol followed the Helsinki declaration and compliance to participate in the study was obtained from written informed consent before starting the experiment. All the data were analyzed and reported anonymously.

D. Heart rate analysis

Arousal of physical and mental fatigue was estimated from cardiac activity measured with the chest band. In particular, heart rate was analyzed starting from RR interval time series, which is the series of time of occurrence of heart beats [19]. It collects time intervals occurring between consecutive heart beats, where the occurrence of a heart beat is detected from R peaks [20]. Thus, the RR series is defined as

$$RR = \{RR_k; k = 1, \ldots, N\} = \{R_{k+1} - R_k; k = 1, \ldots, N\},$$

where R_k is the instant of occurrence of the k−th beat and N is the number of beats occurring during a measurement session.

While heart rate undergoes normal physiological oscillations, the way it varies provides relevant information on the autonomic neural regulation of the heart and the circulatory system, which, ultimately, is influenced by fatigue states. To this end, assessment of HRV turns out to be useful. HRV is typically analyzed considering a set of established metrics in the time and frequency domain [19], [20]. Time-domain metrics include statistical indices such as the mean value (*mean RR*) and the standard deviation (*SDNN*) of the RR series, the root mean square of successive differences (*RMSSD*), and the percentage number of consecutive (normal) intervals differing by more than 50 ms in the entire recording (*pNN50*). Frequency-domain metrics account for RR series power in the low frequency band (*LF*, $0.04 − 0.15$Hz), the high frequency band (*HF*, $0.15 − 0.40$Hz) and their ratio (*LF/HF ratio*).

III. RESULTS AND DISCUSSION

Data were analyzed to investigate whether the four considered experimental conditions could be detected from cardiac activity.

To this end, measured RR series were first visually inspected to check for the presence of measurement noise or ectopic beats. Ectopic beats were replaced with interpolated intervals computed as the average over a ten intervals window, centered on the ectopic beat [21]. Then, the HRV metrics introduced in Sec. II-D were computed, for each test participant, in the four experimental conditions: rest, mental fatigue, physical fatigue, and mental and physical fatigue at

HRV metrics	Rest	Mental fatigue	Physical fatigue	Mental and physical fatigue	p-value
mean RR [ms]	800.0 ± 154.0	752.2 ± 142.9	677.8 ± 121.5	696.7 ± 123.1	0.0028*
SDNN [ms]	95.9 ± 68.5	63.0 ± 20.0	64.3 ± 19.6	52.4 ± 20.3	0.0002**
RMSSD [ms]	85.1 ± 105.6	42.4 ± 18.5	33.3 ± 20.1	32.4 ± 21.7	0.0007**
pNN50 [%]	27.0 ± 17.9	19.6 ± 14.8	9.7 ± 9.3	9.7 ± 8.3	0.0000**
LF [ms^2]	46.1 ± 8.3	45.4 ± 6.7	48.0 ± 6.8	47.3 ± 7.1	0.5323
HF [ms^2]	53.9 ± 8.3	54.5 ± 6.7	52.0 ± 6.8	52.7 ± 7.1	0.5323
LF/HF ratio	0.9 ± 0.3	0.9 ± 0.3	0.9 ± 0.3	0.9 ± 0.3	0.5743

* The null hypothesis that all the means are equal is rejected with significance level $p < 0.05$.
** The null hypothesis that all the means are equal is rejected with significance level $p < 0.01$.

	R-M	R-P	R-(M&P)	M-P	M-(M&P)	P-(M&P)
mean RR		**	*			
SDNN	**	**	**			
RMSSD	*	**	**			
pNN50		**	**	*	*	

* The null hypothesis that the means of the pair are equal is rejected with
significance level $p < 0.05$.
** The null hypothesis that the means of the pair are equal is rejected
with significance level $p < 0.01$.

the same time. The achieved results were analyzed considering different statistical tests, for a thorough understanding of relationship between cardiac activity and fatigue conditions.

A. Analysis of variance (ANOVA)

To assess whether difference could be found among these metrics we carried out a one-way multivariate analysis of variance (ANOVA). Table I reports the results of such analysis. Specifically, therein we report the mean value and standard deviation for time-domain and frequency-domain metrics at rest and under fatigue, computed over all test participants. The right column in the table reports the p-value for ANOVA. These results show that different fatigue conditions reflect in variations in all the HRV time-domain metrics. On the contrary, statistically significant differences are not found in frequency-domain metrics.

Overall, these results show that there is a statistically significant difference between the means of the HRV time-domain metrics. In other words, results in Table I suggest that, on average, at least one of the considered experimental conditions is different from the others, for all time-domain metrics. However, this analysis does not provide details on which conditions are different from each other.

B. Multiple pairwise comparisons

To further investigate differences existing among the experimental conditions, we carried out a post-hoc analysis

based on multiple pairwise comparisons. The Tukey's honestly significant difference test was applied [22] and results are reported in Table II. In the table, columns denote all the pairs among experimental conditions: R denotes rest condition, while M, P and (M&P) denote the mental, physical and joint mental and physical fatigue condition, respectively. For each of them, we report whether a statistically significant difference was found, for each time-domain metrics, with significance level $\alpha = 0.05$ ($*$) or $\alpha = 0.01$ ($**$). Frequency-domain metrics were not considered since the ANOVA test returned that differences among conditions are not significant. Results in Table II show that rest condition can be distinguished from the three fatigue conditions, quite accurately. All the metrics returned statistically significant difference between rest and physical fatigue (R-P) and rest and joint mental and physical fatigue R-(M&P). Moreover, metrics *SDNN* returned noticeable statistically significant differences ($p < 0.01$) for all the pairs R-M, R-P and R-(M&P). Similarly, *RMSSD* proved quite effective to distinguish any fatigue condition from rest. Metrics *mean RR* returned statistically significant differences between rest and physical fatigue (R-P), and rest and joint mental and physical fatigue R-(M&P); no significant differences were found between rest and mental fatigue (R-M). Metrics *pNN50* proved more sensitive to characterize mental fatigue: specifically, statistically significant differences were observed when comparing mental to physical fatigue (M-P) and to mental and physical fatigue induced together M-(M&P). Notably, none of the considered metrics exhibited significant differences between the conditions of physical fatigue and joint mental and physical fatigue (P-(M&P), rightmost column in Table II). It can be suggested that the presence of physical fatigue induces greater variability in heart rate than mental fatigue, thus concealing this last when the two sources of fatigue are jointly present.

As a summary, from Table II it is possible to conclude that the selected time-domain metrics for HRV analysis are sensitive enough to detect the lack of any kind of fatigue and the presence of physical fatigue. Mental fatigue can be discriminated from rest condition, as shown also in previous works [12], [23]. Nevertheless, at least according to our experimental protocol, it is hard to detect mental fatigue in

the presence of additional sources of physical fatigue.

C. Empirical cumulative distribution functions

These considerations are corroborated by the analysis of the empirical cumulative distribution functions of RR series and HRV metrics. While ANOVA and multiple pairwise comparisons tests compare metrics means only, the empirical distribution functions provide a complete statistical characterization of samples of a random variable. Figure 2 reports the empirical distribution functions for the RR series (Fig. 2(a)) and all the considered HRV metrics (Fig. 2(b) to Fig. 2(h)). In all the plots, the $x-$axis collects the sorted measured values for the corresponding variable X, and the $y-$axis reports, for a given x, the probability $F_X(x)$ to observe values less or equal to x: $F_X(x) = P(X \leq x)$ [24]. In the figure, panel 2(a) collects the RR intervals measured for all the test participants in the four experimental conditions. The figure shows that in the rest condition it is much more likely to measure greater RR intervals than in the physical fatigue condition, with these two conditions being well separated. In other words, greater RR intervals are much more likely to be observed during rest (blue curve) than during exposure to physical fatigue (yellow). Moreover, also the joint mental and physical fatigue condition (red) is quite well separated from rest, whereas shorter RR intervals are likely to be observed either during physical fatigue or mental and physical fatigue than during mental fatigue or rest. The fact that the curves for physical and joint mental and physical fatigue lie close in Fig. 2(a) explains why none of the considered HRV metrics exhibit significant differences in the pairwise comparison P-(M&P), as reported in the rightmost column of Table II. Conversely, the condition relative to mental fatigue (red curve) lies in the middle, exhibiting moderate separation from the other conditions.

The other panels in the figure refer to the considered HRV metrics and, for each metrics, collect the values computed for each test participant. Hence, the curves are traced considering 30 (one per participant) samples, computed per metrics in each condition. In the case of *mean RR*, Fig. 2(b) shows that this metrics provides greatest separation for rest and physical fatigue, with the other two conditions in the middle. This confirms the results in the first row of Table II. As regards *SDNN* (Fig. 2(c)), the greatest difference is found between rest and joint mental and physical fatigue, while mental fatigue and physical fatigue result in similar probabilities to achieve similar values. Curves in Fig. 2(c) show a similar behavior for *RMSSD*, with rest being the most separated condition. A similar behavior is found for physical and joint and mental physical conditions, which cannot be discriminated at all by this metrics. The curve for mental fatigue lies in the middle, between rest and those for physical and joint mental and physical fatigue. As confirmed in Table II, significant differences can be found between rest and mental fatigue. Then, as regards *pNN50*, whose empirical cumulative functions are shown in Fig. 2(e), it exhibits a similar behavior to *RMSSD*, but greater distances can be found among the conditions. In particular, also for this metrics it is not possible

to find any differences between physical and joint and mental physical conditions: as discussed above, the effect of mental fatigue adding to physical fatigue is minor. However, these two condition are well separated from both rest and mental fatigue.

Finally, the remaining panels in figures from 2(f) to 2(h) show the empirical cumulative distribution functions for the frequency-domain metrics. Here it can be noted that all the four experimental conditions exhibit the same behavior, thus making it impossible to discriminate them. This is confirmed by the high p-values reported in Table I for these metrics.

IV. CONCLUSIONS

In this paper we considered the problem of detecting mental and physical fatigue in a transparent and unobtrusive manner. This was motivated by the need to assess operator's state and condition in industrial workplaces in order to promote human wellbeing. Ultimately, the idea is that of achieving an accurate operator's model that can be used as an input in collaborative scenarios, such as collaborative robotic cells, to tune the behavior of the system and make it adaptive with respect to operator's state. Within this broad objective, in this paper we focused on the detection of mental and physical fatigue conditions by means of physiological monitoring. A wearable device, namely a thoracic belt, was considered to this end, as it provides unobtrusive measurement of cardiac activity. HRV was then extracted and analysed by computing standard time- and frequency-domain metrics. An experimental protocol was designed such that test participants were exposed to mental and physical fatigue.

Results showed that time-domain metrics exhibited statistically significant differences among four conditions: namely rest, mental fatigue, physical fatigue and joint mental and physical fatigue. However, it was found that, in the presence of physical fatigue, the presence of additional mental fatigue could not be detected by any metrics. A detailed analysis of the empirical distribution functions showed, for each metrics, the conditions that exhibit the greatest differences and, hence, that can be distinguished more accurately.

As a limitation of the present work, it is noteworthy that it represents a preliminary study, where physical and mental fatigue have been induced in a laboratory setting. Future works will consist in considering more realistic stressors that are typically occurring in factory shop floors. Moreover, to refine the detection of fatigue, multimodal physiological monitoring will be considered as well, merging information recorded with other wearable devices, such as electrodermal or muscular activity. Similarly, information related to the current task could be used, if available, to enhance accuracy of fatigue detection.

REFERENCES

[1] E. Commission, D.-G. for Research, Innovation, M. Breque, L. De Nul, and A. Petridis, *Industry 5.0 : towards a sustainable, human-centric and resilient European industry*. Publications Office, 2021.
[2] V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications," *Mechatronics*, vol. 55, pp. 248–266, 2018.
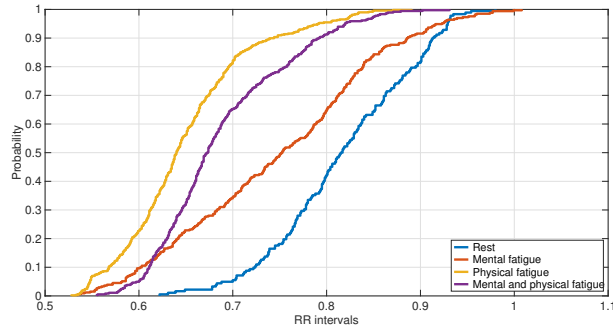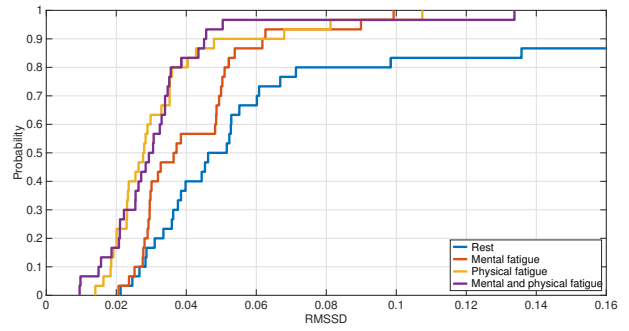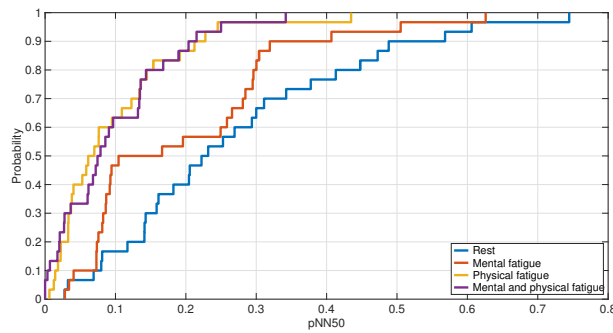
(a) RR intervals series.
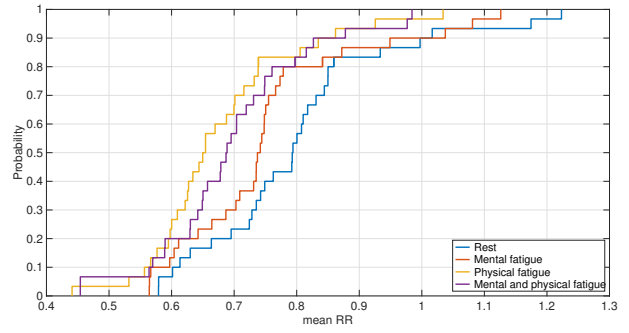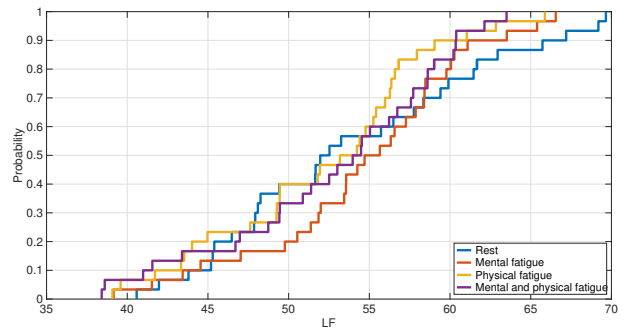
(b) HRV time-domain metrics *mean RR*.
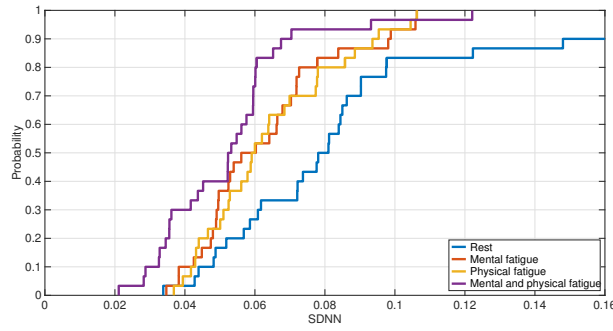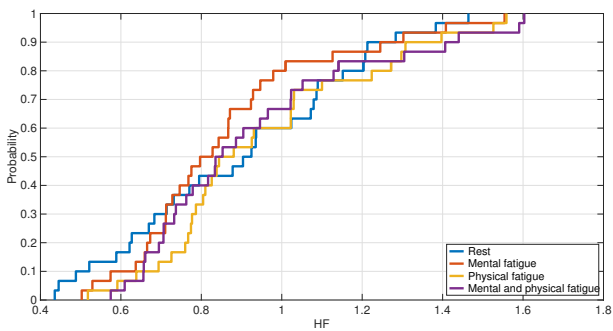
(c) HRV time-domain metrics *SDNN*.

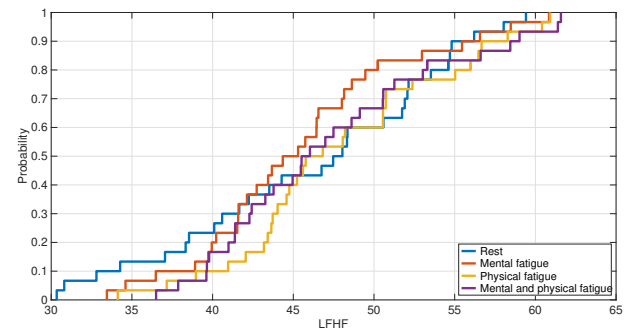(d) HRV time-domain metrics *RMSSD*.

(e) HRV time-domain metrics *pNN50*.

(f) HRV frequency-domain metrics *LF*.

(g) HRV frequency-domain metrics *HF*.

(h) HRV frequency-domain metrics *LFHFratio*.

Fig. 2.   Empirical cumulative functions for RR intervals series and HRV metrics, in the four experimental conditions.

[3] A. Bettoni, E. Montini, M. Righi, V. Villani, R. Tsvetanov, S. Borgia, C. Secchi, and E. Carpanzano, "Mutualistic and adaptive human-machine collaboration based on machine learning in an injection moulding manufacturing line," *Procedia CIRP*, vol. 93, pp. 395–400, 2020.

[4] A. Pupa, W. Van Dijk, and C. Secchi, "A human-centered dynamic scheduling architecture for collaborative application," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4736–4743, 2021.

[5] C. T. Landi, V. Villani, F. Ferraguti, L. Sabattini, C. Secchi, and C. Fantuzzi, "Relieving operators' workload: Towards affective robotics in industrial scenarios," *Mechatronics*, vol. 54, pp. 144–154, 2018.

[6] V. Villani, L. Sabattini, F. Loch, B. Vogel-Heuser, and C. Fantuzzi, "A general methodology for adapting industrial hmis to human operators," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 1, pp. 164–175, 2019.

[7] Z. S. Maman, Y.-J. Chen, A. Baghdadi, S. Lombardo, L. A. Cavuoto, and F. M. Megahed, "A data analytic framework for physical fatigue management using wearable sensors," *Expert Systems with Applications*, vol. 155, p. 113405, 2020.

[8] S. Gillinov, M. Etiwy, R. Wang, G. Blackburn, D. Phelan, A. M. Gillinov, P. Houghtaling, H. Javadikasgari, and M. Y. Desai, "Variable accuracy of wearable heart rate monitors during aerobic exercise," *Med Sci Sports Exerc*, vol. 49, no. 8, pp. 1697–1703, 2017.

[9] I. Hettiarachchi, S. Hanoun, D. Nahavandi, J. Iskander, M. Hossny, and S. Nahavandi, "Towards More Accessible Physiological Data for Assessment of Cognitive Load - A Validation Study," in *Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018*. IEEE, oct 2019, pp. 3045–3050.

[10] V. Villani, L. Sabattini, C. Secchi, and C. Fantuzzi, "A framework for affect-based natural human-robot interaction," in *27th IEEE Int. Symp. Robot and Human Interactive Communication (RO-MAN)*, IEEE, Ed., 2018.

[11] V. Mishra, G. Pope, S. Lord, S. Lewia, B. Lowens, K. Caine, S. Sen, R. Halter, and D. Kotz, "The Case for a Commodity Hardware Solution for Stress Detection," in *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers - UbiComp '18*. ACM Press, 2018, pp. 1717–1728.

[12] V. Villani, M. Righi, L. Sabattini, and C. Secchi, "Wearable devices for the assessment of cognitive effort for human–robot interaction," *IEEE Sensors Journal*, vol. 20, no. 21, pp. 13 047–13 056, 2020.

[13] R. Castaldo, P. Melillo, U. Bracale, M. Caserta, M. Triassi, and L. Pecchia, "Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis," *Biomedical Signal Process. Control*, vol. 18, pp. 370–377, 2015.

[14] P. Schmidt, A. Reiss, R. Dürichen, and K. V. Laerhoven, "Wearable-based affect recognition - a review," *Sensors*, vol. 19, no. 19, p. 4079, 2019.

[15] J. R. Stroop, "Studies of interference in serial verbal reactions," *Journal of Experimental Psychology*, vol. 18, no. 6, pp. 643–662, 1935.

[16] W. K. Kirchner, "Age differences in short-term retention of rapidly changing information," *Journal of Experimental Psychology*, 1958.

[17] D. J. Plews, B. Scott, M. Altini, M. Wood, A. E. Kilding, and P. B. Laursen, "Comparison of heart-rate-variability recording with smartphone photoplethysmography, polar H7 chest strap, and electrocardiography," *International Journal of Sports Physiology and Performance*, vol. 12, no. 10, pp. 1324–1328, nov 2017.

[18] N. Dahlstrom and S. Nahlinder, "A comparison of two recorders for obtaining in-flight heart rate data," *Applied psychophysiology and biofeedback*, vol. 31, no. 3, pp. 273–279, 2006.

[19] Task Force of The European Society of Cardiology & The North American Society of Pacing and Electrophysiology, "Heart rate variability - Standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, pp. 354–381, 1996.

[20] G. D. Clifford, F. Azuaje, and P. McSharry, Eds., *Advanced Methods and Tools for ECG Data Analysis*. Artech House, Inc., 2006.

[21] G. D. Clifford and L. Tarassenko, "Quantifying errors in spectral estimates of HRV due to beat replacement and resampling," *IEEE Trans. Biomedical Engineering*, vol. 52, no. 4, pp. 630–638, 2005.

[22] R. G. Miller, *Simultaneous statistical inference*. Springer-Verlag New York Inc., 1981.

[23] I. Hettiarachchi, S. Hanoun, D. Nahavandi, J. Iskander, M. Hossny, and S. Nahavandi, "Towards More Accessible Physiological Data for Assessment of Cognitive Load - A Validation Study," in *Proceedings - 2018 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2018*. IEEE, oct 2019, pp. 3045–3050.

[24] H. Kobayashi, B. L. Mark, and W. Turin, *Probability, random processes, and statistical analysis: applications to communications, signal processing, queueing theory and mathematical finance*. Cambridge University Press, 2011.