

Article

Assessment of Implicit and Explicit Measures of Mental Workload in Working Situations: Implications for Industry 4.0

Michele Mingardi ^{1,†}, Patrik Pluchino ^{1,2,†} , Davide Bacchin ^{1,2} , Chiara Rossato ¹  and Luciano Gamberini ^{1,2,*}

¹ Human Inspired Technology Research Centre, University of Padova, 35121 Padova, Italy; michele.mingardi@studenti.unipd.it (M.M.); patrik.pluchino@unipd.it (P.P.); davide.bacchin.2@phd.unipd.it (D.B.); chiara.rossato@unipd.it (C.R.)

² Department of General Psychology, University of Padova, 35131 Padova, Italy

* Correspondence: luciano.gamberini@unipd.it; Tel.: +39-049-8276605

† The authors equally contributed to this work.

Received: 31 July 2020; Accepted: 11 September 2020; Published: 15 September 2020



Abstract: Nowadays, in the context of Industry 4.0, advanced working environments aim at achieving a high degree of human–machine collaboration. This phenomenon occurs, on the one hand, through the correct interpretation of operators’ data by machines that can adapt their functioning to support workers, and on the other hand, by ensuring the transparency of the actions of the system itself. This study used an ad hoc system that allowed the co-registration of a set of participants’ implicit and explicit (I/E) data in two experimental conditions that varied in the level of mental workload (MWL). Findings showed that the majority of the considered I/E measures were able to discriminate the different task-related mental demands and some implicit measures were capable of predicting task performance in both tasks. Moreover, self-reported measures showed that participants were aware of such differences in MWL. Finally, the paradigm’s ecology highlights that task and environmental features may affect the reliability of the various I/E measures. Thus, these factors have to be considered in the design and development of advanced adaptive systems within the industrial context.

Keywords: implicit/explicit measures; human–machine interaction; symbiotic system; mental workload; eye tracking; heart rate; electrodermal activity; NASA-TLX; working tasks

1. Introduction

The widespread distribution of cutting-edge software and hardware products in the workplace requires more and more individuals to interact with technological systems. Most of these tools are designed and developed to ease the accomplishment of tasks (e.g., avoiding information pollution). They support workers both at the physical and cognitive level, based on the paradigm of the internet of things (IoT), which empowers the fast and flexible information exchange among different technological devices through their digital connection. The IoT concept, applied to the industrial environment, especially in the context of Industry 4.0, results in an interconnected system of physical entities characterized by communication exchanges, data processing, and/or sensing capabilities [1]. The goal of IoT exploitation is to produce a wide range of work-related improvements, such as increased work safety and security, productivity, and efficiency of resource management.

Humans are often part of such systems, and whenever the interaction between humans and machines involves a form of mutual support, it could be defined as symbiotic. Licklider first coined the term man–computer symbiosis to describe this partnership characterized by cooperative interaction

between the two coupled agents, as happens in nature between two living organisms that draw mutual support [2]. When these conditions are met, human beings are able to receive more customized support in performing tasks, and technological devices improve their functioning due to continuous human–computer interactions (HCI) [3].

In this perspective, the standard human–computer communication is explicit and observable, and users can intentionally communicate and act within the system, for example, through a graphical interface. At present, several forms of implicit interaction are spreading, in terms of non-voluntarily types of communication from the human agent towards the computer, to improve HCI [4,5]. This type of interaction is characterized by the machine’s acquisition of users’ data that cannot be controlled or manipulated by them. For instance, bio-signals are usually considered to reflect the users’ psychophysiological state [5,6] and could be acquired by utilizing non-invasive sensors. In symbiotic systems, these users’ collected data could be shared among all the computational devices connected through the network. Once it detects the operator’s need for support (e.g., an extreme level of stress), the system could tailor its interface and functions by implementing real-time adaptive features [7,8]. These concepts were modeled by Gamberini and Spagnolli ([3]; see Figure 1). They postulated that a symbiotic system collects meaningful data from both the environment and the human, employing a series of devices such as biosensors (data acquisition). These data are elaborated and interpreted by the computational tools, exploiting artificial intelligence or machine learning algorithms, resulting in a reliable and flexible user’s model (data elaboration and interpretation). Lastly, the machine personalizes its actions and decisions on the actual user’s needs and desires as they are programmed (execution of automated actions sequence). This sequence of events occurs iteratively. Thus, the symbiotic systems would be more precise in interpreting the users’ intentions and behaviours and supporting them in achieving their goals. As a consequence, the performance of the human agent and overall proficiency would improve.

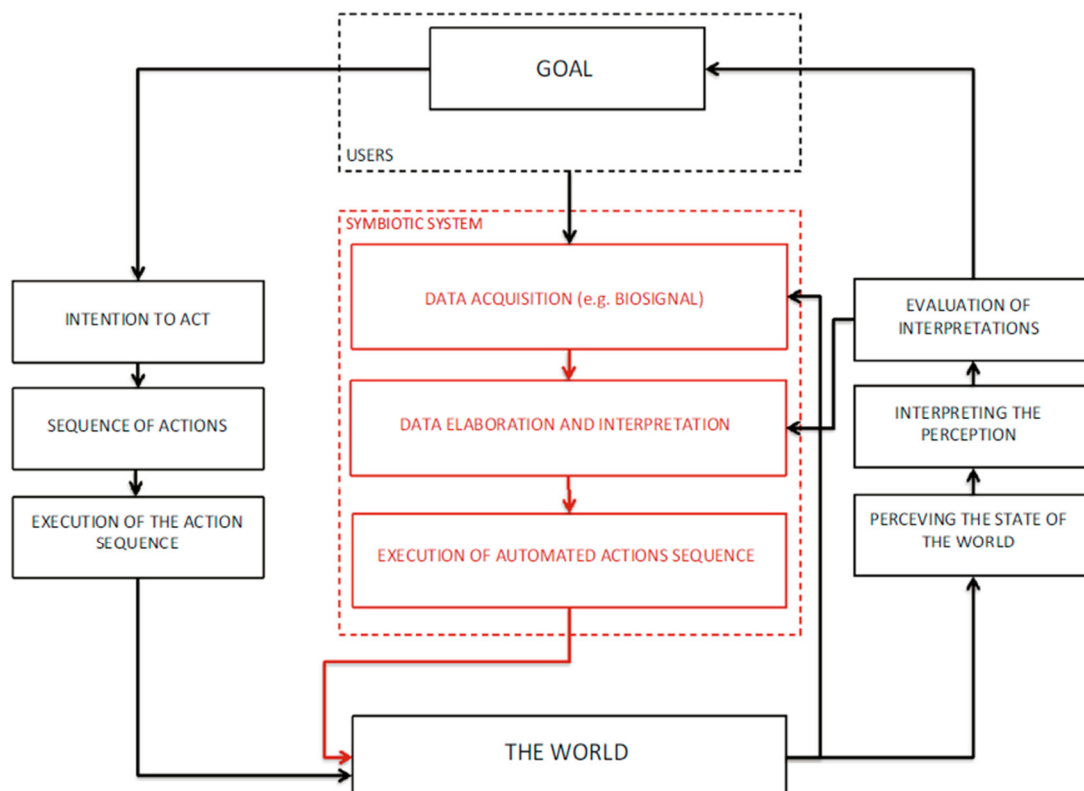


Figure 1. Model of a symbiotic system.

Symbiotic systems present two crucial limits: the interpretability of the users' data and the transparency of the system's actions.

Regarding the former, it must be considered that a monitored measure (e.g., heart rate; HR) has no univocal interpretation. The behavioral and psychophysiological measures may reflect different internal states related to the specific task that the user is accomplishing. For instance, the electrodermal activity or visual behaviour can reveal different internal states, such as variations in the level of stress [9,10], affective state (i.e., arousal and valence) [11], or mental workload [12,13]. A reductionist approach could lead to the misunderstanding of data, causing a failure to support users when needed, or implementing adaptive features when they are not necessary. To overcome this problem, the combination of various sources of information could be the most suitable solution. It is possible to pair users' implicit and explicit (I/E) data to achieve the richest level of comprehension [5].

Regarding the second limit, namely the system transparency, it is relevant to clarify to the users the reasons and the modalities of the machine's actions that modify the "state" of the world. This characteristic allows the users to carefully decide which data to provide to the devices and how to use the information coming from the devices [14]. Whenever this transparency is lacking, the system becomes a black box, and the user is not capable of understanding the causes of modifications in the system's functioning. To test this construct, we could compare if the users' perceptions (i.e., self-reported data acquired offline) match the detected changes in the I/E data.

The exploitation of these user's personal and sensitive data raises a series of issues related to ethics, privacy violations, and misuse of the processed data, which are intrinsically connected to the symbiotic system's main advantage, namely, its capability to act bypassing the user [14,15].

To consider the critical aspects of the symbiotic system mentioned above, we needed to identify a reliable construct that could be assessed based on implicit, explicit, and subjective measures [16]. Within the work context [8], the impact of mental workload (MWL) on performance has been analyzed [17–19], especially regarding its optimization for users operating in the environment of so-called Industry 4.0 [20].

In the ergonomics domain, MWL is defined as the mental demands imposed on workers by accomplishing tasks, and refers to the operators' cognitive resources needed to meet the working requests [21]. One of the most interesting features of the MWL is that it can be exploited to evaluate how users' responses are affected by the difficulty of a task [17].

Recent literature reviews have reported a series of implicit measures of MWL that are affected by variations in the imposed level of mental workload [12,22], such as the heart rate (HR), which increases with higher task demands [23] and multitasking conditions [24]. Another example is the skin conductance (SC, or electrodermal activity; EDA) which has been shown to increase in different situations such as complex driving environments [21] and the presence of secondary tasks [18,25]. Considering eye-related metrics, modulations due to a high level of mental workload are shown in pupil diameter (i.e., increment [20]), fixation duration and frequency (i.e., decrement [25]), saccadic frequency (decrement [26]), blink duration (decrement [22]) and frequency (increment [27,28]), and the nearest neighbor index (NNI) [13,29,30]. MWL is generally assessed explicitly considering the user's performance, such as time on task and accuracy [31].

Concerning the self-reported measures of MWL, one of the commonly employed tools is the NASA task load index (NASA-TLX) questionnaire [32]. This comprises six sub-scales, eventually providing a single score. Some studies utilize the overall scoring (i.e., computed considering all the sub-scales), the so-called task load index. Recently, Galy and colleagues [33] have pointed out the importance of evaluating each dimension of the NASA-TLX questionnaire independently (e.g., mental demand, physical demand, effort, etc.). By doing so, it is possible to assess how the single sub-scale scores may be differently affected by the mental workload related to a specific task. In experiments in which mental demand is increased, participants reported higher perceived mental demand, effort, and frustration, and lower subjective perception of their task performance [34–36].

By integrating these multiple sources of information (i.e., I/E), we aim to comprehend how to improve context-aware interaction and lay the foundations for the future design and development of a fully integrated symbiotic system.

The Study

The present research could be considered an initial phase of the design and development of a symbiotic system that can detect an increase in the user's MWL and, on this basis, adapt its operating to reduce the task demand. As the first step toward such purpose, this study's main objective was to define a set of implicit and explicit measures that could be exploited to monitor variations of the MWL experienced by the users due to the manipulation of the task difficulty. A second purpose was to investigate if participants were aware of the different levels of mental load. The assignment was carried out in an ad hoc built ecological setup, similar to a real assembly line, to increase the outcomes' potential generalizability. Indeed, this ad hoc system (described in Section 2.3, Equipment and Material) allowed us to simultaneously observe, and subsequently compare, all the measures under investigation in the present study.

Based on the two main limits of symbiotic systems (i.e., data interpretability and system transparency), the study explored the following research questions:

RQ1. Do the implicit and explicit measures reflect the variation in the MWL due to the different experimental tasks in the same fashion?

RQ2. Do the implicit and explicit measures reflect the subjective perceptions of participants regarding the MWL imposed by the different tasks?

For research question 1, we expected that both implicit (i.e., ocular, cardiac, and electrodermal activity), and explicit (i.e., task accuracy) measures would be affected, in a highly ecological paradigm, by different levels of MWL elicited through the manipulation of task difficulty.

For research question 2, we predicted that the findings related to the implicit and explicit measures (e.g., eye tracking and task accuracy) would match participants' subjective perception (e.g., NASA-TLX scores) in reflecting the different levels of mental workload.

2. Materials and Methods

2.1. Participants

Thirty individuals ($F = 16$) with a mean age of 32.03 years ($SD = 12.02$) took part in the experiment. All participants presented normal or corrected-to-normal vision. None of the participants had heart conditions. We did not give any kind of compensation for partaking in the trial. Indeed, participation was voluntary.

2.2. Design

The experiment followed a within-participants design. An independent variable, the task difficulty, was manipulated: single- (screwing) vs. dual-task (screwing and backward counting). The order of the tasks was counterbalanced across the participants.

2.3. Equipment and Materials

The workstation was designed and developed within the present study to consider the highest possible number of implicit and explicit user data. It was composed of a perforated blackboard (Figure 2a,b), an eye-tracking pair of glasses (Pupil Labs, Figure 2d), and an amplifier for the monitoring of psychophysiological signals (ProCompt system, Figure 2c).

The handmade blackboard consisted of a hollow parallelepiped with a square base ($60 \times 60 \times 50$ cm) and four telescopic supports, to allow the adjustment of its relative height. Two of the four vertical sides had 50 holes arranged in 5 parallel horizontal rows, with threaded inserts in each hole, able to hold hexagonal-head steel bolts (8×25 mm). The entire system could register every voluntary

human-workstation interaction (duration and frequency), from when the participant collected a bolt to when s/he completely tightened it on the blackboard. Indeed, a Makey Makey board (Makey Makey LLC ©; JoyLabz, Santa Cruz, CA, USA), installed inside the blackboard and connected to a laptop, allowed this information to be sent to a portable computer. These data were recorded using E-Prime software (v.2.0; Psychology Software Tools ©; Pittsburgh, PA, USA). The participants, the board, and a metal drawer containing the screws were part of an electrical circuit. A bracelet worn by participants was an important part of the system. When s/he touched one of the circuit components, the ensuing electrical contact sent a trigger to the connected computer to record the information of the interaction which had just occurred. Two GoPro Hero 4 cameras, one mounted in front of the participant (i.e., over the blackboard; see Figure 2b) and the other on a lateral position, allowed the continuous monitoring of the participants. The GoPro and the eye-tracker world-camera sampling frequency was set to 30 frames per second to permit an offline synchronization of the video-recordings.

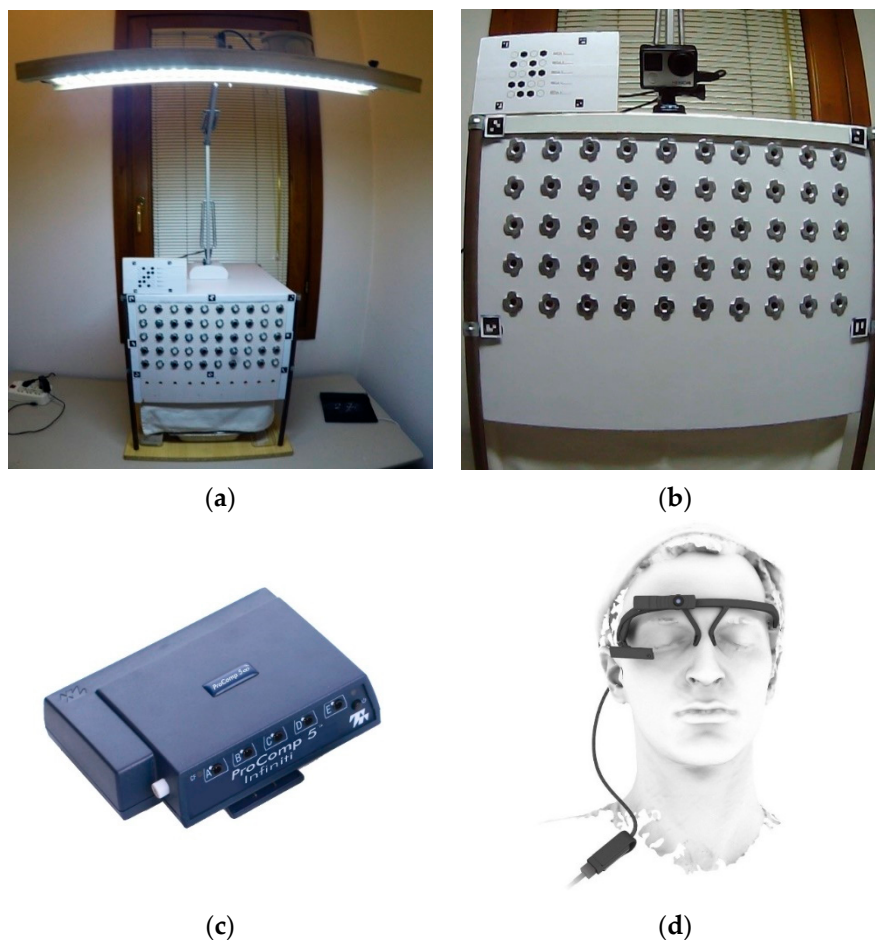


Figure 2. Components of the ad hoc workstation. (a) Blackboard distant view; (b) blackboard front view; (c) ProComp5 Infiniti; (d) Pupil Labs glasses.

A cinematographic clapper was utilized for synchronizing the measures collected from the human–workstation interactions and the implicit measures. This tool was also interfaced, through the Makey Makey board, to the E-Prime software. This clapper created a detectable electrical contact, which was then synchronized with the frame corresponding to the closure. Therefore, the closing of this instrument was carried out in the cameras' visual field to synchronize the video streams and other devices. Finally, the workstation was equipped with a lighting system, consisting of six strips of SMD 5050 LEDs, emitting a total of 1100 lumens of cold white light (4500 K), positioned on an

adjustable support, anchored to the upper base of the blackboard, to maintain the same level of lighting in the laboratory.

We utilized a pair of binocular eye-tracking glasses (i.e., Pupil Labs; Pupil Labs GmbH ©, Berlin, DE; Figure 2d) in the experiment. This tool is capable of gathering data on fixations, blinks, saccades (duration and frequency), and pupil diameter. It was connected to an MSI laptop (Intel Core i7-6700HQ, screen resolution 1920 × 1080). The software Pupil Capture (Pupil Labs GmbH ©, Berlin, DE) allowed the system calibration and the data recording (sampling frequency: 120 Hz). The software Pupil Player (Pupil Labs GmbH ©, Berlin, DE) was utilized to export the eye-tracking data.

Lastly, a Procomp5 Infiniti amplifier (Thought Technology Ltd.; Montréal, QC; Figure 2c) and its software (i.e., BioGraph Infiniti) permitted the recording of psychophysiological data. The software ran on a second MSI laptop (in the same configuration as the aforementioned laptop). The heart rate (HR) was acquired using a blood volume pulsimeter positioned on the left middle finger. The electrodermal activity (EDA) was collected using two electrodes placed on the annular and index fingers of the same hand. The hand was placed still on the table and was not involved in any task.

2.4. Tasks

The main task was conceived ecologically, using the blackboard described in the previous section to perform a real manual screwing task. In the single-task, the participants collected 50 bolts of two different colours (25 white and 25 black), one at a time. They placed them according to a specific sequence in the perforated blackboard (Figure 2a,b). The sequence was depicted on a small structure located above the blackboard's top left (see Figure 2b, upper-left; Figure 3).

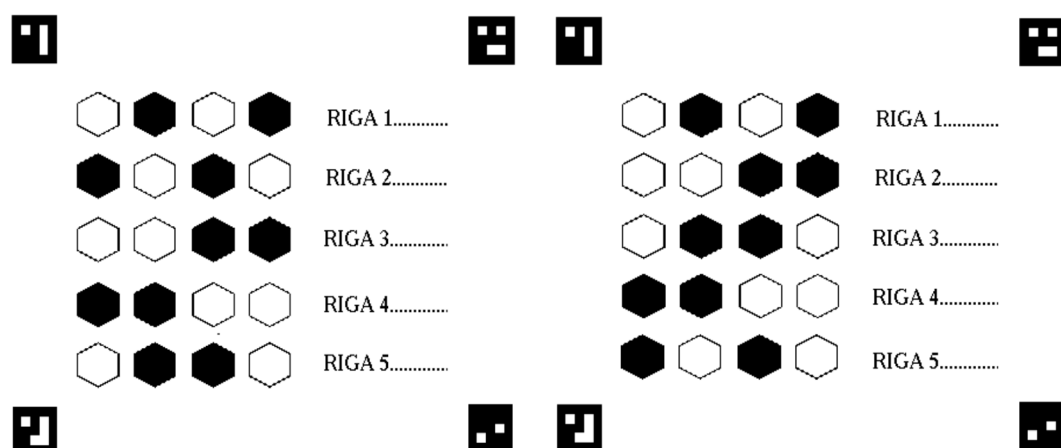


Figure 3. Single-task model (left); dual-task model (right). The black and white small pictures that surround both models are eye-tracker markers.

The models comprised five sequences (one for each row) of four hexagons (i.e., the “heads” of the bolts). Considering the primary task (i.e., screwing), in both experimental sessions, participants had to fill in the blackboard from left to right (each row) and from up to down following the model's sequences (Figure 3). Participants picked the bolts from a drawer, covered by a small curtain (Figure 2a) placed below the blackboard.

In the dual-task, users had to also deal with an interference task (i.e., backward counting). It consisted of continuously subtracting 7, starting from 1000, up to the completion of the main task (i.e., screwing). Participants were asked to be accurate and fast as much as possible while accomplishing both the experimental tasks. The backward counting task was chosen as an interfering task because to be performed it involves multiple cognitive resources (e.g., executive functions, attention, short- and long-term memory, and procedural and working memory) and this increases the overall general task demands. Furthermore, it does not explicitly require visual processing.

2.5. Procedure

The experiment was performed in a quiet and isolated room, comprising different phases. In the beginning, the participants were administered an informative note and gave informed consent. Then, they carried out two preliminary tests. A backward counting test was conducted, in which participants, starting from 900, had to subtract 7 twenty times to complete the task. This task served as a familiarization with the interference task. Then, participants had to accomplish a Corsi test that evaluated their visuospatial memory. The Corsi test allowed the eventual exclusion of participants with reduced visuospatial memory (lower than 3). At this stage, the experimenters administered a questionnaire concerning demographic information (e.g., gender, education, occupation) and familiarity with ordinary manual working tasks. Then, participants were helped by the researchers to wear all the sensors (see Equipment and Materials section for more details). The experimenters carried out the eye-tracker calibration. An audio file containing the instructions of the first task (e.g., single-task, low difficulty) was presented. If participants were unsure of something regarding the task, they could ask the experimenters. Then, a 5-min baseline of psychophysiological measurements was performed. After the completion of the first task, participants filled out the NASA-TLX questionnaire. The researchers double-checked all the sensors and devices in preparation for the second task. A second audio file with the instructions for the second task (e.g., dual-task, high difficulty) was presented. Participants had to perform the screwing task and, simultaneously, the backward counting task (i.e., interference). The dual-task ended when the screwing task was completed. Both experimental sessions were video-recorded.

The whole experiment lasted about one hour.

2.6. Measures

We consider the following implicit, explicit, and subjective metrics.

Explicit

- Task accuracy, the number of errors in the screwing task (mean, s), e.g., bolts placed into the wrong position or dropped.

Implicit:

- Time on task, the time needed to accomplish the screwing task (mean, s);
- Fixation duration (mean, ms) and frequency (per min);
- Nearest neighbor index (NNI, mean), namely, the dispersion of fixations in the visual field;
- Blink duration (mean, ms) and frequency (per min);
- Saccade duration (mean, ms) and frequency (per min);
- Pupil diameter (mean, mm);
- Heart rate (mean, beats per min);
- Electrodermal activity (mean, micro Siemens, μS).

Subjective:

- Overall task load (NASA-TLX questionnaire); the six-subscales were considered separately.

3. Results

Data were analyzed utilizing RStudio software [37]. We carried out non-parametric tests in the case of data that were not normally distributed. Whenever needed, the p -values were adjusted applying the Benjamini-Hochberg (BH) correction for multiple comparisons [38]. Considering the Wilcoxon tests, the reported values of N were equal to the pairs of measures that showed differences.

3.1. Task Accuracy

A Wilcoxon test was performed considering the accuracy of the main task (i.e., screwing), as a function of task difficulty. An effect emerged ($N = 28$, $W = 333$, $p < 0.01$, $r = 0.54$). Participants made more mistakes in the dual-task ($M = 4.3$, $SD = 2.5$) compared to the single-task ($M = 2.7$, $SD = 2.7$; see Figure 4).

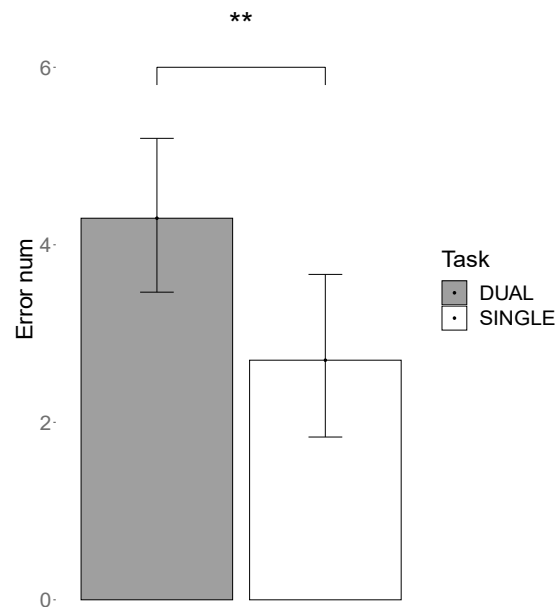


Figure 4. Mean number of errors as a function of task difficulty. ** $p < 0.01$.

3.2. Time on Task

A t -test was conducted concerning the time needed to accomplish the tasks as a function of task difficulty. An effect was shown ($t(29) = 7.391$, $p < 0.001$, $r = 0.80$). Participants needed more time for carrying out the dual-task ($M = 873.36$ s, $SD = 128.34$) compared to the single-task ($M = 692.97$ s, $SD = 112.31$; see Figure 5).

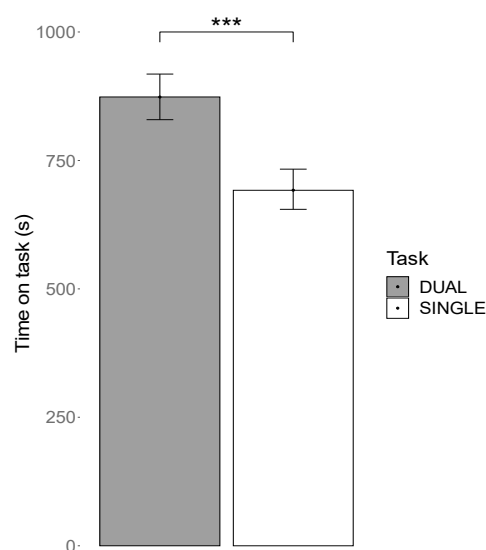


Figure 5. Mean time on task (s) as a function of task difficulty. *** $p < 0.001$.

3.3. Fixation Duration and Frequency

A paired t -test was carried out considering the mean duration of fixation as a function of task difficulty. A difference emerged ($t(29) = -3.992$, $p < 0.001$, $r = 0.60$). A shorter mean duration was shown in dual-task ($M = 431.1$ ms, $SD = 128.34$) compared to the single-task ($M = 488.4$ ms, $SD = 120.56$; see Figure 6a).

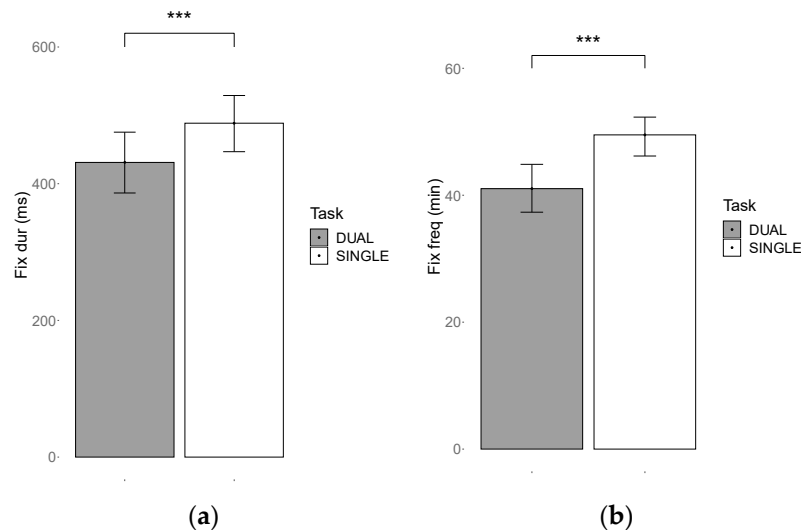


Figure 6. (a) Mean fixation duration and (b) frequency as a function of task difficulty. *** $p < 0.001$.

A Wilcoxon test was performed regarding the mean frequency of fixation. A difference emerged ($N = 30$, $W = 11$, $p < 0.001$, $r = 0.97$). In the single-task ($M = 49.5$, $SD = 9.03$), participants made, on average, more fixations compared to the dual-task ($M = 41.04$, $SD = 11.27$; see Figure 6b).

3.4. Blink Duration and Frequency

A Wilcoxon test was conducted regarding the average duration of blinks, and a difference emerged ($N = 30$, $W = 348$, $p < 0.05$, $r = 0.44$). The average blink length was higher in the dual-task ($M = 215.26$ ms, $SD = 23.22$) than in the single-task ($M = 208.1$ ms, $SD = 17.12$; see Figure 7a). A Wilcoxon test was performed, considering the mean frequency of blinking, and a difference emerged ($N = 30$, $W = 387$, $p < 0.001$, $r = 0.60$). The average number of blinks was higher in the dual-task ($M = 11.11$, $SD = 11.93$) than in the single-task ($M = 7.75$, $SD = 8.5$; see Figure 7b).

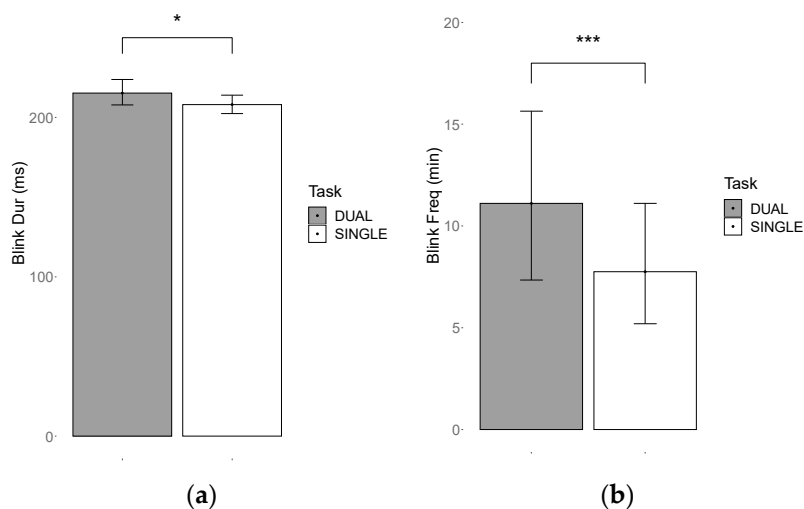


Figure 7. (a) Mean blink duration and (b) frequency as a function of task difficulty. * $p < 0.05$. *** $p < 0.001$.

3.5. Saccade Frequency

Concerning the average frequency of saccades, a difference was shown ($N = 30$, $W = 60$, $p < 0.001$, $r = 0.69$). The mean number of saccades was lower in the dual-task ($M = 19.67$, $SD = 5.98$) compared to the single-task ($M = 23.28$, $SD = 5.14$; see Figure 8).

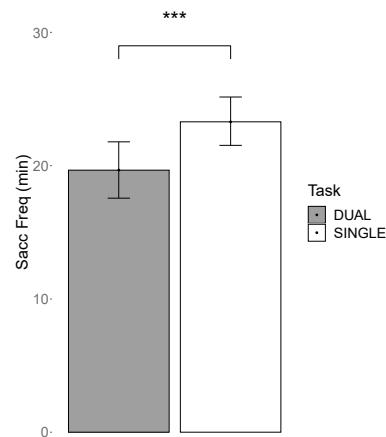


Figure 8. Mean saccade frequency as a function of task difficulty. *** $p < 0.001$.

3.6. Heart Rate

The analysis of the mean heart rate, as a function of task difficulty, showed a difference ($N = 30$, $W = 104$, $p < 0.01$, $r = 0.48$). Accomplishing the dual-task increased the heart rate of participants ($M = 91.83$, $SD = 15.06$) compared to the single-task ($M = 88.57$, $SD = 14.83$; see Figure 9).

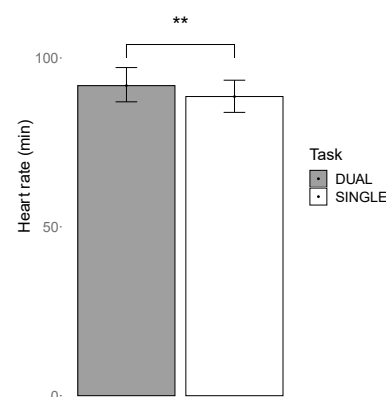


Figure 9. Mean heart rate as a function of task difficulty. ** $p < 0.01$.

3.7. EDA

A Wilcoxon test was carried out concerning the mean electrodermal activity. No differences emerged as a function of task difficulty ($N = 30$, $W = 198$, $p > 0.05$). Furthermore, the mean EDA was evaluated employing a Friedman test as a function of time ($X^2(3) = 74.64$, $p < 0.001$; see Figure 10). We then considered in the following analyses the values obtained at the baseline ($M = 2.07$, $SD = 2.28$), during the first task ($M = 3.35$, $SD = 2.81$), the resting phase ($M = 3.41$, $SD = 2.85$), and the second task ($M = 4.03$, $SD = 3.03$). The series of Wilcoxon post hoc tests (BH correction; [38]) showed that in comparison to the baseline all the other experimental phases showed higher levels of EDA (respectively, first task, $N = 30$, $W = 0$, $p < 0.001$, $r = 0.87$; resting phase, $N = 30$, $W = 0$, $p < 0.001$, $r = 0.87$; and second task, $N = 30$, $W = 0$, $p < 0.001$, $r = 1.1$). Furthermore, the mean EDA associated with the second task was higher than both the results related to the first task ($N = 30$, $W = 11$, $p < 0.001$, $r = 0.83$) and the one of the resting phase ($N = 30$, $W = 3$, $p < 0.001$, $r = 1$). No differences emerged between the first task and the resting phase ($N = 30$, $W = 280$, $p > 0.05$).

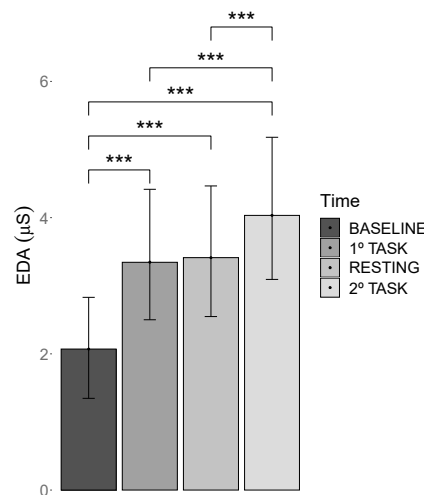


Figure 10. Mean electrodermal activity (EDA) as a function of time. *** $p < 0.001$.

3.8. NASA-TLX

A series of Wilcoxon tests, considering four out of a total of six sub-scales (i.e., mental demand, performance, effort, frustration), highlighted differences as a function of task difficulty. With regard to the perception of mental demand ($N = 27$, $W = 378$, $p < 0.001$, $r = 0.83$), participants considered the dual-task more challenging ($M = 74.5$, $SD = 16.6$) than the single-task ($M = 37.5$, $SD = 19.2$; see Figure 11a).

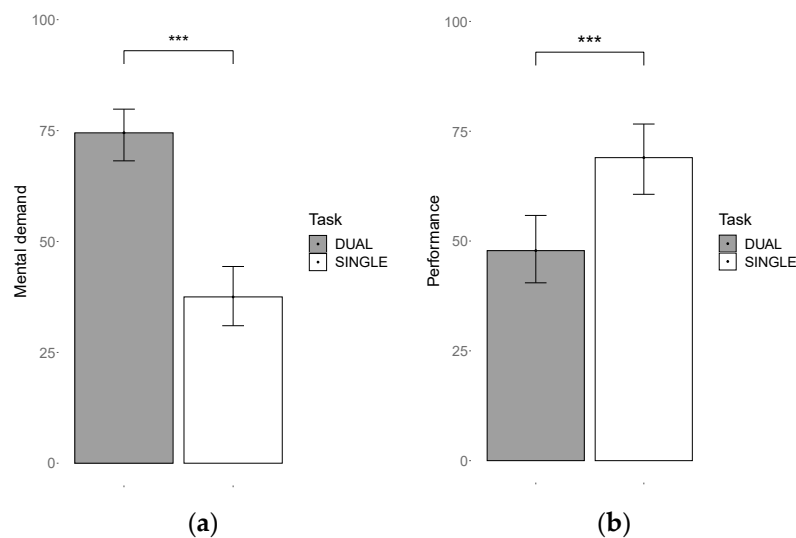


Figure 11. (a) Mean perceived mental demand and (b) perceived performance as a function of task difficulty. *** $p < 0.001$.

Furthermore, participants reported a higher perceived performance ($N = 25$, $W = 36$, $p < 0.001$, $r = 0.62$) in the single- ($M = 69$, $SD = 21.67$) than in the dual-task ($M = 47.83$, $SD = 21.06$; see Figure 11b).

In addition, the perceived effort showed a difference due to task difficulty ($N = 27$, $W = 350$, $p < 0.001$; $r = 0.71$). Participants perceived a higher effort in the dual- ($M = 72.83$, $SD = 18.83$) than in the single-task ($M = 48.33$, $SD = 21.67$; see Figure 12a).

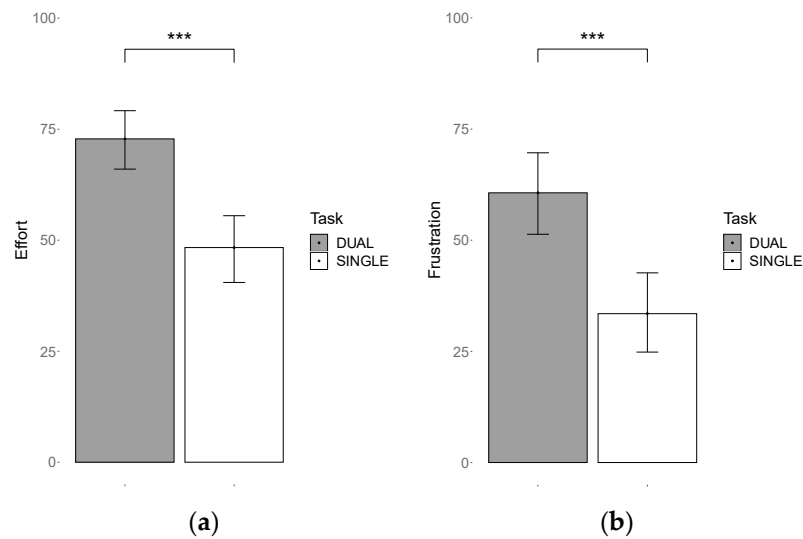


Figure 12. (a) Mean perceived effort and (b) perceived frustration as a function of task difficulty.

*** $p < 0.001$.

Finally, participants experienced a different level of frustration considering the difficulty of the tasks ($N = 26$, $W = 322.5$, $p < 0.001$; $r = 0.68$). Participants assigned a higher level of frustration to the dual-task ($M = 60.67$, $SD = 27.16$) than to the single-task ($M = 33.5$, $SD = 25.90$; see Figure 12b).

3.9. Measures with Non-Significant Differences

Detailed pieces of information regarding the other dependent variables, which did not show significant differences as a function of task difficulty, are reported in Table 1.

Table 1. Means and standard deviations of each variable that did not show a significant difference between the two conditions, the corresponding statistical tests, and p -values. Note: ST = single-task; DT = dual-task.

Variables	ST M (SD)	DT M (SD)	Tests	p
NNI	0.82 (0.058)	0.79 (0.063)	$t(29) = -2.07$	0.047
Saccade Duration (ms)	110.92 (18.41)	111.51 (26.74)	$N = 30$, $W = 220$	0.81
Pupil Diameter (mm)	4.26 (0.75)	4.32 (0.77)	$t(29) = 1.58$	0.13
EDA (mS)	3.71 (2.82)	3.67 (3.06)	$N = 30$, $W = 198$	0.48
Physical Demand (NASA-TLX)	48.33 (25.91)	41.17 (24.62)	$N = 20$, $W = 65$	0.14
Temporal Demand (NASA-TLX)	41.33 (21.13)	49.67 (24.28)	$N = 24$, $W = 213.5$	0.071

3.10. Multiple Linear Regressions

A series of multiple linear regressions was carried out to test whether the I/E measures could predict the scores obtained in the scales of the NASA-TLX that showed differences as a function of task difficulty (i.e., mental demand, performance, effort, and frustration). The I/E measures that discriminated the two levels of task difficulty (i.e., task accuracy, time on task, fixation duration, fixation frequency, blink duration, blink frequency, saccade frequency, and heart rate), the task, and all two-way interactions (i.e., I/E measures–task) were inserted as predictors in the regression models. The outcomes of the regression model with the mental demand as the dependent variable indicated that there was a significant collective effect between the considered predictors ($F(16, 43) = 4.684$, $p < 0.001$, $R^2 = 0.49$). However, the individual predictors and the interactions did not significantly predict the mental demand scores (all $p_s > 0.05$). Similarly, the regression model which considered the frustration scores ($F(16, 43) = 2.455$, $p < 0.001$, $R^2 = 0.28$) highlighted a collective significant effect but no significant effect of individual predictors or interactions (all $p_s > 0.05$). Considering the other two

regression models, no collective significant effect between the predictors was shown (i.e., performance: $F(16, 43) = 1.862, p > 0.05, R^2 = 0.19$; effort: $F(16, 43) = 1.859, p > 0.05, R^2 = 0.19$).

A second multiple linear regression was performed to check if the implicit measures (i.e., time on task, fixation duration, fixation frequency, blink duration, blink frequency, saccade frequency, and heart rate) could predict the explicit one (i.e., task accuracy), also considering the task and all two-way interactions between I/E measures and the task. Results of the regression model indicated that there was a collective significant effect between the considered predictors ($F(15, 44) = 4.582, p < 0.001, R^2 = 0.48$). We further examined the individual predictors. The task ($B = 26.57, t = 2.335, p < 0.05$), the heart rate ($B = -0.052, t = -2.046, p < 0.05$), and the interactions respectively between fixation duration $\times$ task ($B = -0.23, t = -3.613, p < 0.001$) and blink Frequency $\times$ Task ($B = -0.159, t = -2.241, p < 0.05$) were able to significantly predict the task accuracy. In Figure 13, the significant interactions that emerged in the regression model are depicted.

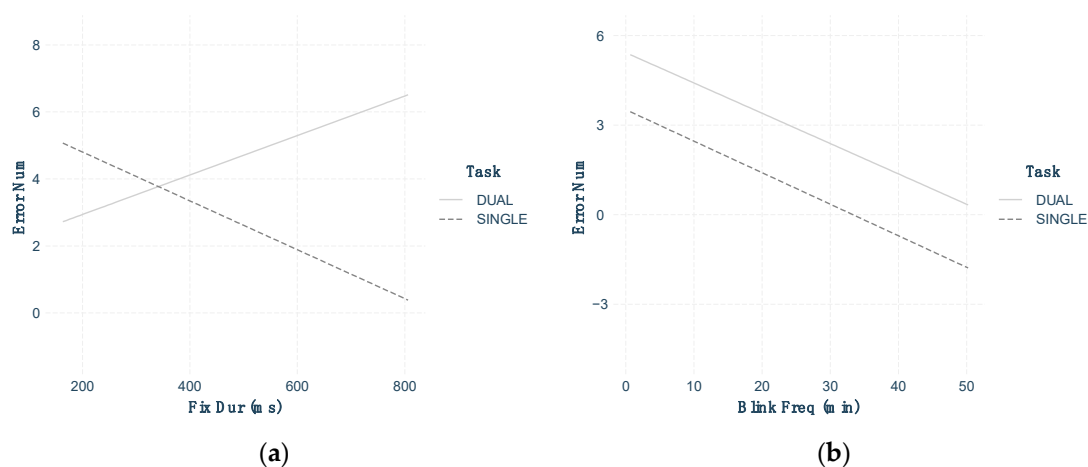


Figure 13. Interactions of the regression model: fixation duration $\times$ task (a) and blink frequency $\times$ task (b).

4. Discussion

This study first aimed to investigate the feasibility of combining I/E measures to understand if they could provide univocal interpreted pieces of information in an ecological task (i.e., repetitive work) characterized by the manipulation of the imposed mental workload. The second aim was to analyze whether the I/E measures' variations matched the perceptions of users about the workload experienced. Therefore, we intended to verify the participants' awareness regarding the mental demand imposed by the different tasks, because the adaptive changes in the functioning of future symbiotic systems will be based on the modification of I/E measures. Thus, in the case of a match between I/E measures and users' perceptions, the user will comprehend the actions implemented by the system (i.e., transparency). We have defined these two issues in terms of data interpretability and system transparency. We believe that whenever they will be solved, as in a symbiotic system, the HMI collaboration will be certainly improved. Regarding the interpretability, we compared the results concerning the ability of these measures to discriminate between two distinct levels of MWL evoked by two tasks of different levels of difficulty (i.e., single- vs. dual-task). The majority of the considered I/E measures (i.e., psychophysiological, behavioural, performance) were able to distinguish between the difficulty (and thus the MWL) of the two tasks.

Although the screwing task does not appear to be extremely difficult (i.e., a limited number of errors in both experimental sessions), when participants had to deal with the interference task (i.e., dual-task condition), they made more errors (i.e., 4.3 vs. 2.7). Furthermore, the outcomes related to this explicit measure reflected one of the considered implicit measures of performance (i.e., time on task: dual-task (DT), 873.36 s; single-task (ST), 692.97 s).

In our ecological paradigm, some of the indices concerning ocular behaviour have proven to be equally useful in discriminating the MWL levels, thus demonstrating once again the I/E binomial correspondence. These findings are also in line with the literature, considering laboratory settings, where a shorter duration and a lower number of fixations [21,23,26], lower blinks, and a lower saccade rate [22,26–28,39,40] were shown in dual-task conditions. Therefore, these ocular indices could be used in real-world working tasks with characteristics similar to the present ones.

Regarding the psychophysiological measures, the HR has also shown the ability to discriminate between the two levels of task difficulty, proving its reliability to monitor changes in the user's MWL. During the dual-task, the higher level of HR is consistent with previous studies that considered different tasks in terms of imposed mental load [22,23,41].

Some of the considered measures show contrasting results with respect to the literature. In the present research, the blink duration showed an increment in the higher mental load condition (i.e., DT), whereas previous studies reported a decrement in such circumstances [22,24,42]. Reviewing the DT experimental sessions' video-recordings, several participants were backward counting, keeping their eyes closed for prolonged time intervals. This occurrence affected the average blink duration, as has been noticed in internally focused tasks [43]. Furthermore, the interference task (i.e., backward counting) did not add visual demands on the participants, which have been proven as a factor that shortens blink duration [42,44,45].

The expected outcome considering the NNI was a higher dispersion of fixations in the dual-task (i.e., an NNI close to 1). The absence of a difference could be due to the participants' strategy, as observed in the dual-task session's videos. Indeed, while carrying out the screwing task, several participants were fixating unique points (i.e., clustering of fixations) on the blackboard to perform the interference task simultaneously. This visual behavior could influence the NNI value, reducing it in the dual-task (as clustering brings the value towards zero) compared to a situation of cognitive overload, in which this value is around 1 (randomness of fixation dispersion across the visual field; [13]).

Considering the pupil diameter, it was not sensitive in discriminating the experimental conditions, which likely did not present an extreme variation in terms of mental demands (also corroborated by the low level of errors committed in both conditions).

Therefore, the peculiar features of the proposed tasks highlighted how ocular behaviour could be influenced by the task to be performed. At the same time, they allow us to highlight the different levels of reliability of the eye-tracking measures. This suggests the importance of task type in the data interpretability and points out its importance as a factor that must be considered in the design and implementation of future symbiotic systems.

Concerning the EDA, no difference emerged as a function of task difficulty. This result contrasts with other studies that highlighted a large skin response in conditions of high mental load [18,21,25]. Evaluating the EDA as a function of time, the analysis showed a summation effect. Indeed, the signal did not return to baseline in the resting period, which followed the first task, and it increased further during the second task. The reason for this may have been the influence of ambient temperatures on the EDA signal [46]. Despite the use of a ventilation system, the room temperature could not be kept constant. This occurrence may result in a confounding factor when assessing the difference in mean EDA as a function of task difficulty.

This result, influenced by the paradigm's ecology, underlines how the working environment's characteristics play a critical role in the correct interpretation of the data. Usually, it is not possible to maintain a constant ambient temperature within industrial working settings. The EDA analysis outcomes highlighted that it could be strongly affected by the conditions of the surrounding environment whenever they are not actively controlled.

To analyze the data on a more individual level, we performed a multiple linear regression to investigate if the task accuracy (i.e., our explicit measure) could be predicted from the implicit measures and their interactions with the task. The regression outcomes showed that the task accuracy could be predicted considering the task (i.e., ST vs. DT), the HR, and considering the fixation duration and the

blink frequency in different ways in ST and DT (i.e., interactions). As the performance indices strongly indicate MWL levels [47], these results show how the HR and ocular indices can be interpretable in association with an explicit measure (i.e., task accuracy) to understand the experienced MWL level. Focusing on the HR, an increment in this measure, in both ST and DT, will reflect a decrement in task accuracy, and this occurrence is in line with our previous HR analysis as a function of task difficulty and the literature [23]. Considering the fixation duration and the task (Figure 13a), the regression model predicts that in DT when the mean fixation duration is very short (around 200 ms), we should expect the best task performance, whereas when it is longer (i.e., higher than ~400 ms), the task performance would be impaired. The opposite pattern is shown in ST. Indeed, the best task performance is predicted to occur in the case of a longer mean fixation duration (i.e., longer than ~400 ms), whereas the worst performance would occur in the case of a mean fixation duration around 200 ms. Concerning the blink frequency (Figure 13b), the model predicts the best performance in both tasks when the mean frequency/min is high (i.e., higher than 25–30 blink/min). Nevertheless, when the mean frequency of blinking is lower than ~20, as in the present experiment, the prediction is that the number of errors in DT should be around double that of ST. This last occurrence is in line with the analysis of task accuracy as a function of task difficulty (i.e., 4.3 errors in DT and 2.7 in ST).

The second problem that this study has addressed is that of system transparency. To test it, we analyzed users' perceptions to verify the correspondence between them and the I/E measurements used. Participants' subjective perceptions were assessed using the NASA-TLX questionnaire. The results demonstrated that the users were aware of the differences in the level of MWL linked to the experimental conditions. Indeed, mental demand, effort, frustration (i.e., all higher in the dual-task), and the perceived performance (i.e., lower in the dual-task) reflected this awareness. These outcomes are in line with the literature. In particular, an increased level of MWL has been correlated with dual-tasks [48] and tasks of very high difficulty [34–36]. These findings showed a consistent trend insofar as the results of the analyses of the subjective measures (i.e., NASA-TLX) matched the I/E measures, confirming our second hypothesis.

To deepen this topic, we also performed a multiple linear regression trying to predict the subjective results, starting from combining the I/E measures, the task, and the interactions of I/E measures–task. In the regression analyses, the scores of the four NASA-TLX scales (e.g., mental demand) were not predicted by the individual factors or the interactions. Nevertheless, it is necessary to consider that the participants were aware only of their performance accuracy, which was related to their NASA-TLX scores. Combining I/E measures to predict the subjective scores provided to the NASA-TLX questionnaire resulted in the absence of an effect, because the regression model included only one explicit variable and several implicit ones as predictors.

5. Conclusions

We designed and developed an ecological paradigm comprising an ad hoc workstation connected with devices for acquiring I/E data from users.

Some limitations of the study could be pointed out. First of all, to be able to generalize these results to real-world situations, the temporal length of the experimental sessions should be similar to the duration of real working activities (i.e., at least 1 h–2 h per condition). Thus, there will be the possibility to analyze changes in cognitive and psychophysiological measures over time due to tiredness and drowsiness. Second, insofar as the participants involved were mainly young adults, it is not possible to generalize the results to different workforce categories such as senior workers. Such workers may present different levels of acceptance [49], considering the wearable devices employed (i.e., eye trackers, surface electrodes). We will consider these aspects in subsequent studies.

However, the data collected showed that some implicit measures (i.e., eye behavior and heart activity) could be matched with the proposed explicit measure. This combination could be extremely informative about the type of data that can be exploited by a potential symbiotic system. Indeed, the considered measures proved their feasibility in monitoring variations in MWL, ensuring a

proper interpretation of users' data. The present findings pointed out that, considering the current experimental tasks, a symbiotic system could exploit the combination of I/E variables (i.e., task accuracy, HR, fixation duration, and blink frequency) to detect situations of high MWL while a participant is accomplishing the ST and DT. Indeed, in the ST, a high MWL (and relatedly, a lower task accuracy) will be reflected by a higher HR, a shorter fixation duration, and a lower blink frequency. In contrast, in the DT, a high MWL (and relatedly, a lower task accuracy) will be reflected by a higher HR, a longer mean fixation duration, and a lower mean blink frequency. By doing so, based on a different mean fixation duration, the system will be also capable of understanding whether the individual is in a condition of high MWL while performing the ST (i.e., shorter mean duration) or the DT (i.e., prolonged mean duration). In this way, the system will be capable of implementing a set of adaptive features to support properly, in both tasks, the individual who is experiencing a condition of high MWL. Moreover, these findings, as previously discussed in regard to data interpretability, underline once again that the type of task, as well as the environmental context, crucially influences the predictions based on these I/E measures.

Furthermore, the correspondence between the majority of I/E measures and the subjective ones (i.e., participants' perception of the MWL) suggested that the functioning of cutting-edge systems that will implement adaptive features based on these users' data will be transparent to the operators. The users will understand the changes that the system will operate in consequence of the variations in their internal state (MWL) linked to the working activities. Thus, the human-machine mutual understanding will result in a coordinated and fluent interaction, whereby the technology will support the users whenever needed [50].

In future works, we will implement devices for cognitive and psychophysiological monitoring in an adaptive workstation equipped with a cobot (i.e., that will adjust its operating to I/E measures related to the internal state of users). These experiments will be carried out first in laboratory settings and then in entirely ecological situations (i.e., inside a factory) to establish which design approach would be the most suitable with regards to the different characteristics of the task and the environment. A specific set of measures, in response to which the system will implement adaptive functions, will be grounded on features of both the tasks and environment. For instance, based on the results of this study, when the temperature of the working environment is non-constant, the EDA should not be considered.

Once the correct data interpretability and a high level of system transparency can be fully achieved, man and machine will have taken a fundamental step towards reaching a mutual understanding, which is necessary for the development of a symbiotic system.

Author Contributions: Conceptualization, M.M., D.B., P.P. and L.G.; methodology, M.M., D.B.; software, M.M.; validation, M.M.; formal analysis, P.P., D.B., M.M.; investigation, M.M., D.B.; resources, L.G.; data curation, M.M., D.B.; writing—original draft preparation, P.P., M.M., D.B.; writing—review and editing, P.P., D.B., C.R., M.M.; visualization, P.P.; supervision, L.G.; project administration, L.G.; funding acquisition, L.G. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by a grant from MIUR (Dipartimenti di Eccellenza DM 11/05/2017 n. 262) to the Department of General Psychology and partially supported by the European Commission under Grant [ID: 826266; Co-Adapt project].

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Thoben, K.D.; Wiesner, S.A.; Wuest, T. "Industrie 4.0" and Smart Manufacturing—a Review of Research Issues and Application Examples. *Int. J. Autom. Technol.* **2017**, *11*, 4–16. [[CrossRef](#)]
2. Licklider, J.C.R. Man-computer symbiosis. *IRE Trans. Hum. Factors Electron.* **1960**, *HFE-1*, 4–11. [[CrossRef](#)]
3. Gamberini, L.; Spagnoli, A. Towards a definition of symbiotic relations between humans and machines. In *International Workshop on Symbiotic Interaction*; Springer: Cham, Switzerland, 2017; Volume 9961, pp. 1–4. [[CrossRef](#)]

4. Schmidt, A. Implicit Human Computer Interaction through Context. In *Personal and Ubiquitous Computing*; Springer: London, UK, 2000; Volume 4, pp. 191–199. [\[CrossRef\]](#)
5. George, L.; Lécuyer, A.; An, A.L. An Overview of Research on “Passive” Brain-Computer Interfaces for Implicit Human-Computer Interaction. In Proceedings of the International Conference on Applied Bionics and Biomechanics ICABB 2010—Workshop W1 “Brain-Computer Interfacing and Virtual Reality”, Venise, Italy, 14–16 October 2010.
6. Rani, P.; Sarkar, N.; Smith, C.A.; Kirby, L.D. Anxiety Detecting Robotic System—Towards Implicit Human-Robot Collaboration. *Robotica* **2004**, *22*, 85–95. [\[CrossRef\]](#)
7. Birrell, S.; Young, M.; Stanton, N.; Jennings, P. Using Adaptive Interfaces to Encourage Smart Driving and Their Effect on Driver Workload. In *Advances in Intelligent Systems and Computing*; Springer: Cham, Switzerland, 2017; Volume 484, pp. 31–43. [\[CrossRef\]](#)
8. Maior, H.A.; Wilson, M.L.; Sharples, S. Workload Alerts-Using Physiological Measures of Mental Workload to Provide Feedback during Tasks. *ACM Trans. Comput. Interact.* **2018**, *25*. [\[CrossRef\]](#)
9. Liu, Y.; Du, S. Psychological Stress Level Detection Based on Electrodermal Activity. *Behav. Brain Res.* **2018**, *341*, 50–53. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Giannakakis, G.; Pediaditis, M.; Manousos, D.; Kazantzaki, E.; Chiarugi, F.; Simos, P.G.; Marias, K.; Tsiknakis, M. Stress and Anxiety Detection Using Facial Cues from Videos. *Biomed. Signal Process. Control* **2017**, *31*, 89–101. [\[CrossRef\]](#)
11. Kassem, K.; Salah, J.; Abdrabou, Y.; Morsy, M.; El-Gendy, R.; Abdelrahman, Y.; Abdennadher, S. DiVA: Exploring the Usage of Pupil Diameter to Elicit Valence and Arousal. *ACM Int. Conf. Proc. Ser.* **2017**, 273–278. [\[CrossRef\]](#)
12. Charles, R.L.; Nixon, J. Measuring Mental Workload Using Physiological Measures: A Systematic Review. *Appl. Ergon.* **2019**, *74*, 221–232. [\[CrossRef\]](#)
13. Duchowski, A.T. *Eye Tracking Methodology: Theory and Practice*, 3rd ed.; Springer: Cham, Switzerland, 2017. [\[CrossRef\]](#)
14. Spagnolli, A.; Conti, M.; Guerra, G.; Freeman, J.; Kirsh, D.; van Wynsberghe, A. Adapting the System to Users Based on Implicit Data: Ethical Risks and Possible Solutions. In *International Workshop on Symbiotic Interaction*; Springer: Cham, Switzerland, 2017; Volume 9961, pp. 5–22. [\[CrossRef\]](#)
15. Jacucci, G.; Spagnolli, A.; Freeman, J.; Gamberini, L. Symbiotic Interaction: A Critical Definition and Comparison to Other Human-Computer Paradigms. In *International Workshop on Symbiotic Interaction*; Springer: Cham, Switzerland, 2015; pp. 3–20. [\[CrossRef\]](#)
16. Mattsson, S.; Fast-Berglund, A.; Akerman, M. Assessing Operator Wellbeing through Physiological Measurements in Real-Time-Towards Industrial Application. *Technologies* **2017**, *5*, 61. [\[CrossRef\]](#)
17. Jaquess, K.J.; Lo, L.C.; Oh, H.; Lu, C.; Ginsberg, A.; Tan, Y.Y.; Lohse, K.R.; Miller, M.W.; Hatfield, B.D.; Gentili, R.J. Changes in Mental Workload and Motor Performance Throughout Multiple Practice Sessions Under Various Levels of Task Difficulty. *Neuroscience* **2018**, *393*, 305–318. [\[CrossRef\]](#)
18. O’Keeffe, K.; Hodder, S.; Lloyd, A. A Comparison of Methods Used for Inducing Mental Fatigue in Performance Research: Individualised, Dual-Task and Short Duration Cognitive Tests Are Most Effective. *Ergonomics* **2020**, *63*, 1–12. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Wickens, C.D. Mental Workload: Assessment, Prediction and Consequences. In *Communications in Computer and Information Science*; Springer: Cham, Switzerland, 2017; Volume 726, pp. 18–29. [\[CrossRef\]](#)
20. Van Acker, B.B.; Bombeke, K.; Durnez, W.; Parmentier, D.D.; Mateus, J.C.; Biondi, A.; Saldien, J.; Vlerick, P. Mobile pupillometry in manual assembly: A pilot study exploring the wearability and external validity of a renowned mental workload lab measure. *Int. J. Ind. Ergon.* **2020**, *75*, 102891. [\[CrossRef\]](#)
21. Foy, H.J.; Chapman, P. Mental Workload Is Reflected in Driver Behaviour, Physiology, Eye Movements and Prefrontal Cortex Activation. *Appl. Ergon.* **2018**, *73*, 90–99. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Tao, D.; Tan, H.; Wang, H.; Zhang, X.; Qu, X.; Zhang, T. A Systematic Review of Physiological Measures of Mental Workload. *Int. J. Environ. Res. Public Health* **2019**, *16*, 2716. [\[CrossRef\]](#) [\[PubMed\]](#)
23. De Rivecourt, M.; Kuperus, M.N.; Post, W.J.; Mulder, L.J.M. Cardiovascular and Eye Activity Measures as Indices for Momentary Changes in Mental Effort during Simulated Flight. *Ergonomics* **2008**, *51*, 1295–1319. [\[CrossRef\]](#) [\[PubMed\]](#)

24. Fournier, L.R.; Wilson, G.F.; Swain, C.R. Electrophysiological, Behavioral, and Subjective Indexes of Workload When Performing Multiple Tasks: Manipulations of Task Difficulty and Training. *Int. J. Psychophysiol.* **1999**, *31*, 129–145. [\[CrossRef\]](#)
25. Mehler, B.; Reimer, B.; Coughlin, J.F.; Dusek, J.A. Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *Transp. Res. Rec.* **2009**, *2138*, 6–12. [\[CrossRef\]](#)
26. Holmqvist, K. *Eye Tracking: A Comprehensive Guide to Methods and Measures*; OUP Oxford: Oxford, UK, 2011.
27. Faure, V.; Lobjois, R.; Benguigui, N. The Effects of Driving Environment Complexity and Dual Tasking on Drivers' Mental Workload and Eye Blink Behavior. *Transp. Res. Part F Traffic Psychol. Behav.* **2016**, *40*, 78–90. [\[CrossRef\]](#)
28. Tsai, Y.-F.; Viirre, E.; Strychacz, C.; Chase, B.; Jung, T.-P. Task Performance and Eye Activity: Predicting Behavior Relating to Cognitive Workload. In *Aviation, Space, and Environmental Medicine*; Aerospace Medical Association: Alexandria, VA, USA, 2007; Volume 78, pp. B176–B185.
29. Di Nocera, F.; Terenzi, M. A Random Glance at the Flight Deck: Pilots' Scanning Strategies and the Real-Time Assessment of Mental Workload. *J. Cogn. Eng. Decis. Mak.* **2007**, *1*, 271–285. [\[CrossRef\]](#)
30. Sturman, D.; Wiggins, M.W.; Auton, J.C.; Loft, S.; Helton, W.S.; Westbrook, J.I.; Braithwaite, J. Control Room Operators' Cue Utilization Predicts Cognitive Resource Consumption during Regular Operational Tasks. *Front. Psychol.* **2019**, *10*, 1967. [\[CrossRef\]](#)
31. Young, M.S.; Brookhuis, K.A.; Wickens, C.D.; Hancock, P.A. State of Science: Mental Workload in Ergonomics. *Ergonomics* **2015**, *58*, 1–17. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Hart, S.G. NASA-TLX: 20 Years Later. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2006**, *50*, 904–908. [\[CrossRef\]](#)
33. Galy, E.; Paxion, J.; Berthelon, C. Measuring Mental Workload with the NASA-TLX Needs to Examine Each Dimension Rather than Relying on the Global Score: An Example with Driving. *Ergonomics* **2018**, *61*, 517–527. [\[CrossRef\]](#)
34. Mansikka, H.; Virtanen, K.; Harris, D. Comparison of NASA-TLX Scale, Modified Cooper–Harper Scale and Mean Inter-Beat Interval as Measures of Pilot Mental Workload during Simulated Flight Tasks. *Ergonomics* **2019**, *62*, 246–254. [\[CrossRef\]](#)
35. Lowndes, B.R.; Forsyth, K.L.; Blocker, R.C.; Dean, P.G.; Truty, M.J.; Heller, S.F.; Blackmon, S.; Hallbeck, M.S.; Nelson, H. NASA-TLX Assessment of Surgeon Workload Variation across Specialties. *Ann. Surg.* **2020**, *271*, 686–692. [\[CrossRef\]](#)
36. Law, K.E.; Lowndes, B.R.; Kelley, S.R.; Blocker, R.C.; Larson, D.W.; Hallbeck, M.S.; Nelson, H. Surgeon Workload in Colorectal Surgery: Perceived Drivers of Procedural Difficulty. *J. Surg. Res.* **2020**, *245*, 57–63. [\[CrossRef\]](#)
37. Urbanek, S.; Bibiko, H.-J.; Stefano, M.L. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
38. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B.* **1995**, *57*, 289–300. [\[CrossRef\]](#)
39. Nakayama, M.; Takahashi, K.; Shimizu, Y. The Act of Task Difficulty and Eye-Movement Frequency for the “Oculo-Motor Indices”. In Proceedings of the 2002 Symposium on Eye Tracking Research & Applications, New Orleans, LA, USA, 25–27 March 2002; pp. 37–42. [\[CrossRef\]](#)
40. Wolkoff, P.; Nøjgaard, J.K.; Troiano, P.; Piccoli, B. Eye Complaints in the Office Environment: Precorneal Tear Film Integrity Influenced by Eye Blinking Efficiency. *Occup. Environ. Med.* **2005**, *62*, 4–12. [\[CrossRef\]](#)
41. Splawn, J.M.; Miller, M.E. Prediction of perceived workload from task performance and heart rate measures. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2013**, *57*, 778–782. [\[CrossRef\]](#)
42. Veltman, J.A.; Gaillard, A.W.K. Physiological Indices of Workload in a Simulated Flight Task. *Biol. Psychol.* **1996**, *42*, 323–342. [\[CrossRef\]](#)
43. Walcher, S.; Körner, C.; Benedek, M. Looking for ideas: Eye behavior during goal-directed internally focused cognition. *Conscious. Cogn.* **2017**, *53*, 165–175. [\[CrossRef\]](#)
44. Veltman, J.A.; Gaillard, A.W.K. Physiological Workload Reactions to Increasing Levels of Task Difficulty. *Ergonomics* **1998**, *41*, 656–669. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Benedetto, S.; Pedrotti, M.; Minin, L.; Baccino, T.; Re, A.; Montanari, R. Driver Workload and Eye Blink Duration. *Transp. Res. Part F Traffic Psychol. Behav.* **2011**, *14*, 199–208. [\[CrossRef\]](#)

46. Boucsein, W. *Electrodermal Activity*, 2nd ed.; Springer Science & Business Media: New York, NY, USA, 2012.
47. Bommer, S.C.; Fendley, M. A theoretical framework for evaluating mental workload resources in human systems design for manufacturing operations. *Int. J. Ind. Ergon.* **2018**, *63*, 7–17. [[CrossRef](#)]
48. Rubio, S.; Díaz, E.; Martín, J.; Puente, J.M. Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods. *Appl. Psychol.* **2004**, *53*, 61–86. [[CrossRef](#)]
49. Marangunić, N.; Granić, A. Technology Acceptance Model: A Literature Review from 1986 to 2013. *Univers. Access Inf. Soc.* **2015**, *14*, 81–95. [[CrossRef](#)]
50. Liang, C.; Proft, J.; Andersen, E.; Knepper, R.A. Implicit Communication of Actionable Information in Human-AI Teams. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).