2019

H-Workload 2019: 3rd International Symposium on Human Mental Workload: Models and Applications (Works in Progress)

Luca Longo
Technological University Dublin, luca.longo@tudublin.ie

Maria Chiara Leva
Technological University Dublin, maria.leva@tudublin.ie

H-WORKLOAD
models & applications
November 14th and 15th
Rome 2019

The Third International Symposium on Human Mental Workload: Models and Applications

Works in progress

SAPIENZA
UNIVERSITÀ DI ROMA

Contents

Preface

This book contains the work-in-progress research contributions presented at the Third International Symposium on Human Mental Workload, models and applications. It presents recent developments in the context of theoretical models of mental workload and practical applications. Additionally, it aims to stimulate and encourage discussion on mental workload, its measures, dimensions, models, applications and consequences. It is a topic that demands a multidisciplinary approach, spanning across Human Factors, Computer Science, Psychology, Neuroscience, Statistics and Cognitive Sciences.

From the content of these research contributions, it is clear that mental workload, as a multidimensional and multifaceted construct, is still under definition, development, and investigation. This is one of the reasons why mental workload is today a keyword used and abused in life sciences, as pointed by Prof. Fabio Babiloni. However, despite the difficulty in precisely defining and modeling it, the capacity to assess human mental workload is a key element in designing and implementing information-based procedures and interactive technologies that maximize human performance. Some of the articles published in this book applied psychological subjective self-reporting measures, others made use of primary task measures and some a combination of these. Physiological measures in general, and more specifically electroencephalography (EEG), have been gaining a more prominent role, thanks to advances in data-gathering technology as well as a growing availability of computational power and classification techniques offered by the discipline of artificial intelligence. This is also reflected in the present book where half of the chapters focus on the development of novel models of mental workload employing data-driven techniques, borrowed from machine learning. However, one of the key issues in modeling mental workload employing automated learning techniques is that, although it often leads to accurate and robust models, they lack explanatory capacity. This problem is fundamental if we want to define mental workload for the fields of human factors, human–computer interaction, and in general for human-centered designers. Thus, we believe that future research efforts on mental workload modeling should employ a mix of measures as well as qualitative and quantitative research methods to not only assess mental workload but also to understand its meaning and implications on the individuals and our approach toward work and life.

Dr. Luca Longo, Technological University Dublin
Dr. Maria Chiara Leva, Technological University Dublin

Identifying predictive EEG features for cognitive overload detection in assembly workers in Industry 4.0

Jessica Morton[1], Pieter Vanneste[2], Charlotte Larmuseau[2], Bram B. Van Acker[1], Annelies Raes[2], Klaas Bombeke[1], Frederik Cornillie[2], Jelle Saldien[1], Lieven De Marez[1]

[1] imec-mict-UGent, Platteberg 11,
9000 Gent, Belgium
[2]imec-itec-KU Leuven, Etienne Sabbelaan 51,
8500 Kortrijk, Belgium

Abstract. Industry 4.0 will be characterized by far-reaching production automation because of recent advancements in robotics and artificial intelligence. As a consequence, a lot of simple, repetitive assembly tasks will no longer be performed by factory workers, but by machines. However, at the same time, consumers demand more and more personalized products, increasing the need for human assembly workers who can adapt quickly to new and more complex assembly procedures. This need for adaptation is most likely to increase the cognitive workload and potentially *overload* assembly workers that were already having a hard time during more traditional assembly work. Several studies have tried to identify this cognitive overload in the EEG signal, but many failed because of poor experimental measurement procedures, bad data quality and low sample sizes. In this paper, we therefore designed a highly controlled lab experiment to collect EEG data of a large number of participants (N=46) performing an assembly task under various levels of cognitive load (low, high, overload). This systematic approach allowed us to study which EEG features are particularly useful and valid for cognitive overload assessment in the context of assembly work.

Keywords: Industry 4.0, assembly work, cognitive overload, EEG

1 Introduction

1.1 Assembly work in Industry 4.0 and cognitive load

Industry 4.0 or "smart factories" of the future will be characterized by wide-scale automatization, connectivity and AI-driven technology, resulting in a manufacturing process that will become more and more efficient [1,2,3]. It is beyond any doubt that many jobs involving simple, repetitive tasks will disappear in favor of robots or at least cobots (i.e., machines that physically interact with human workers). However, at the same time, it is expected that customer demand will push the industry towards increasing product variety to allow for broad product personalization [4,5,6]. For example, in car manufacturing, it is more common that customers have the ability to decide on design specifications compared to the past.

Hence, amidst this evolution stands the human assembly worker who will need to operate more and more in a flexible way and will be required to constantly adjust his or her skills to changing job demands and technology [7,8,9]. Since it is not unlikely that this increasing complexity and need for flexibility will make it harder for this human worker to do the job in a proper way, it is highly important to accurately measure *cognitive load* and explore ways to avoid or reduce this load from a cognitive ergonomic point of view [10,11]. In this paper, we therefore focus on cognitive load detection in the context of personalized assembly work.

Throughout the history of (cognitive) ergonomics, the construct of cognitive load has been playing a substantive role in the prevention of occupational error, safety hazard, and negative (physical) stress caused by overload [4,12]. Cognitive load is a multi-dimensional, rather than a unitary construct and covers working memory processes ranging from attention and perception to memory and decision making [13]. Originally, the concept of cognitive load evolved from early work in the instructional and educational research field, eventually coming together in a widely-applied theory called cognitive load theory (CLT) [14,15,16]. Resonating with the multidimensional nature of the cognitive load concept, cognitive load measures are equally various in nature. In general, the literature converges towards assessing cognitive load based on subjective self-reporting and psychophysiological measurements [13,17,18,19,20,21].

Whereas a lot of research has been done on how to accurately question people about their cognitive load using questionnaires and in-depth interviews, there is still a lot of work to be done with respect to using psychophysiological data to assess cognitive load. Interestingly enough, recent innovations and advancements in wearable technology have led to low-cost, easy-to-wear, energy-efficient devices to measure electrical activity at the human scalp. Therefore, it is expected that cognitive load measurement based on psychophysiological EEG data will become very prominent in the future [11,22,23,24,25]. Being able to rather noninvasively measure brain activity in a real-world context in a relatively cheap way has triggered the interest of both the industry and academic cognitive load community. As a consequence, there are already many studies available in which researchers looked at the relationship between cognitive load and changes in the EEG signal [for a review, 26,27,28]. Unfortunately, many of these studies do not succeed in obtaining valid and reliable conclusions because of methodological flaws in design, issues with poor experimental measurement procedures or settings, bad data quality, and low sample sizes. This is not surprising, since inter-individual differences in EEG recordings can be very high and signals are prone to artefacts caused by technical malfunctioning, facial muscle activity and static noise coming from other electrical sources in the assessment setting. For this reason and because replication is an important characteristic of scientific research, we chose to take one step back and study predictive EEG features for cognitive overload in a highly controlled lab setting instead of at the factory floor right away (although it is beyond any doubt that the latter should be the end goal). By choosing this approach, we hope to overcome the aforementioned problems.

1.2 EEG and cognitive load

There are basically two approaches to analyze EEG data. First, spectral analysis of oscillatory activity can be used to convert time series data (electrical current fluctuating over time) to frequency domain data (the *frequencies* that represent these fluctuations). By separating the signal into different frequency "bands" (i.e., delta, theta, alpha, beta, and gamma, representing slower to faster signals), different cognitive and affective processes can be monitored [29]. The most interesting finding with respect to cognitive load measurements is that alpha activity suppression (decrease in power of frequencies oscillating between 8 and 12 Hz at parietal regions) has found to be associated with increasing task difficulty and load across a wide variety of tasks [30,31].

The second approach to analyze EEG data is to look at the event-related potentials, representing the changes in mean voltage preceding or following a stimulus or action of interest (hearing a sound or pressing a button). By averaging over several trial repetitions, this analysis is focuses on the specific stimulus-related activity and decreases the impact of any activity that is not related or within the time window of the occurrence of the event (which increases the signal-to-noise (SNR) ratio) [32]. Interestingly enough, some ERP components can reflect the extent to which cognitive function and sensory processing are affected by mental workload. The amplitudes of ERP components such as N1, N2, P2 and P3 are expected to be reduced when a primary task becomes more demanding and workload increases [31,33]. An auditory oddball paradigm, in which sounds of different frequencies are presented, allows us to study the high demand on general processing resources reflected in the ERP components. With an irrelevant-probe technique this can be done in a non-intrusive way, without interference on the task flow [34,35,36]. Contrary to the standard ERP design, in this technique the ERP-eliciting stimuli (sounds) are presented without requiring participants to actively attend or respond to them, thus not co-varying with task demands.

1.3 The current study

As mentioned before, we wanted to focus on cognitive load detection in the context of personalized assembly work using the EEG method. More precisely, the main goal is the collection and in-depth analysis of both performance measures, subjective measures and psychophysiological measures, in a highly controlled lab context that overcomes some of the issues that previous studies had to deal with. The main difference with previous work is that next to the low and high load condition, we also introduced a condition in which cognitive overload was induced. In the current study, load was induced in an experimental setting by manipulating complexity levels of a set of Tangram tasks combined with working memory load (i.e., remembering visual stimuli). The cognitive overload condition included the most difficult Tangram puzzles and the greatest amount of stimuli to remember. We expected that the majority of the participants would not be able to succeed in these tasks and that this would be accompanied by feelings of despair, giving up and being discouraged in completing the task. Also, the manipulation with Tangram puzzles was used in order to have a representative task for assembly performance, which requires similar spatial intelligence skills. Another additive value in this study is that all three load conditions have an equal length of duration, keeping the data balanced for statistical comparisons. All conditions lasted for 10 minutes, which is a substantial amount of time to be able to measure cognitive load. Additionally, baselining was carefully conducted with 4 minute measures in rest state before and after the experimental block with load conditions. Finally, this study had a multimodal approach with additional sensors in order to explore other potential and less studied cognitive load markers (i.e., heart rate, skin temperature, galvanic skin response, electro-ocular activity, motion analysis, facial video analysis). Also, additional EEG features such as band power activity of other band frequencies, other event-related components, the time-frequency spectrogram, spectral entropy, individual alpha peak frequency, and auto-correlation can be explored in this dataset. All these features are beyond the scope of this manuscript and will be analyzed in the future.

With this optimized research design, we aimed to investigate the following hypotheses:

H1: The induced cognitive load by manipulation of complexity levels of the Tangram task is also reflected in subjective ratings of mental investment. The more complex the task, the more mental investment will be reported.
H2: Task performance on the Tangram task will be reduced in the cognitive overload condition, compared to the high and low load condition.
H3: Alpha power is decreased at parietal electrode sites when performing Tangram tasks that induce a cognitive overload compared to Tangram tasks that induce low or high load.
H4: The auditory processing of sounds presented during the performance of the Tangram task that induces cognitive overload can be reflected in a decreased N2 amplitude, when compared to the Tangram tasks that induce low or high load.

2 Method

2.1 Participants

This research got the approval of the ethics committee of the Faculty of Political and Social Sciences at Ghent University. In addition, all participants read and agreed to sign an informed consent with information about the procedure, purpose, voluntary participation, right to decline, access and storage of data.

In this study, 46 participants aged between 19 and 40 years old (M = 25.8, SD = 4.19) were recruited based on a questionnaire inquiring education, hair type, and other requirements via different social media channels (Facebook, the channel of the public library and the channel of the University).

Each session had a duration of approximately 90 minutes. We strived for a more or less equal number of male (N=21) and female (N=25) participants. The participants differed somewhat regarding their background in education: 11 participants had secondary education as highest degree, 6 participants completed a professional bachelor, 4 participants completed an academic bachelor, 24 participants completed an academic master, and 1 participant owned a PhD as highest degree.

To control for prior experience and knowledge, participants were first asked about their experience with Tangram puzzles. Most participants (67 percent) indicated to be rather inexperienced regarding the Tangram task to be conducted in the experiment, while 11 percent were neutral and 22 percent indicated to have had some amount of experience with Tangram puzzles. Additionally, a spatial ability test was conducted with an adapted version of the Revised Minnesota Task Load Index [37]. Only 20 of the total of 64 questions were included, still covering the entire difficulty range. The histogram in Figure 1 shows the results on the spatial ability test, indicating a desired variance.

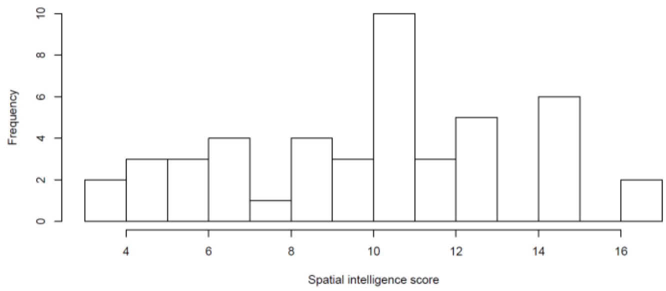


Fig. 1. Histogram of participants' average score on the spatial intelligence test (max score 20).

2.2 Research design & procedure

Design. In this experiment, a within-subjects design was used in which each participant was exposed to all experimental conditions (i.e., three levels of cognitive load: respectively a low, high, and overload level). As mentioned before, the length of the Tangram task was kept equal for all experimental conditions (i.e., 10 minutes). Thus, the experiment consisted of three phases for which a counterbalanced design, with 6 possible orders, was used to exclude possible learning effects and order effects. Baselines were measured before the first and after the last experimental phase. The independent variable was the induced cognitive load.

Procedure. At first, participants filled in the informed consent and the pretest measuring their spatial ability. Next, the testing equipment was prepared (i.e., external electrodes on mastoids and the EEG set). After the set-up, each participant got detailed instructions about the experimental procedure in a systematic way. Before starting the first experimental phase, a resting state measurement was conducted for the baseline. Participants subsequently opened and closed their eyes, each for 2 minutes. Next, the main experimental phase started in which participants spent 10 minutes in each condition. After each condition, participants completed a one-page questionnaire gauging perceived load, perceived affective states, and memory of visual stimuli. Finally, after completing all the experimental conditions, a post baseline measurement was conducted. The final step consisted of a participant briefing and the clean-up. Figure 2 shows the experimental setting when performing the Tangram task.

Fig. 2. Experimental setting when performing the Tangram task.

2.3 Materials & questionnaires

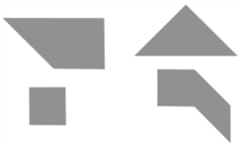
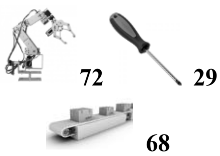
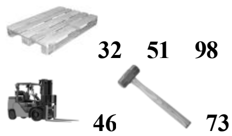
Materials. The aim of the experimental design was to induce different levels of cognitive load, which should allow us to identify physiological parameters that are explanatory for cognitive load. Because each method of inducing cognitive load may have different shortcomings, a combination of different methods was employed.

The first method to induce cognitive load was by manipulating the complexity of the task. In the low load phase, participants assembled a Tangram puzzle of which the contours of each of the seven pieces were individually visible. In the high load phase, three pairs of two pieces touched each other, so only the surrounding contour of the pair was visible. In the overload phase all seven pieces touched each other, which created only one surrounding contour, and making it even more challenging to find the correct assembly. Two different versions were created for each load phase in which the order of the Tangram puzzles was randomly shuffled.

In order to induce additional cognitive load, the participants' working memory was addressed by asking them to remember visual stimuli simultaneously while performing the assembly task. Participants were asked to write down the stimuli they remembered after each phase. Two different kinds of stimuli were alternately presented: pictures representing a tool that is typically used in industry (such as a safety helmet, a conveyor belt or a drilling machine) or a two-digit number. The number of stimuli that had to be remembered differed for each phase. During the low load phase, two pictures and two numbers were presented. During the high load phase, three pictures and three numbers were presented. And finally, five pictures and five numbers had to be remembered in the overload phase.

The third way to experimentally vary the level of cognitive load was to include background sounds and noise. For generalization to real-life assembly work, ambient factory floor sounds were played in the background. Additionally, two different sounds that differed in frequency were presented for the ERP analysis. About every 5 seconds (with some jitter to avoid predictability and rhythmic effects) a beep tone was played. 80 percent of these beeps were standard sounds with a low tone, while 20 percent were deviant oddballs with a higher pitch. This manipulation allowed us to study sensory processing of the sounds under different levels of cognitive load.

Table 1. Methods for inducing three different levels of cognitive load.

<table>
<tr><th>Low cognitive load</th><th>High cognitive load</th><th>Cognitive overload</th></tr>
<tr><td colspan="3">For each condition participants perform a series of Tangram tasks for 10 minutes, built up analogously as the examples below.</td></tr>
<tr><td>The contours of all seven pieces are each individually visible (no touching sides).
</td><td>Three pairs of pieces have touching sides. The contour of the seventh piece is visible.
</td><td>All pieces can have multiple touching sides.
</td></tr>
<tr><td colspan="3">Visual stimuli that are alternately shown on a computer screen in front of the participant have to be remembered while performing the Tangram task.</td></tr>
<tr><td>Two pictures related to industry and two two-digit numbers.
</td><td>Three pictures related to industry and three two-digit numbers
</td><td>Five pictures related to industry and five two-digit numbers.
</td></tr>
</table>

Performance. Tangram task performance was measured by the number and percentage of correctly assembled puzzles and the percentage of remembered visual stimuli.

Questionnaires. The subjective experience of cognitive load was measured by a continuous scale (ranging from 0 to 100) and a Likert scale (ranging from 1 to 7). Based on an adapted version of the NASA TLX questionnaire task complexity and mental investment were inquired [38].

2.4 Apparatus & analysis

The EEG was acquired with a Biosemi ActiveTwo measurement system (BioSemi, Amsterdam, Netherlands), using 64 Ag-AgCl scalp electrodes attached to a standard international 10–20 system cap. Two additional external electrodes were attached to the left and right mastoids, which were used for offline re-referencing. Signals were amplified and digitized with a sampling rate of 1024 Hz. Triggers were sent through a serial port via Psychopy, an open-source application for a wide range of neuroscience, psychology and psychophysics experiments, written in Python language [39]. The recording computer received these triggers for the start of both baselines, all experimental conditions and every trial a standard or oddball sound was presented.

EEG analysis was performed in Python with MNE, an open-source Python software for exploring, visualizing, and analyzing human neurophysiological data, and custom-made code [40,41]. The raw EEG data preprocessing included re-referencing to the mastoid channels, interpolation of bad channels, and a bandpass filter with a high cut-off frequency of 1 Hz and a low cut-off frequency of 45 Hz to eliminate movements and electric noise. The preprocessed data was also normalized by subtracting the average baseline activity, measured when participants relaxed with their eyes open for 2 minutes. Finally, for ease of analyzing purposes and processing speed, data was downsampled to 100 Hz.

Next, the pre-processed signal was transformed to the frequency domain with Fourier Transform for power analysis (focus on alpha oscillations). The power spectral density (PSD) was computed using Welch's method [42]. The Python function scipy.signal.welch computed an estimate of the PSD by averaging consecutive Fourier transform of small windows of the signal (segments of 2 seconds) without overlapping, resulting in a frequency resolution of 0.50 Hz. Absolute alpha bandpower was calculated by taking the absolute mean of the power for the frequency band within its range of 8 to 12 Hz.

Finally, the preprocessed signal was kept in the time domain for the analysis of the event-related component (i.e., the amplitude of the N2 component). All standard sound and oddball sound trials were epoched with a time window of [-200,500], and combined in an overall value at electrodes C3 and C4 at the central region.

3 Results

3.1 Self-reported cognitive load & task performance

Task complexity. The experimental manipulation in terms of complexity was as desired. Participants indicated the low load condition as the least complex and the overload condition as the most complex, with the high load condition in between, $F(2,84) = 172.04$, $p < .001$, $\eta_p^2 = .79$ (see Figure 3A). No significant correlation was established between spatial intelligence of the participants and how complex they perceived the task, $r = -.11$, $p = .19$.

Mental investment. The results indicate that there is a significant main effect for the different conditions on mental investment measured on the continuous scale, $F(2,90) = 196.62$, $p < .001$, $\eta_p^2 = .81$ (see Figure 3B). The more load that was induced in the Tangram task, the more mental investment participants reported. Also, there is no significant correlation between participants' spatial intelligence and their experienced mental investment during the task, $r = -.13$, $p = .13$.

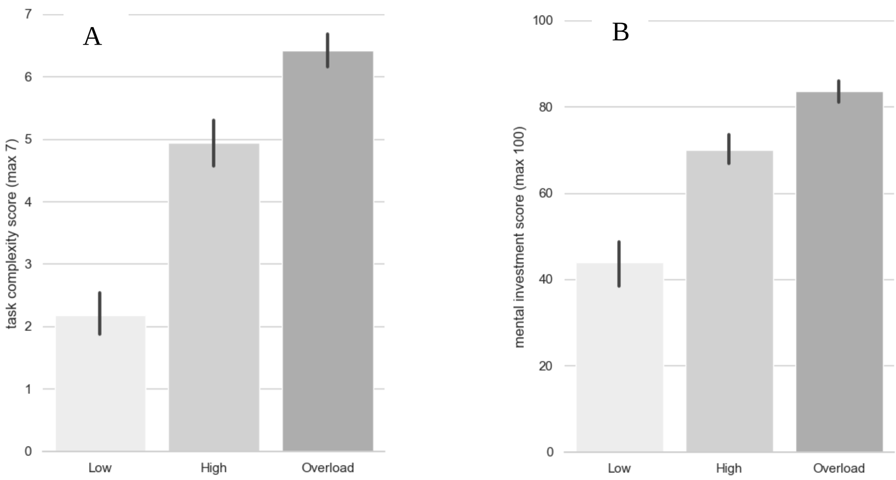


Fig. 3. A) Task complexity score (max score 7) and B) Mental investment score (max score 100) rated by the participants after each experimental condition.

Task performance. As expected, results show that the amount of correctly assembled Tangram puzzles significantly differs across conditions, F(2,90) = 539, $p < .001$, $\eta_p^2 = .92$ (see Table 2, Figure 4A). A smaller amount of Tangram puzzles were correctly assembled with the increasing task complexity. There is a significant correlation between the spatial intelligence and the number of correctly assembled Tangram puzzles ($r = .40$, $p < .001$). In a similar way, the proportion of remembered stimuli decreased with increasing complexity, F(2,90) = 51.68, $p < .001$, $\eta_p^2 = .54$ (see Table 3, Figure 4B).

Table 2. Comparison of all conditions with pairwise t-tests for performance Tangram task results (*p*-adjusted Holm).

Cond A	Cond B	T	*p*	Hedges g
low	high	20.680	<.001	2.849
low	overload	34.215	<.001	7.797
high	overload	10.858	<.001	2.416

Table 3. Comparison of all conditions with pairwise t-tests for performance Memory task results (*p*-adjusted Holm).

Cond A	Cond B	T	*p*	Hedges g
high	low	-3.812	<.001	-0.827
high	overload	4.596	<.001	0.894
low	overload	10.652	<.001	1.881

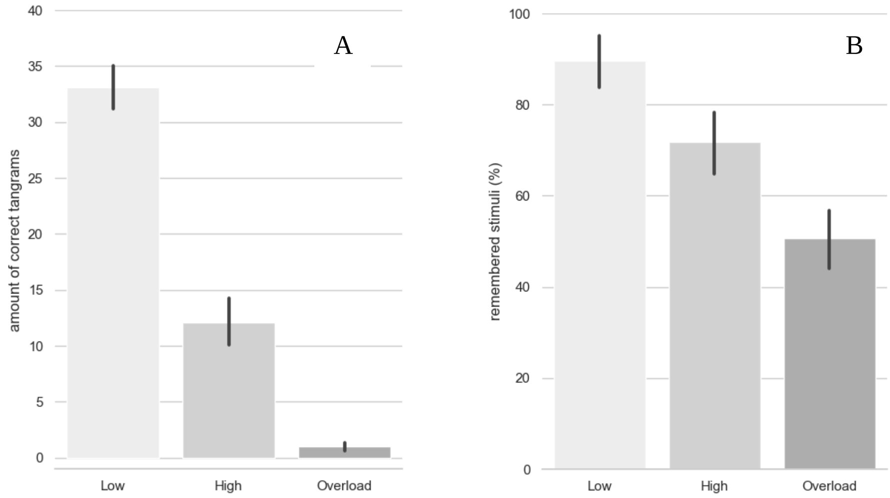


Fig. 4. A) Amount of correctly assembled Tangram puzzles and B) percentage of remembered visual stimuli in experimental load conditions.

3.2 Alpha power

Absolute mean power. As expected, alpha activity in the region of interest (i.e., the four selected electrodes in the parietal region: Pz, POz, P1 and P2) differed between conditions in a one-way ANOVA repeated measures test, $F(4,176) = 45.12$, $p < .001$, $\eta_p^2 = .51$. Table 4 summarizes the pairwise t-tests and Holm adjusted p-values. Both baseline conditions showed more alpha power activity compared to the cognitive load conditions. More importantly, alpha power differed between load conditions, $F(2,88) = 4.70$, $p = 0.01$, $\eta_p^2 = .07$. A lower amount of alpha power in parietal electrodes was found for the overload condition when compared to the high load condition, and similarly when compared to the low load condition. High load condition also showed a lower amount of alpha power when compared to the low load condition (see Figure 5).

Table 4. Comparison of all conditions with pairwise t-tests for alpha activity results (*p*-adjusted Holm).

Cond A	Cond B	T	*p*	Hedges g
base post	base pre	10.997	<.001	0.217
base post	overload	16.359	<.001	1.289
base post	low	14.386	<.001	1.176
base post	high	16.246	<.001	1.220
base pre	overload	11.962	<.001	0.995

base pre	low	10.227	<.001	0.886
base pre	high	11.724	<.001	0.933
overload	low	-5.473	<.001	-0.140
overload	high	-2.695	0.008	-0.057
low	high	3.499	0.001	0.080

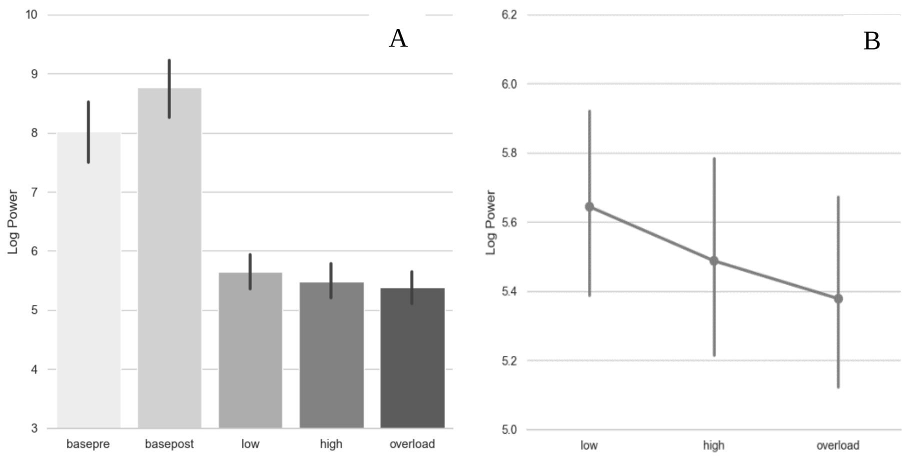


Fig. 5. A) Alpha log power activity results for all conditions and B) experimental load conditions only.

Differences in alpha activity power were also observed when looking closer at the separate electrodes, $F(3,132) = 38.59$, $p = 0.00$, $\eta_p^2 = .47$ (see Table 5 and Figure 6). Especially the POz electrode showed to have greater alpha power overall, without taking load conditions into account. There were no differences between electrodes in predicting the decreased alpha power effect in the different conditions, $F(6,264) = 1.25$, $p = 0.28$.

Table 5. Comparison of all electrodes with pairwise t-tests for alpha activity results (*p*-adjusted Holm).

Electr A	Electr B	T	*p*	Hedges g
P1	P2	-1.715	0.089	-0.071
P1	POz	-24.151	<.001	-0.587
P1	Pz	-11.141	<.001	-0.135
P2	POz	-10.936	<.001	-0.488
P2	Pz	-1.315	0.191	-0.057
POz	Pz	17.729	<.001	0.462

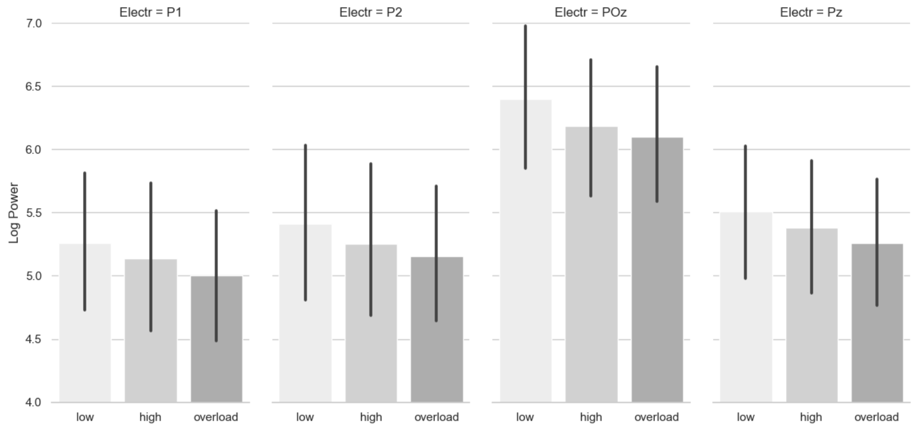


Fig. 6. Alpha log power activity results for experimental load conditions at four selected electrode sites in the parietal region (Pz, POz, P1, P2).

3.3 Auditory event-related potentials

The auditory processing of standard and oddball sounds is measured by the N2 component, averaged at central electrode sites C3 and C4. There were no significant differences found between the three load conditions, $F(2,84) = 0.50$, $p = 0.61$ (see Figure 7 and Figure 8).

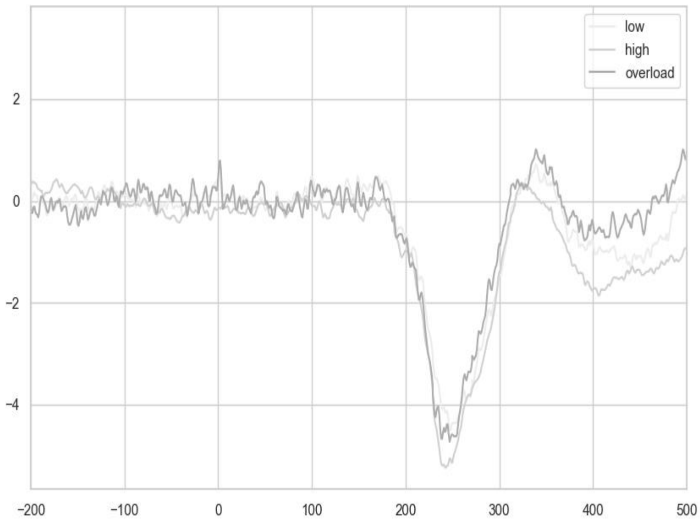


Fig. 7. ERP plot for experimental load conditions averaged for two electrodes at the central region (C3+C4).

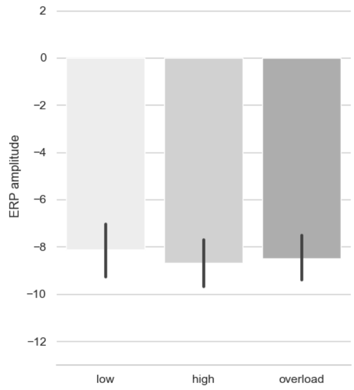


Fig. 8. ERP amplitude results for experimental load conditions averaged for two electrodes at the central region (C3+C4).

3.4 Additional correlation analysis

Correlation analysis indicated that no correlation was found between Tangram performance and alpha log power activity in load conditions. But percentage of memory performance and alpha log power activity were significantly correlated, only in the high load condition, r = -.37, p = .01. This indicated that only in the high load condition, reduced alpha power activity is correlated with a greater amount of visual stimuli participants could remember (see Table 6).

Table 6. Correlation analysis results for two performance features (amount of correctly assembled Tangram puzzles and percentage of visual stimuli remembered) with alpha log power activity in all experimental load conditions. Values in brackets indicate p value for each correlation.

Performance feature	Low	High	Overload
Tangram	-0,13 (.41)	-0,16 (.28)	-0,11 (.49)
Memory (%)	-0,17 (.25)	-0,37 (.01)	0,01 (.93)

4 Discussion

The current study focused on the detection of cognitive load and overload in the context of personalized assembly work with psychophysiological sensors, performance measures, and subjective measures. We manipulated cognitive load in three conditions (i.e., low, high, overload) by creating different complexity levels of a dual task, which included a Tangram puzzle task and a working memory load (i.e., remembering visual stimuli). This task was performed with a large sample size and in a highly controlled lab context to overcome some of the methodological issues that previous studies had to deal with. The experimental design also allowed us to conduct considerable baseline measures and compare load conditions that lasted for an equal and substantial length of time. Finally, while this manuscript only focuses on EEG features, other sensors (i.e., heart rate, skin temperature, galvanic skin response, electro-ocular activity, motion analysis, facial video analysis) were implemented in the current study and will also be explored as potential cognitive load markers in the future.

Our successful manipulation of cognitive load was reflected in task performance and subjective rating results. First, the majority of the participants were not able to succeed on the dual task in the overload condition, assembling almost none of the puzzles and remembering only a small amount of the presented visual stimuli. Second, on subjective ratings they indicated that the task in the overload condition was the most complex and required the greatest amount of mental investment.

More importantly, results for the EEG features alpha power activity and auditory response (i.e., N2 amplitude) are partially in line with expectations. A greater amount of cognitive load was indicated by reduced alpha power activity at parietal electrodes, especially at the POz electrode. On the contrary, no significant effect was found on the auditory response. The reduced alpha power effect found in the current study validates this EEG feature as a marker for estimating cognitive load, in line with previous research [10,11,31,43,44]. However, we expected there would be a greater effect on alpha power in the overload condition when compared to the other load conditions. This could be due to the task being too complex and overwhelming, making participants give up and not staying motivated to invest mental effort and resources anymore. Participants confirmed that these puzzles were too difficult and some believed they were actually unsolvable, which was reflected in the nervous laughs and freeze reactions. Regarding the assembly work context, this could be reflected in dropout, bad quality and errors because operators are becoming apathetic to the task performance [19,45]. Motion analysis of the videos or additional EEG features that study the EEG signal over time (i.e., time frequency spectrogram, auto-correlation) could provide more insights.

Also, the small effect size in these findings indicates that the alpha power may be not sensitive enough for differentiating between different levels of cognitive load. The differentiation between conditions in resting state (i.e., pre and post baselines) and conditions that require mental effort (i.e., experimental load conditions) was more pronounced than the comparison amidst only load conditions. The baselines had a distinctly lower amount of alpha power when compared to the load conditions. Consequently, the real-time differentiation between cognitive load versus overload in an assembly work context is challenging, especially when using alpha power activity as a deciding marker.

The results regarding the sensory processing with ERP analysis could not validate the N2 amplitude as a marker for estimating cognitive load. The sensory processing of the presented sounds was similar in all load conditions. First of all, factory noise and the ERP-eliciting sounds were presented in order to create additional cognitive load and reflect the ambience of assembly work for all conditions. Because a dual task was already created for manipulation of cognitive load, also attending to these sounds would have been too difficult. That way we would not have been able to create a low load condition. Consequently, in this ERP paradigm participants did not have to actively attend to the standard and oddball sounds. The N2 amplitude was possibly not sensitive enough as a marker of cognitive load because it was not part of the task flow. Another possible feature for measuring the sensory processing without overt action or attention to the presented sensory stimuli is the mismatch negativity (MMN) component and could be explored in the future [34]. The MMN indicates the event-related response to sudden changes in auditory stimuli. Finally, the low amount of trial repetitions is another possible confound in our ERP paradigm. Even though conditions lasted for 10 minutes, presenting a sound every second would have been too interfering with the primary task. Additionally, because an ERP design requires enough repetitive trials for filtering out noise and obtaining reliable conclusions, it may even be unsuitable in real-time assessment of cognitive load [11].

With regard to the finding that alpha power activity did not correlate with performance measures, we can remark that this spectral power feature may not be sensitive enough to discriminate on an aggregated level. A lot of information is lost because values for alpha power activity are averaged for the whole duration of the condition. The investigation of lower (8-10Hz) and upper (10-12Hz) alpha bands could give more detailed insights on specific frequency effects that are not distinct when only looking at the broad alpha range [44, 46].

As previously mentioned, we would like to explore other potential markers for cognitive load in our EEG dataset. Exploring the power activity in time and auto-correlation analysis of the raw data could possibly unravel more in-depth insights about fluctuations or recurring patterns in the signal, especially when synchronized with video motion analysis. Future research will also focus on other aggregated EEG features such as the alpha peak power frequency, the frequency bin where maximum power (i.e., local peak) is found within the 8-12 Hz range [44, 47]. We expect to find lower peak frequency values in the overload condition, indicating "less integrated and interconnected feedback loops among brain areas" [47, p.419] compared to lower load conditions. This is also reflected in the deteriorated task performance results. Another approach for a more nuanced analysis of our results is the use of advanced machine learning techniques in order to classify 'overload' within conditions based on the data from multiple sensors.

We can conclude that our results encourage to measure and evaluate EEG features for estimating cognitive load in a highly controlled lab setting with experimental design. The results from the current study validated alpha power activity as a potential marker for estimating cognitive load, while the auditory N2 response failed to differentiate between load conditions. Our future research focusing on the real-time measurement of cognitive load will aim to validate these findings with a wearable EEG headset that is applicable in the assembly work context. Especially alpha power activity at the POz electrode will be considered as a potential marker of load. However, researchers should be aware that aggregated EEG features (i.e., alpha power and N2 amplitude) at group level are not sensitive enough for detecting cognitive *overload*. Other features analyzed over time, the use of a longitudinal design and training a statistical load model with data from several individual sessions may be preferred for a more nuanced approach in the exploration of other potential markers for estimating cognitive (over)load.

References

1. ElMaraghy, H., AlGeddawy, T., Azab, A., ElMaraghy, W.: Change in Manufacturing – Research and Industrial Challenges. Keynote Paper, In: ElMaraghy H (editor). Enabling Manufacturing Competitiveness and Economic Sustainability. Proceedings of the 4th International Conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV 2011). pp. 2-9. Springer, Montreal (2011).
2. Kaare, K.K., Otto, T.: Smart Health Care Monitoring Technologies to Improve Employee Performance in Manufacturing. Procedia Engineering 100, 826 – 833 (2015).
3. Qin, J., Liu, Y., & Grosvenor, R.: A categorical framework of manufacturing for industry 4.0 and beyond. Procedia CIRP, 52, 173–178 (2016).
4. Brolin, A., Thorvald, P., Case, K.: Experimental study of cognitive aspects affecting human performance in manual assembly, Prod. Manuf. Res. 5(1), 141–163 (2017).
5. United Nations Industrial Development Organization (UNIDO): Emerging Trends in Global Manufacturing Industries: challenges, opportunities and policy responses (2017).
6. Takayama, L., Ju, W., Nass, C.: Beyond Dirty, Dangerous and Dull: What Everyday People Think Robots Should Do. In Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), Amsterdam, pp. 25–32 (2008).
7. Longo, F., Nicoletti, L., Padovano, A.: Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. Computers & Industrial Engineering 113, 144–159 (2017).
8. Peruzzini, M., Gregori, F., Luzi, A., Mengarelli, M., & Germani, M.: A social life cycle assessment methodology for smart manufacturing: The case of study of a kitchen sink. Journal of Industrial Information Integration, 7 , 24–32 (2017). https://doi.org/10.1016/j.jii.2017.04.001
9. Xu, L. Da, Xu, E. L., & Li, L.: Industry 4.0: state of the art and future trends. International Journal of Production Research , 7543 , 1–22 (2018). https://doi.org/10.1080/00207543.2018.1444806
10. Gevins, A., Smith, M.E.: Neurophysiological measures of cognitive workload during human-computer interaction. Theoretical Issues in Ergonomics Science 4, 1-2, 113–131 (2003).
11. Kosch, T., Funk, M., Schmidt, A., & Chuang, L. L.: Identifying Cognitive Assistance with Mobile Electroencephalography: A Case Study with In-Situ Projections for Manual Assembly. Proceedings of the ACM on Human-Computer Interaction, 2(EICS), 11 (2018).
12. Wickens, C. D.: Mental workload: assessment, prediction and consequences. In L. Longo & M. C. Leva (Eds.), Human mental workload: models and applications: First International Symposium, H-WORKLOAD 2017, Dublin, pp. 18–29 (2017). https://doi.org/10.1007/978-3-319-61061-0_2
13. Young, M., Brookhuis, K., Wickens, C., & Hancock, P.: State of science: mental workload in ergonomics. Ergonomics, 58 (1), 1–17 (2014). https://doi.org/10.1080/00140139.2014.956151
14. Sweller, J.: Cognitive load during problem solving: effects on learning. Cognitive Science, 12 (1), 257–285 (1988). https://doi.org/10.1016/0364-0213(88)90023-7
15. Sweller, J.: Cognitive load theory, learning difficulty, and instructional design. Learning and Instruction, 4 (4), 295–312 (1994). https://doi.org/10.1016/0959-4752(94)90003-5
16. Sweller, J., van Merrienboer, J. J. G., Paas, F.: Cognitive architecture and instructional design. Educational Psychology Review , 10 (3), 25–296 (1998). https://doi.org/10.1023/A:1022193728205
17. Cain, B.: A review of the mental workload literature. In Report No.: RTO-TR-HFM-121-Part-II. (pp. 1–34). Ontario, WT: Defence Research and Development Canada Toronto Human System Integration Section (2007).

18. O'Donnell, R. D., Eggemeier, F. T.: Workload assessment methodology. In K. R. Boff, L. Kaufman, J. P. Thomas (Eds.), Handbook of perception and human performance. Vol. 2. Cognitive processes and performance. (p. 42/1-42/9). Wiley Interscience, New York (1986).
19. Van Acker, B.B., Parmentier, D.D., Vlerick, P., Saldien, J.: Understanding mental workload: from a clarifying concept analysis toward an implementable framework. Cognition, Technology & Work , 20 (3), 351–365 (2018). https://doi.org/10.1007/s10111-018-0481-3
20. Wickens, C. D., Hollands, J. G.: Engineering psychology and human performance (3rd ed.). Upper Saddle River, NJ: Prentice-Hall (2000).
21. Young, M.S., Stanton, N.A.: Mental workload: theory, measurement and application. In W. Karwowski (Ed.), International encyclopedia of ergonomics and human factors (Vol. 1, pp. 507–509). Taylor & Francis, London (2001).
22. Berka, C., Levendowski, D.J., Cvetinovic, M.M, Petrovic, M.M, Davis, G, Lumicao, M.N., et al.: Real-Time Analysis of EEG Indexes of Alertness, Cognition, and Memory Acquired With a Wireless EEG Headset. International Journal of Human-Computer Interaction 17 (2), 151–170 (2004).
23. Gevins, A., Smith M.E.: Electroencephalography (EEG) in neuroergonomics. Neuroergonomics: The Brain at Work, 15–31 (2006).
24. Hogervorst, M.A., Brouwer, A.-M., van Erp, J.B.F.: Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. Front. Neurosci. 8–322 (2014).
25. Gougelet, R.J.: Neural Oscillation Dynamics of Emerging Interest in Neuroergonomics. In: Ayaz, H., Dehais, F. (Eds.), Neuroergonomics: The Brain at Work and Everyday Life (pp. 87-91), Elsevier (2019).
26. Antonenko, P., Paas, F., Grabner, R., van Gog, T.: Using Electroencephalography to Measure Cognitive Load. Educational Psychology Review, 22 (4), 425–438 (2010). https://doi.org/10.1007/s10648-010-9130-y
27. Brouwer, A.M., Zander, T.O., van Erp, J.B.F., Korteling, J.E., Bronkhorst, A.W.: Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. Front Neurosci, 9–136 (2015). doi: 10.3389/fnins.2015.00136.
28. Charles, R.L., Nixon, J.: Measuring mental workload using physiological measures: a systematic review. Appl. Ergon. 74, 221–232 (2019).
29. Basar, E., Basar-Eroglu, C., Karakas, S., & Schürmann, M.: Oscillatory brain theory: a new trend in neuroscience. IEEE Engineering in Medicine and Biology Magazine: The Quarterly Magazine of the Engineering in Medicine & Biology Society, 18 (3), 56–66 (1999). https://doi.org/10.1109/51.765190
30. Luck, S.J.: An Introduction to the Event-Related Potential Technique, 2nd Edn. Cambridge, MA: MIT Press (2014).
31. Brouwer, A.M., Hogervorst, M.A., van Erp, J.B.F., Heffelaar, T., Zimmerman, P.H., Oostenveld, R.: Estimating workload using EEG spectral power and ERPs in the n-back task. Journal of Neural Engineering 9, 4 (2012).
32. Roy, R.N., Charbonnier, S., Campagne, A., Bonnet, S.: Efficient mental workload estimation using task-independent EEG features. Journal of Neural Engineering 13, 2 (2016).
33. Ullsperger, P., Freude, G., Erdmann, U. Auditory probe sensitivity to mental workload changes: an event-related potential study. Int. J. Psychophysiol. 40, 201–209 (2001). doi: 10.1016/S0167-8760(00)00188-4
34. Kramer A.F., Trejo, L.J., Humphrey, D.: Assessment of mental workload with task-irrelevant auditory probes. Biol Psychol. 40, 83-100 (1995)
35. Allison, B.Z., Polich, J.: Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. Biol. Psychol. 77, 277–283 (2008). doi: 10.1016/j.biopsycho.2007.10.014
36. Miller, M.W., Rietschel, J.C., McDonald, C.G., Hatfield, B.D.: A novel approach to the physiological measurement of mental workload. Int. J. Psychophysiol. 80, 75–8 (2011).

37. Stinissen, J.: Revised minnesota paper form board test. Vorm AB - Leuvense aanpassing van vorm MA en MB, Likert, R., Quasha, W. (pp. 1–7). Amsterdam: Swets & Zeitlinger, B.V. (1977)
38. Hart, S. G., Staveland, L. E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In P. A. Hancock, N. Meshkati (Eds.), Human mental workload. (pp. 139–183). Elsevier Science Publishers B.V., Amsterdam (1988).
39. Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., et al.: PsychoPy2: experiments in behavior made easy. Behavior Research Methods (2019). doi: 10.3758/s13428-018-01193-y
40. Gramfort, A., Luessi, M., Larson, E., Engemann, D. Strohmeier, D., Brodbeck, C., et al.: MNE software for processing MEG and EEG data. NeuroImage, 86, pp. 446–460 (2014). https://doi.org/10.1016/j.neuroimage.2013.10.027
41. Gramfort, A., Luessi, M., Larson, E., Engemann, D. Strohmeier, D., Brodbeck, C., et al.: MEG and EEG data analysis with MNE-Python, Front Neurosci, 7, (2013). https://doi.org/10.3389/fnins.2013.00267
42. Welch, P. D.: The Use of Fast Fourier Transform for the Estimation of Power Spectra: A Method Based on Time Averaging Over Short, Modified Periodograms. IEEE Transactions on Audio and Electroacoustics (1967). https://doi.org/10.1109/TAU.1967.1161901
43. Jensen, O., Gelfand, J., Kounios, J., Lisman, J.E.: Oscillations in the Alpha Band (9-12 Hz) Increase with Memory Load during Retention in a Short-term Memory Task. Cerebral Cortex 12, 8 (2002). https://doi.org/10.1093/cercor/12.8.877
44. Klimesch, W.: EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res Brain Res Rev, 29, 169–195 (1999).
45. Matthews, G., Campbell, S. E., Falconer, S., Joyner, L. A., Huggins, J., Gilliland, K., et al.: Fundamental dimensions of subjective state in performance settings: task engagement, distress, and worry. Emotion, 2, 315–340 (2002). https ://doi.org/10.1037/1528-3542.2.4.315
46. Fink, A., Grabner, R.H., Neuper, C., Neubauer, A.C.: EEG alpha band dissociation with increasing task demands. Cognitive Brain Research, 24, 252–259 (2005).
47. Samaha, J., Postle, B.R.: The Speed of Alpha-Band Oscillations Predicts the Temporal Resolution of Visual Perception. Curr. Biol., 25(22), 2985–90 (2015). doi: 10.1016/j.cub.2015.10.007

Towards affective robotics: wearable devices for the assessment of cognitive effort

Massimiliano Righi, Valeria Villani, Lorenzo Sabattini, Cristian Secchi

Authors are with the Department of Sciences and Methods for Engineering (DISMI), University of Modena and Reggio Emilia, Reggio Emilia, Italy
name.surname@unimore.it

Abstract. Cognitive effort has been studied for a long time with the purpose of integrating this measure in human-machine/robot interaction by modulating the system behaviour and simplifying the interaction tasks. To this end, the question is whether standard wearable devices are reliable enough to be adopted in this regard. The purpose of the present paper is to analyze the accuracy in detection of cognitive effort of some commercial wearable devices, such as armbands, wristbands and chest straps. In an experiment setting, thirty participants were exposed to an increase in their cognitive effort by means of some common stressors. Two wearable devices, of the above mentioned categories, were used all together to evaluate a change in the participants' heart activity. The trend of heart rate variability (HRV) reflected the change of subjects' cognitive effort. In particular, we found that the analysis of HRV measured by the chest strap provides the most accurate detection of cognitive effort. Nevertheless, also measurements by the smartwatch are slightly sensitive to cognitive effort.

1 INTRODUCTION

This paper studies the problem of monitoring a subject's cognitive effort while interacting with a complex robotic system. The ultimate aim is that of detecting when the user is overwhelmed, simplifying the interaction task when it gets too complex, accordingly.

Recent advances in robotics have determined the introduction of service robots in everyday life scenarios and the use of collaborative robots working close and together with human operators in industries [20]. In particular, collaboration between workers and collaborative robots on the factory shop is becoming the new frontier in industrial robotics [12]. This has the positive outcome that human capabilities and skills are enhanced and complemented by those of the robot. However, as a drawback, it implies that non-expert users (i.e., users without expertise focused on robotics) are requested to use complex systems. In order to guarantee an effective use of robots in such contexts, it is, then, of paramount importance that a smooth interaction is enabled that allows ease of use, efficiency and user's satisfaction.

To this end, approaches based on affective robotics have been proposed. Affective robotics consists in enhancing the interaction of a human with a robot by recognizing her/his affect. Monitoring and interpreting nonverbal communication can provide important insights about a human interacting with the robot and, thus, implicit feedback about the interaction can be achieved. Accordingly, the aim of affective robotics is relieving users cognitive burden when the task to accomplish overloads her/his mental capabilities, adapting the behaviour of the robot and implementing a sufficient level of autonomy. While approaches based on affective robotics are largely considered for socially interacting robots, they are still unexplored in the case of service and industrial robots, where the interaction task and the environmental conditions, such as noise, stress, time constraints, overload the cognitive and emotional burder of the user. In this context, methodologies for affective interaction with service and industrial robots are needed. These aim at increasing the safety and efficiency of the interaction system by tuning the level of autonomy of the robotic system, in order to assist the human operator. Furthermore, it would be useful to achieve this goal by using non-dedicated multi-purpose devices, such as smartwatches or wristbands for activity tracking, which are easy to be worn and used by the operator and do not interfere with the interaction task.

Moving along these lines, in this paper, we address the problem of monitoring a subject's cognitive workload in approaches of affective robotics. We aim at detecting cognitive workload by means of quantitative measurement of subject's physiological parameters. In particular, we consider the assessment of user's cognitive effort by means of commercial non-invasive wearable devices that do not limit the freedom of movement, and, hence, could be truly used in real operative scenarios. Specifically, cognitive effort is estimated from the analysis of cardiac activity, measured by a smartwatch and a chest strap.

The results of this study have been applied to the industrial use case addressed in the COMPLEMANT experiment [1], which is part of the EU H2020 HORSE project [2]. The use case refers to a robotized injection moulding production activity, characterized by such a high production pace that forces the operator to work under an external, and very fast, pace determinant with consequent effects on the cognitive demand and on the quality of the output.

The paper is organized as follows. In Sec. 2 we discuss the background of cognitive workload estimation by means of the analysis of cardiac activity and related works with respect to the use wearable devices for this purpose. Therein, the contribution of this paper with respect to the state of the art is highlighted. Then, in Sec. 3 the experimental analysis we performed is introduced and the achieved results are discussed in Sec. 4. Finally, Sec. 5 follows with some concluding remarks.

2 BACKGROUND AND CONTRIBUTION

Physiological signals have been widely used to estimate human affective state, focus, attention and intent. Among them, hear rate (HR) is the most suited to be considered in real operational environments since it can easily be measured by commercial portable and non-invasive wearable devices, such as smartwatches and wristbands or chest straps for activity tracking.

From HR, heart rate variability (HRV), which is the variation over time of the interval between consecutive heart beats, can be derived. It is an established quantitative index for the non-invasive assessment of autonomic nervous system function. It has been widely shown that cognitive processing influences HRV [3, 9, 15]. The effect of stress on HRV is due to the fact that cognitive effort is one of the factors contributing to sympathetic stimulation, which is associated with the low frequency range of HR.

To quantify HRV, several metrics are computed from the analysis of RR interval time series [4, 17]. Such metrics are typically extracted from the time and frequency domain. In particular, the most common statistical time domain metrics are: the mean value and the standard deviation, denoted by *mean RR* and *SDRR* in the following, 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*). As regards the frequency domain metrics, the most used ones are the power in the low frequency band (*LF*, $0.04 - 0.15Hz$), the power in the high frequency band (*HF*, $0.15 - 0.40Hz$) and their ratio (*LF/HF ratio*).

Unfortunately, RR interval time series are seldom provided as output by commercial wearable devices: HR, expressed in terms of averaged number of beats per minute (BPM), is the only output of most inexpensive devices for the recording of cardiac activity. Some studies in the literature have proved that HR is sufficient to discriminate a variation in cardiac activity during moderate and intense cognitive load [6, 8, 10]. However, information about cardiac activity is largely lost in the BPM measure due to the moving window averaging that lies under its construction.

To this end, in this paper we aim at assessing the accuracy in detecting the physiological response to cognitive effort from HRV analysis through the use of wearable commercial devices. In particular, we consider a chest strap and a smartwatch in order to a assess how the recording location (chest and wrist) and the technology (electrodes or photoplethysmography-derived) of measurements affects the accuracy of cognitive effort detection.

3 MATERIALS AND METHODS

The aim of our experiments was to assess how accurately cognitive effort can be detected when using commercial wearable devices worn at the chest and wrist. To this end, test participants were exposed to cognitive effort and their cardiac activity was recorded and analyzed.

3.1 Test subjects

A total of $N = 30$ users (8 females, 22 males, age 26.58 ± 5.11 y.o.) were enrolled in the experiment. Participants are researchers working at our engineering department, in different research fields. All of them were completely new to the experimental task and goals. Compliance to participate in the study was obtained from written informed consent during the description of the experiment. All the data were analyzed and reported anonymously. No participant reported to have any cardiovascular problems that may have influenced the procedures carried out.

3.2 Measurement devices

We used the following devices: a Polar H10 (Polar Electro Inc., Bethpage, NY, USA) as chest strap and a Samsung Gear S1 (Samsung, South Korea) as a smartwatch.

Polar H10 is an electrode based chest strap that replaced the discontinued H7. These chest straps have been reported as trustworthy devices to detect BPM and RR data [6,10,11]. The smartwatch uses photoplethysmography technology, but, being a wrist-worn device, is more susceptible to motion artifacts [16].

The two devices were used simultaneously during the experimental sessions. The chest strap was connected through Bluetooth to a laptop PC running Ubuntu 16.04.5, while the smartwatch was connected over Wi-Fi socket to the same PC.

3.3 Test protocol

The tests were conducted in a laboratory environment in which light and temperature were kept constant during the entire test. Participants were asked to seat on a height-adjustable chair in front of a 19" monitor's screen. Before recordings, participants were asked to wear the two sensors previously sanitized with alcohol. The Polar H10 was applied with the addition of an ECG gel on the strap as suggested on the manual. After sensors placement, participants were introduced to the test by reading the instructions that described the types of quizzes at which they where asked to respond. The experimenter was seated next to the participants to help them with the instructions.

The experimental protocol consisted of two phases, during which participants were exposed to rest and cognitive effort, for an overall duration of 12 minutes. In the following, the two conditions are denoted by rest and stress, for ease of notation. During the rest phase, each participant was asked to relax while data were recorded. Then, the study design implied that the stress phase had both a cognitive aspect (mental effort) and an emotionally stressful aspect (annoying music, presence of the experimenter and performance related stress elicited by the fact that the answers collected to establish a ranking among participants). To induce cognitive effort, a combination of memory tasks, mathematics tasks and visual tasks were used[1]: Stroop test, fast counting, math calculation and the 2-back test. These stressors are well known in the literature and customarily used for assessment of cognitive effort [3,14,15,18]. Quizzes were presented in the order depicted in Fig. 1. Every session was presented for one minute including 3 seconds of interval between each one.

After the presentation of quizzes, at the end of the stress phase, participants were asked to answer the NASA Task Load Index [7]. The questionnaire investigated subject's perception of cognitive effort, physical fatigue, temporal demand, performance, effort and frustration during the test, in a rating scale from 1 to 10.

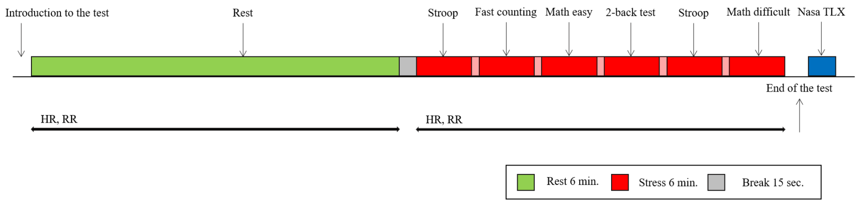


Fig. 1: Experimental protocol. RR were acquired during both rest and stress phases. The stress phase consisted of six different quizzes organized as depicted.

3.4 Data processing

Firstly, raw HRV data were pre-processed using the detrending tool presented in [19]. Secondly, each RR series was visually inspected to verify the absence of acquisition noise or by ectopic beats. Ectopic beats were replaced with interpolated intervals (five intervals) calculated between the previous and next validated RR intervals [5]. Following the visual inspection of the obtained RR series, some participants' data were discarded due to the presence of movement artifacts caused by the worn location. For the data coming from Samsung Gear S1, 6/30 participants were removed from the analysis while, for Polar H10, 2/30 participants were discarded. Participants were not asked to repeat the test to avoid familiarization with the test itself.

As regards HRV analysis, the Welch's periodogram was considered for frequency domain parameters.

3.5 Statistical analysis

Recorded data were analyzed using MATLAB Statistics and Machine Learning Toolbox. The tools presented in [13, 21] were also used.

[1] A video of the tests presented to study participants can be found at `https://drive.google.com/file/d/1d2pPdWYV1icRLV9ejYRklpPUCdkfWgYv/view?usp=sharing`.

First, normal distribution and homogeneity of variance was assessed with Fisher's test. In order to investigate whether subjects response during the stress phase differs from the one during the rest phase we ran a paired-samples t-test.

Moreover, one-way multivariate analysis of variance (ANOVA) was considered to explore possible between-subject differences on the cardiovascular metrics of HRV (dependent variables) between rest and different stages of stress (i.e., after 3 minutes and at the end of the stress phase) (independent variables).

Possible sex differences in cardiovascular activity were not considered. Bonferroni's Post-hoc-test was used to explore further significant differences between the two phases. The significance level was set at $p < 0.05$.

4 RESULTS AND DISCUSSION

The answers given by the study participants to the NASA TLX questionnaire confirmed that the devised test shown in Fig. 1 succeeded in eliciting cognitive effort. Specifically, the arithmetic mean of the NASA TLX answers showed high values in a scale from 0 to 10 for all those questions concerning cognitive effort of the experiment: mental demand 7.10, temporal demand 7.43, effort 7.13. Physical fatigue was rated low (average value 3.37), as expected, whereas subjective perception of performance was rated 5.67 on average.

Building upon this result, in the following we report the results of statistical analyses of RR data, aimed at detecting the different conditions of cognitive effort.

4.1 HRV analysis

The general trend of the RR intervals for the smartwatch Samsung Gear S1 and the chest strap Polar H10 is depicted in Fig. 2.

Results of the HRV analysis are reported in Table 1 and in Table 2. In particular, Table 1 reports the results of the ANOVA considering three conditions: rest (R), the first 3 minutes of the stress phase (S1) and the second 3 minutes of the stress phase (S2). In Table 2 the results of the t-test regarding the comparison between the rest (R) and the whole stress (S) phases are reported.

Most of the cardiovascular metrics we considered exhibited a trend between the rest and stress phases that is in accordance with the literature review presented in [3]. Specifically, in the case of the chest strap, all the time domain metrics, namely *mean RR*, *SDNN*, *rMSSD*, *pNN50*, decreased when stress was increased. This trend was not confirmed for some metrics of the smartwatch: both *rMSSD* and *SDNN* increased with stress. Regarding the frequency domain metrics it can be noted that in general *HF*, *LF* and *LF/HF* decreased during stress. This behavior is consistent with that reported in [3] for *HF* while it disagrees with the trend expected for *LF* and *LF/HF* (an increase in values). However, as argued in [3] there is not a generalized consensus about the trends of these two metrics. For example, 3 out of 8 studies considered in [3] reported a decreasing value of *LF* during stress, according with our trend. This difference was

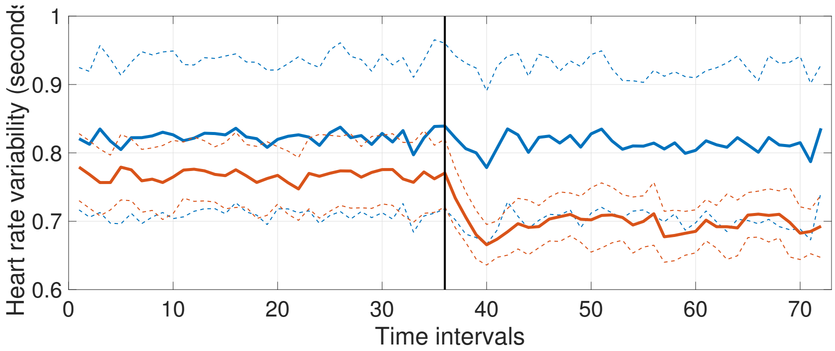


Fig. 2: Simultaneous RR acquisition during the entire test (12 mins) (smartwatch Samsung Gear S1 in blue and chest strap Polar H10 in red). Data are averaged over all the study participants, considering intervals of 10 s duration. Standard deviation is shown in thin dashed lines.

attributed in [3] to the fact that the type of tasks required the user to interact with the mouse; this may have activated a cortical area (associated with physical activity) causing a shift in the autonomic nervous system, which is reflected in a change in the balance between both sympathetic and parasympathetic nervous systems. As a consequence the ratio (LF/HF) was also affected.

The one-way ANOVA with the cardiovascular metrics for HRV analysis as the dependent variables showed some significant changes across conditions from rest (R) to stress (S1 and S2) phases. Mean and standard deviation for the HRV metrics are reported in Table 1. As regards the chest strap Polar H10, the statistical analysis revealed a significant difference between the three phases: $SDNN$: $F(2, 81) = 4.422, p = .015$, $RMSSD$: $F(2, 81) = 4.072, p = .021$, LF: $F(2, 81) = 7.159, p = .001$. Furthermore, the pairwise comparison with the Bonferroni corrected Post-hoc-test revealed a significant difference for the $SDNN$ between R and S1 ($p_{R,S1} = .043$) and R and S2 ($p_{R,S2} = .029$), for the $RMSSD$ between R and S1 ($p_{R,S1} = .046$) and R and S2 ($p_{R,S2} = .047$) and for LF between R and S1 ($p_{R,S1} = .002$) and R and S2 ($p_{R,S2} = .009$). Statistical analysis for the smartwatch Samsung Gear S1 revealed no significant differences. Despite this result, a further Post-hoc-test with Bonferroni correction showed a significant difference for the LF between R and S1 ($p_{R,S1} = .039$) that was not reflected in S2.

Beside the one-way ANOVA, this result was confirmed also by the paired-samples t-test. In Table 2 we report the significant differences between the distribution of different metrics. The analysis of the HRV metrics for the Polar H10 showed a significant difference for all the studied cardiovascular metrics ($p < .001$) except for the LF/HF ratio. On the contrary, no significant difference was found for the Samsung Gear S1.

Table 1: ANOVA - Mean values (standard deviation) of different HRV metrics obtained from the chest strap Polar H10 (top) and from the smartwatch the Samsung Gear S1 (bottom) along the three conditions, rest (R), first part of stress (S1) and second part of stress (S2) (independent values).

	Chest strap		
Mean (SD)	R	S1	S2
Mean RR (ms)	766.97(149.88)	697.67(141.12)	694.68(140.05)
SDNN (ms)	60.65(17.72)	49.02(17.08)$*$	48.34(17.43)$*$
rMSSD (ms)	48.60(21.70)	36.38(16.81)$*$	36.44(16.37)$*$
pNN50 (%)	23.02(15.88)	15.44(13.01)	15.24(12.16)
LF/HF	3.73(1.85)	3.54(2.59)	3.97(3.69)
LF (ms^2)	1881.33(927.93)	977.18(916.19)$*$	1085.67(1076.62)$*$
HF (ms^2)	776.23(801.78)	416.79(555.17)	458.28(644.79)

	Smartwatch		
Mean (SD)	R	S1	S2
Mean RR (ms)	818.35(114.91)	809.12(144.34)	805.77(145.06)
SDNN (ms)	119.61(23.17)	120.58(33.06)	120.22(34.51)
rMSSD (ms)	139.88(27.98)	144.47(43.85)	140.52(42.51)
pNN50 (%)	68.70(5.36)	68.82(8.69)	66.90(7.71)
LF/HF	0.90(0.28)	0.71(0.32)	0.99(0.69)
LF (ms^2)	7559(12078.11)	4031.29(2473.53)$*$	4881.65(3387.19)
HF (ms^2)	8738.23(12792.08)	6800.88(4270.98)	5931.03(3302.72)

$*$ Significantly different from R ($p < 0.05$)
Significantly different from S1 ($p < 0.05$)

Table 2: t-TEST - Mean values (standard deviation) of different HRV metrics obtained from the chest strap the Polar H10 (top) and from the smartwatch Samsung Gear S1 (bottom) along the two phases, rest (R) and stress (S).

	Chest strap		
Mean (SD)	R	S	p
Mean RR (ms)	766.97(149.88)	696.17(140.57)	$p < 0.01$
SDNN (ms)	60.65(17.72)	49.06(16.36)	$p < 0.01$
rMSSD (ms)	48.60(21.7)	36.70(16.08)	$p < 0.01$
pNN50 (%)	23.02(15.88)	15.34(12.31)	$p < 0.01$
LF/HF	3.73(1.85)	3.46(2.56)	NS
LF (ms^2)	1881.33(927.93)	1096.30(942.00)	$p < 0.01$
HF (ms^2)	776.23(801.78)	454.11(547.14)	$p < 0.01$

	Smartwatch		
Mean (SD)	R	S	p
Mean RR (ms)	818.35(114.91)	807.40(144.59)	NS
SDNN (ms)	119.61(23.17)	120.73(31.88)	NS
rMSSD (ms)	139.88(27.98)	142.51(41.65)	NS
pNN50 (%)	68.70(5.36)	67.82(7.81)	NS
LF/HF	0.90(0.28)	0.88(0.46)	NS
LF (ms^2)	7559.00(12078.11)	4595.41(2490.94)	NS
HF (ms^2)	8738.23(12792.08)	6211.15(3584.39)	NS

4.2 Discussion

The results presented above have shown that the smartwatch is not robust enough to artifacts and, hence, cannot detect the occurrence of cognitive effort. On the contrary, the chest strap proved capable to discriminate the two main phases of our experiment (R and S), although it failed to discriminate between the sub-phases S1 and S2. This result is confirmed by other studies involving the use of the same chest strap, for example [10] where participants were asked to perform a mental arithmetic task. The chest strap was capable to discriminate between the rest phase and the stress phase. However, it is worthwhile noting that the metrics that showed significant difference are different to the ones of the present study (i.e., *SDNN* and *RMSSD* were not significant).

As regards the smartwatch, the results reported in Tables 1 and 2 are in contrast with our previous finding reported in [18], where a statistically significant difference was found, on the t-test, for HRV analysis on *mean RR* values between the two conditions of rest and stress. We can argue that this result might be due to the movement artifacts caused by the body location where the device is worn. In [18] the smartwatch was accurately put on by the experimenter avoiding areas close to bones and other tissues, while, in this study, participants were asked to wear the device by themselves, in order to simulate a more realistic procedure. Moreover, the accuracy of the photoplethysmography sensor readings strictly depend on how tight is the smartwatch bracelet.

5 CONCLUSIONS

In this paper we considered the problem of estimating subject's cognitive effort by means of commercial non-invasive wearable devices that measure cardiac activity. The ultimate goal is that exploiting information about user's status when interacting with a robot, in order to adapt the behaviour of the robot when it becomes too complex for the user. In this regard, the aim of the study presented in this paper was to investigate the accuracy of different wearable devices to discriminate between different levels of cognitive effort compared to a baseline. Moreover, we were interested in assessing how, both, the body location and the technology of these devices may have an influence on the detection of the HRV on subjects involved in cognitive tasks. A commercial smartwatch and a commercial chest strap were compared to this end.

The results suggest that the adopted mental tasks successfully induced a cognitive load on participants with a variation of their cardiac status during the experiment. The chest band Polar H10 was able to detect this change: a much greater accuracy, compared to the smartwatch, was shown by the chest band, as expected, given the fact that it relies on a more reliable measurement technology and it is intrinsically less sensitive to motion artifacts.

As future studies, we will assess whether HR information, i.e., averaged BPM data, is a reliable estimator for subject's cognitive effort. Indeed, this information is provided by commercial wearable devices more frequently than RR time interval series. Moreover, we are interested in using these devices for on-line detection of mental fatigue during interaction tasks, in order to use them in applications of affective robotics.

ACKNOWLEDGMENT

The research is carried out within the Experiment "COllaborative robot aMPLifying and Extending huMAN capabiliTies" (COMPLEMANT), which is part of the "Smart integrated Robotics system for SMEs controlled by Internet of Things based on dynamic manufacturing processes" (HORSE) project, funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 680734.

References

1. COllaborative robot aMPLifying and Extending huMAN capabiliTies - (COMPLEMANT), http://www.horse-project.eu/COMPLEMANT
2. Smart integrated Robotics system for SMEs controlled by Internet of Things based on dynamic manufacturing processes - (HORSE), http://www.horse-project.eu
3. Castaldo, R., Melillo, P., Bracale, U., Caserta, M., Triassi, M., Pecchia, L.: Acute mental stress assessment via short term HRV analysis in healthy adults: A systematic review with meta-analysis. Biomedical Signal Process. Control 18, 370–377 (2015)
4. Clifford, G.D., Azuaje, F., McSharry, P. (eds.): Advanced Methods and Tools for ECG Data Analysis. Artech House, Inc. (2006)
5. Clifford, G.D., Tarassenko, L.: Quantifying errors in spectral estimates of HRV due to beat replacement and resampling 52(4), 630–638 (2005)
6. Gillinov, S., Etiwy, M., Wang, R., Blackburn, G., Phelan, D., Gillinov, A.M., Houghtaling, P., Javadikasgari, H., Desai, M.Y.: Variable accuracy of wearable heart rate monitors during aerobic exercise. Medicine and Science in Sports and Exercise 49(8), 1697–1703 (aug 2017)
7. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Advances in Psychology, vol. 52, pp. 139–183 (1988)
8. Hettiarachchi, I., Hanoun, S., Nahavandi, D., Iskander, J., Hossny, M., Nahavandi, S.: 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. pp. 3045–3050. IEEE (oct 2019)
9. Kulic, D., Croft, E.A.: Affective state estimation for human–robot interaction. IEEE Trans. Robotics 23(5), 991–1000 (2007)
10. Mishra, V., Pope, G., Lord, S., Lewia, S., Lowens, B., Caine, K., Sen, S., Halter, R., Kotz, D.: 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. pp. 1717–1728. ACM Press, New York, New York, USA (2018)
11. Plews, D.J., Scott, B., Altini, M., Wood, M., Kilding, A.E., Laursen, P.B.: Comparison of heart-rate-variability recording with smartphone photoplethysmography, polar H7 chest strap, and electrocardiography. International Journal of Sports Physiology and Performance 12(10), 1324–1328 (nov 2017)

12. Probst, L., Frideres, L., Pedersen, B., Caputi, C.: Service innovation for smart industry: human–robot collaboration. Tech. rep., Publications Office of the European Union, Luxembourg (2015)
13. Ramshur, J.T.: Design, evaluation, and application of heart rate variability analysis software (HRVAS). Ph.D. thesis, University of Memphis (2010)
14. Rani, P., Sarkar, N., Smith, C.A., Kirby, L.D.: Anxiety detecting robotic system - towards implicit human-robot collaboration. Robotica 22(1), 85–95 (2004)
15. Rani, P., Sims, J., Brackin, R., Sarkar, N.: Online stress detection using psychophysiological signals for implicit human-robot cooperation. Robotica 20(06), 673–685 (2002)
16. Spierer, D.K., Rosen, Z., Litman, L.L., Fujii, K.: Validation of photoplethysmography as a method to detect heart rate during rest and exercise. Journal of Medical Engineering and Technology 39(5), 264–271 (jul 2015)
17. 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. 17, 354–381 (1996)
18. Villani, V., Sabattini, L., Secchi, C., Fantuzzi, C.: A framework for affect-based natural human-robot interaction. In: IEEE (ed.) 27th IEEE Int. Symp. Robot and Human Interactive Communication (RO-MAN) (2018)
19. Villani, V., Fasano, A.: Fast detrending of unevenly sampled series with application to hrv. In: Computing in Cardiology 2013. pp. 417–420. IEEE (2013)
20. Villani, V., Pini, F., Leali, F., Secchi, C.: Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications. Mechatronics 55, 248–266 (2018)
21. Vollmer, M.: A robust, simple and reliable measure of heart rate variability using relative rr intervals. In: 2015 Computing in Cardiology Conference (CinC). pp. 609–612. IEEE (2015)

Subjective mental load of dismounted soldiers as a function of the procedures they follow

Julie Albentosa[1], Alexis Remigereau[1], Françoise Darses[1]

[1] French armed forces biomedical research institute, Neurosciences and cognitive sciences Department, B.P 73,
91220 Brétigny-sur-Orge, France
julie.albentosa@intradef.gouv.fr, alexis.remigereau@hotmail.fr,
francoise.darses@intradef.gouv.fr

Abstract. In the land forces, dismounted soldiers, in particular squad and platoon leaders, are often subject to significant subjective mental load as they have to manage their subordinates in dynamic and risky situations. However, during their training, they are taught procedures to help limit the mental load during a mission. The aim of this study was to assess the subjective mental load associated with each of these procedures. Twenty participants (10 squad leaders and 10 platoon leaders) were asked to evaluate mental load on a 6-point Likert-type scale (0 to 5) for each of 21 established procedures. The analysis found differences for three procedures. Knowing when a procedure becomes more mentally costly could help researchers to identify equipment that could help both squad and platoon leaders carry out their missions.

Keywords: subjective mental load, dismounted soldiers, platoon leader, squad leader.

1 Introduction

The missions undertaken by French infantry soldiers, in particular platoon and squad leaders, are highly complex. A platoon leader has to manage more subordinates than a squad leader (23 to 30, compared to 6 or 7, depending on the mission), and has to develop higher-level tactical maneuvers. A platoon leader commands three or four squad leaders and is responsible for the whole section, while a squad leader only commands two fireteams. Nevertheless, at each level, these soldiers have to communicate with their hierarchy, manage their subordinates, develop tactical maneuvers, make decisions, etc. The mental load can increase rapidly and potentially become excessive. Mental load is a function of task characteristics and individual abilities [1] [2] [4]. Although the tasks carried out by platoon leaders can be assumed to be more demanding than those of squad leaders, this group should be better-able to cope, due to their knowledge and training.

The French army has put in place measures that are designed to help. During their training, soldiers practice procedures known as REFLEX ACTIONS (e.g., *moving forward*), ELEMENTARY ACTIONS (e.g., *positioning oneself*), and ORDER FRAMEWORK (e.g., *DPIF* for Direction, Point to be reached, Itinerary, Formation) until they become automatic. These standard procedures are written in an official doctrine manual and are usually followed in real combat situations. According to the model of Meister [6], in low and high demanding situations, the level of workload is high. Indeed, in low demanding situations, the individual has to make an effort to struggle against his decrease of vigilance, while in high demanding situations the effort is made to process all the information. In moderately situations, compensatory strategies are set up to decrease the level of workload. We therefore assume that during a military mission, the application of the standard procedures could be a strategy to decrease the mental load. However, in a complex and difficult situation, under significant time pressure and exposed to imminent danger, are these procedures still effective?

To the best of our knowledge, the mental load of platoon and squad leaders has been little studied in the literature, although it could help to improve training programs by focusing on highly-demanding tasks. One of the few examples is an American study of platoon and squad leaders. Participants had an exercise that consisted of attacking and securing a command and control installation near to a village. They were then asked to rate their subjective mental effort for various subscales of the MARS scale (Mission Awareness Rating Scale) [5], which are *identifying*, *understanding*, *predicting*, and *deciding.* The study found that *identifying* required a similar mental effort for platoon and squad leaders, while *understanding*, *predicting* and *deciding* required more effort for squad leaders than platoon leaders (Matthews & Beal, 2002). The authors argued that platoon leaders had a broader picture of the mission. However, this study only focused on mental effort, which is a voluntary process that is under the control of the individual [1] [3].

The present paper is an exploratory study which aimed to identify the mental load that squad and platoon leaders associate with various procedures, and understand the reasons why some differences could appear. Our hypothesis is that, whatever the procedure, the mental load should be relatively similar for platoon and squad leaders, as their training should have equipped them in the same way. However, in nonroutine and difficult situations, following certain procedures will be more mentally costly for either squad leaders or platoon leaders. The results of this exploratory study will help us to build scenarios of combat situations that should bring the most costly procedures and therefore a high mental load.

2 Methods

2.1 Participants

Twenty soldiers from several French infantry regiments participated: ten platoon leaders (PL) and ten squad leaders (SL). They were all men with an average age of 30.65 years (PL: M = 33.50, SD = 5.87; SL: M = 27.80, SD = 4.92), average experience in the Army of 8.58 years (PL: M = 10.86, SD = 5.89; SL: M = 6.30, SD = 4.24), and average experience in their current position of 2.47 years (PL: M = 2.09, SD = 1.20; SL: M = 2.85, SD = 0.91). Once the heads of the infantry regiments agreed to their participation, they asked squad and platoon leaders to participate and the volunteers have been retained. All participants signed an informed consent form and prior permission was given by the head of each infantry regiment.

2.2 Materials and procedure

Before individual interviews in a quiet office at their base, a pre-briefing was done in order to explain that the aim of the study was to assess the mental load of squad and platoon leaders in their missions in order to identify the difficulties that they could have and to find solutions to help them. The definition of mental load was explained to them in scientific and in common language. During the interviews, participants were asked to associate a subjective score for mental load on a 6-point Likert-type scale (ranging from 0 to 5) for each of the following 21 procedures they had been trained to apply during a mission as a dismounted soldier: 1) reflex actions: *orienting oneself*, *observing*, *moving forward*, *communicating*, *protecting oneself*, *camouflaging* oneself, *estimating a distance*, *designating a target*, *preparing the weapon*, *reporting*, *maintaining contact*; 2) elementary actions: *moving*, *positioning oneself*, *using his weapon*, 3) order framework: *MOIAP* (Mission, Objective, Itinerary, Action to take, Place of the squad leader), *DPIF* and , *FAFS* (in Front of this direction, At a certain location, Formation, Stop), *IMA* (Installation area, Mission, Action to take), *IMAPAP* (Installation area, Mission, Area of oversight and fire, Particular points, Action to take, Place of the squad leader), *COFF/RCOFF* (Consumption, Objective, Flow, Fire/Rear sight, Consumption, Objective, Flow, Fire), and *KDNOF* (Kind, Distance, Number and kind of grenade, Objective, Fire). It should be noted that the study was carried out in France, and that these acronyms do not exist in the armies of Anglophone countries. Participants were also asked to explain why they attributed their scores. A thematic content analysis was realized on these justifications that have been classified in two main themes drawn by the results: endogenous factors and exogenous factors.

2.3 Statistical analyses

Parametric analyses were applied to the data. A general linear mixed effects ANOVA model tested the main effect of position; the 21 procedures; and the interaction between position and procedure, with respect to subjective mental load. *Post-hoc* analyses were followed by Fisher's Least Significant Difference test.

When subjective mental load scores significantly differed as a function of position and procedure, Chi-square tests were applied on the number of scores justifications, for each concerned procedure. When the Chi-square test was significant, the z-score contribution of each justification was calculated.

For all analyses, significance was set at $p \leq .05$.

3 Results

3.1 Effects of position and procedures on subjective mental load

Position had no significant impact on subjective mental load (F(1,10) = 0.14, *n.s*). However, a significant effect was found for procedure (F(20,200) = 2.56, $p < .001$) and the interaction between position and procedure (F(20,200) = 1.70, $p < .05$) on subjective mental load. Post-hoc comparisons of subjective mental load for platoon and squad leaders for individual procedures revealed that scores were significantly higher for squad leaders than platoon leaders for DPIF (respectively: *M* = 1.90, *SD* = 1.29; *M* = 1.13, *SD* = 1.55) and estimating a distance (respectively: *M* = 2.10, *SD* = 1.10; *M* = 1.10, *SD* = 1.07), and a marginal effect was found for protecting oneself with higer scores for platoon leaders than squad leaders (respectively: *M* = 2.80, *SD* = 1.70; *M* = 1.60, *SD* = 0.52). For brevity, non-significant results are not given here (see Figure 1).

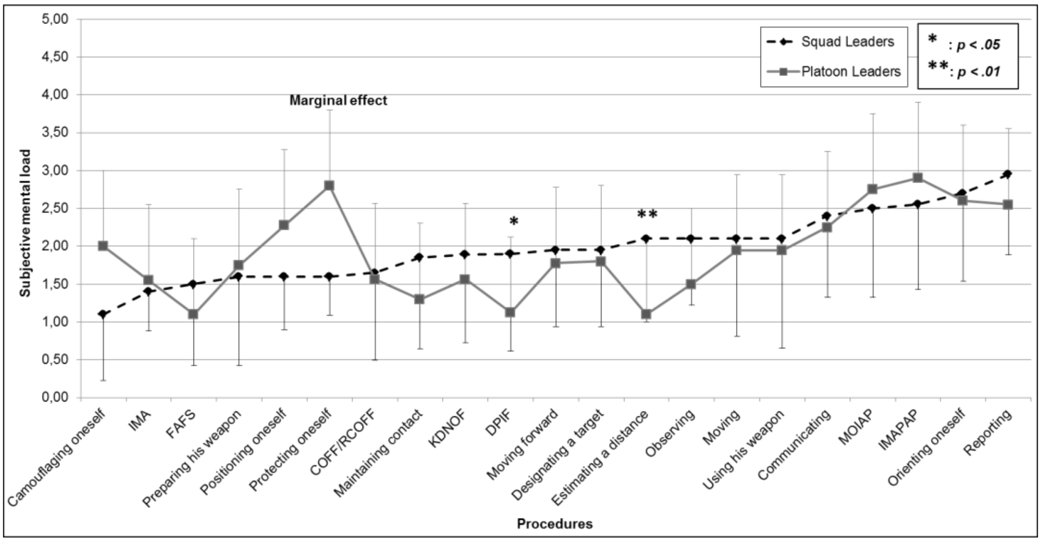


Fig. 1. Subjective mental load (Mean ± Error-Type) for squad and platoon leaders, for each procedure.

3.2 Justifications of subjective mental load scores

As the subjective mental load significantly varied between platoon and squad leaders for *DPIF*, *estimating a distance* and *protecting oneself*, the scores justifications were analyzed for these three procedures. The evocated themes of justifications were divided into three endogenous factors and five exogenous factors, which are respectively: i) level of experience, skills mobilization, confidence in equipment, ii) equipment help, failure or lack of equipment, lack of grenadiers' skills, problematic environment, helping environment.

For *DPIF*, both platoon and squad leaders significantly more justified their subjective mental load by their skills mobilization than by the other types of justification (PL: z-score = 2.84, $p < .01$; SL: z-score = 2.33, $p < .05$).

For *estimating a distance*, there was no justification significantly more expressed than another one. For platoon leaders, they equally justified their mental load by their level of experience, skills mobilization, and equipment help. For squad leaders, they equally justified their mental load by all the justifications except from "confidence in equipment".

For *protecting oneself*, platoon leaders more justified their mental load by "problematic environment" than by the other types of justification (marginal effect: z-score = 1.90, $p = .07$). Squad leaders significantly more justified their mental load by their skills mobilization than by the other types of justification (z-score = 2.41, $p < .05$).

4 Discussion

The results highlight that, overall, subjective mental load was mainly similar for squad and platoon leaders, while all scores were relatively low. This finding could be explained by their experience. Soldiers who have risen through the ranks have gained extensive experience. During their training they have automated most of the procedures they have to follow during a mission. Those who are direct entrants, and lack prior military experience are likely to have the intellectual capacity to apply procedures without making a high mental effort. Furthermore, although platoon leaders have to manage more soldiers than squad leaders, the latter have to give more detailed orders.

It is interesting to note that for *protecting oneself*, platoon leaders tended to have a higher subjective mental load than squad leaders. This could be due to the fact that platoon leaders have more soldiers under their responsibility and they have to think about how to manage the situation, with many tasks to perform. They justified their level of subjective mental load by a problematic environment while squad leaders justified their mental load by their skills mobilization, suggesting that protecting oneself requires a personal involvement. This is interesting to note that platoon leaders gave an external cause while squad leaders gave an internal cause. This refers to the attributional theory of achievement motivation and emotion [7] which describes a causal attribution bias, meaning a trend to explain the causes of our success by internal characteristics (e.g., I am clever) and the causes of our failure by external characteristics (e.g., the task was too hard). Squad leaders reported a higher mental load than platoon leaders for *estimating a distance* and giving the *DPIF* order. Typically, these two procedures are used more by squad leaders, and this group is more likely to use them in complex and difficult situations. For *DPIF*, both of them justified their mental load by their skills mobilization, probably due to the necessity of having a reflection to prepare the order framework without forgetting anything and by being as brief as possible. This is probably more difficult for squad leaders who often have to give this order in complex situations. Indeed, we know that high demanding situations enhance the level of mental load [6]. Furthermore, with respect to *estimating a distance*, platoon leaders are equipped with long-range multifunction binoculars that can display the exact distance to a far-off target, while squad leaders are only equipped with medium-range, infrared binoculars. For this procedure, platoon leaders justified their mental load by two endogenous factors (their level of experience, skills mobilization), and one extraneous factor (equipment help), while squad leaders justified their mental load by two endogenous factors (their level of experience, skills mobilization), and five extraneous factors (equipment help, failure or lack of equipment, lack of grenadiers' skills, problematic environment, helping environment). Again, the attributional theory of achievement motivation and emotion [7] could explain these results.

Although the results are interesting, there are some limitations. First, the study only focuses on mandatory procedures while, in practice, soldiers undertake many other tasks. It would be interesting to evaluate the subjective mental load associated with these other tasks but this is challenging because, for the analyses to be robust, there needs to be a sufficient number of participants who carry out the same tasks. Moreover, subjective mental load scores are highly dependent on the situation in which procedures are executed. For instance, orienting oneself in a calm situation is very different to orienting oneself in a complex and dynamic situation, potentially under fire and with other tasks that must also be completed. A further limitation concerns social desirability. Participants may underestimate their scores in order to maintain a good self-image. It seems likely that this phenomenon is widespread among military personnel. Nevertheless, the analysis makes it possible to compare the behavior of groups within the army, in other words, it highlights if, by nature and independent of the situation, mental loads differ.

Finally, the results of this exploratory study allow targeting the building of simulation scenarios on the elements that can lead to several procedures and particularly to the most mentally costly ones, i.e. *estimating a distance* and giving the *DPIF* order for squad leaders, and *protecting oneself* for platoon leaders. During the simulation scenarios, it would be interesting to objectively assess mental load while actually executing procedures using comprehensive measures (i.e. subjective, objective and physiological) and performance. Physiological measurements could, for instance, identify whether or not mental load is, in fact, underestimated, depending on the procedure. Moreover, these results would also help instructors to target the training on the procedures that are found to be difficult to follow in certain situations.

Acknowledgments. This study was supported by project HUM-1-1009 from the French Directorate General of Armaments. We wish to thank all the contributors and their staff for their participation. We thank the platoon and squad leaders who participated in the study and the heads of the infantry regiments who agreed to their participation. We also thank Commandants Mathieu Catalan and Yvan Demumieux, respectively from the land staff of the army and from the Technical Section of the French Army, who helped us to understand how operations are programmed and how the equipment that is issued is used.

References

1. Brookhuis, K. A., & De Waard, D. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention*, *42*(3), 898-903.
2. Chanquoy, L., Tricot, A., & Sweller, J. (2007). *La charge cognitive. Théorie et applications.* Paris: Armand Colin.
3. Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive–energetical framework. *Biological Psychology, 45*, 73–93.
4. Hockey, G. R. J. (2003). Operator functional state in the analysis of complex performance. In G. R. J. Hockey, A. W. K. Gaillard, & A. Burov (Eds.), *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks* (pp. 3–7). New York: Plenum Press.
5. Matthews, M. D., & Beal, S. A. (2002). *Assessing situation awareness in field training exercises*. Military Academy West Point NY Office of Military Psychology and Leadership.
6. Meister, D. (1976). *Behavioral foundations of system development.* New York: Wiley.
7. Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548-573.

MASTERING THE UNFORESEEN

Resilience and Workload During Air Force Operations. A Psychoanalytical Point of View

Maria Gabriella Pediconi, Glauco Maria Genga, Sabrina Venzi

Maria Gabriella Pediconi
Research Professor of Dynamic Psychology at Urbino University (Italy) and Psychoanalyst of Società Amici del Pensiero 'S. Freud', Milan
maria.pediconi@uniurb.it

Glauco Maria Genga
Psychiatrist, Consultant at Institute of Aerospace Medicine of Milan (Italy) and Psychoanalyst of Società Amici del Pensiero 'S. Freud', Milan glaucomaria.genga@fastwebnet.it

Sabrina Venzi
Psychotherapist and Military Psychologist of Italian Air Force at Institute of Aerospace Medicine of Milan (Italy)
sabrina.venzi@yahoo.it

ABSTRACT

Unforeseen events are an integral part of military pilots' workload. An analysis of how pilots deal with the unforeseen increases knowledge of the Human Factor and provides information about the subjective factors of resilience.
Our work presents an analysis of the experiences of a sample of military pilots, who were asked to tell the most significant, unexpected events of their career. We analysed their subjective perceptions as well as the professional and interpersonal resources they used to face these unexpected events, which occurred during air force operations.
To master the unexpected, military pilots use three main resources: 1) Training, which must be considered the main resource in terms of Technical Skills; 2) Crew Resource Management (CRM), which combines technical training and relational factors in terms of Non-Technical Skills; 3) the individual psychological factor (IPF) that allows the pilots to integrate and put together the conditions met in the field.
The analysis of unforeseen events and the pilots' managing the outcomes of these events allowed us to explore, on the one hand, the specific weight of Operating Pressure (OP) and "Get-Home-Itis Syndrome" (GHS), both of which, if present, complicate the execution of the military task. On the other hand, we observed the importance of the interpersonal factors that are always involved. We have observed that CRM can be explored both as a resource and as a complication. Three emblematic cases will be presented. Resilience can and must be considered the *royal road* that helps us learn from the unexpected. The psychoanalytical point of view is the one that most emphasizes the arduous and inevitable work of the resilient Ego.

Keywords: Unforeseen events, Mental Workload, Resilience, Psychoanalysis

«Ab esse ad posse valet consequentia»
From the fact that something exists, it follows that it is possible
(Oxford Dictionary of Philosophy)

«I wouldn't say that fear is an unknown emotion to us. Rather, fear results when we are conscious of the fact that something unexpected and potentially impossible to deal with, might happen. Nevertheless, with all the training we have done, none of us approaches this flight with fear because we have been prepared to face every possible scenario and we are ready»
(Neil Armstrong)[1]

1. Learning from unforeseen events: hints at previous studies

As we reported in the conclusions of our paper on a sample of firefighters and military pilots last year: «we wished to approach and study the psychic reality - thoughts and affections - of professionals who, in their work, are not only able to face events that present themselves as dangers and threats for themselves and for the whole community and get out unharmed, but show that they do all this with personal satisfaction. And this is something that cannot be imposed by anyone. Therein lies an additional resource and teaching that both these professional categories can offer us.» (Pediconi, Genga, 2018, p. 89) Following the same train of thought (i.e. to deepen our knowledge of the Human Factor (HF) and the factors involved in resilience), in this new article we have examined the different ways in which military pilots deal with unforeseen events in the course of their profession.[2]
We are aware that providing an accurate definition of "unforeseen event" is not always easy. However, attitude and conduct in facing the unexpected remains a sensitive and very important issue for anyone who wants to deal with flight safety or has some responsibility in the aviation world. (Fornette et Al, 2015) We are convinced that psychoanalysis increases knowledge about errors as well as HF, delineating the Ego as a resource, even when managing Human Workload
A premise is first necessary: the profession of aviation pilot is a "young" and recent one, that dates back to a little more than a century (Tiberi, 2011).
Although technological advances have brought flight to previously unthinkable levels, it has been shown that human error is still the main factor responsible for civil and military air accidents (70-80%). (Shappell, Wiegmann, 2000; Hooper, O' Hare, 2013) Numerous studies aimed at ensuring the indispensable protection of lives involved in flight were multiplying, including the introduction of basic personality profiles (Helmreich, 1984) and an increasing attention to Human Mental Workload (Longo, Leva, 2017; Hancock, 2017; Moustafa, Luz, Longo, 2017).
Most researchers agree that, although Mental Workload (MW) is a concept that has intuitive meaning, it is difficult to define. Wickens (2017) suggests that the concept of effort or Mental Workload can be examined from three perspectives: those of measurement, prediction and consequences, even in a subjective way.
Furthermore, we saw that the resilient Ego never works alone, but is always part of a team. Resilience shows, in a very impressive way, the role of the Ego as an adaptive resource. Our paper sought to

[1] These are the words with which Neil Armstrong answered the journalist who asked him if the three astronauts were afraid. (July 15, 1969, eve of the launch of Apollo 11, during the press conference of NASA, youtube.com/watch?v=r_Ct6z9tm0M).
[2] We would like to thank the Chief of Institute of Aerospace Medicine of Milan Dr. Gen. Giuseppe Ciniglio Appiani and Dr. Col. Alessandro Randolfi, Chief of Psychiatric Section, for having hosted this research from the beginning. Furthermore, our thanks go to Col. Riccardo Ferraresi for the precise explanations and kind constant advice on this paper.

strengthen the concept of resilience thanks to the Freudian doctrine. This year we will try to show new connections between the MW, understood as demanding tasks and emergency situations, and resilience in coping unforeseen situations.

2. Unforeseen events and solutions

Mental Workload affects human performance in terms of the amount of mental capacity required to perform a given task. In general, Mental Workload is considered a multidimensional construct involving interactions between task and system demands, the operator (including mental and emotional skills) and the environment. Subjective Mental Workload is multidimensional, and its consistency is the result of interactions between the human, the task, and the environment. (Estes, 2015; Cullen, Cahill, Gaynor, 2017; Longo, Leva, 2018)) Recent works take into account the relationship between the concepts of Mental Workload, situation awareness and operative performance that can support the success of operations even in the case of variations due to circumstances beyond an individual's control. (Borghini et al, 2014)

To master the unexpected, military pilots use three main resources:

1) Training, which must be considered the main resource in terms of Technical Skills;

2) The Crew Resource Management (CRM), which combines technical training and interpersonal factors as far as Non-Technical Skills go; (Helmreich et al., 1999, 2000; Kanki, Helmreich, Anca, 2010)

3) The Individual Psychological Factor (IPF), that consists of the pilots' faculty to integrate and put together conditions met in the field.

Pilots undergo regular check-ups, to insure so that they are operational only on condition that they do not present clinical signs of pathology, including psychopathology. Piloting requires a high level of professionalism, together with the ability to establish and maintain solid interpersonal relationships, in which the competence and reliability of each person plays a leading role.[3] Thanks to the knowledge acquired through training, military pilots become capable of operating in conditions that must be considered "unordinary", i.e. in hostile environments as part of complex, highly automated systems. (Hays, 2002) Military pilots are distinguished by a higher level of risk tolerance than civilian pilots (Sicard at. al. 2003), and unexpected events provide them with a high degree of stress and thus increase their workload. Training pilots to better deal with unforeseen circumstances has become increasingly necessary. (Bourgy 2012, Casner et al. 2013; AAE, 2013).

Several researchers have argued that claim that the performing in surprise situations represents an additional competence area that requires special practice focused on pilots' attentional behavior and sensemaking. (Landman et al., 2017)

In our research, we have identified three types of contingencies, attributable to:

1) Problems with the operation of aircraft;

2) Problems external to the aircraft (meteorological and environmental in a broad sense)

3) Problems due to human error (expressions of HF).

In each of these three types of contingencies, we recognize factors that may play a causal, or more often con-causal, role in influencing the pilots' reactions to contingency.

[3] In Italy, the body responsible for civil aviation safety is the ANSV (Agenzia Nazionale per la Sicurezza del Volo). The Italian Air Force, on the other hand, has an Inspectorate for Flight Safety: since 1991, it is responsible for studying problems related to flight safety and issuing directives concerning the prevention, investigation and legal aspects of accidents involving military aircraft. In addition, the Air Force «has for many years established the Department of Aeronautical and Space Medicine within the Experimental Flight Centre at Pratica di Mare. Its particular task is the aero-physiological training of all personnel in critical flight situations». (Ciniglio Appiani, 2010).

Assuming a typical situation (trained pilot = perfection), we must recognize that there are some factors that limit judgment and consequently interfere with their decision-making process. Training plays a prominent role in determining the level of flight safety, while its lack is a source of error that can prove extremely dangerous. Training contains at least two factors:

CRM: Crew Resource Management

CC: Crew Cooperation

For decades airline pilots had previous military experience, often in single-seater fighter planes. Therefore, the captain was considered to be one who always knew what to do. (Brischetto, 2019) No co-pilot used to dare make comments on the work of the captain. In the early 1980s, however, as a result of serious flight accidents due to lack of cooperation between crew members, things changed dramatically. CRM soon became a means of handling errors.[4] (Helmreich, 1999)

Dealing with the error in a correct way involves accepting it and understanding its causes and effects. Pilots must maintain Situational Awareness, i.e. a constant idea of themselves and the aircraft in relation to the flight environment, external threats and the mission. (Brischetto, 2019)

There are two main factors that have a negative impact on the pilot's performance:

- *Operational Pressure* (Cain, B., 2007; Kanki, B., Helmreich, R., Anca, J., 2010; Academie de l'Aire et de l'Espace, 2013; Cahill, Cullen, Gaynor, 2018)
- *Get Home-Itis Syndrome* **(**Causse, Dehais, Pastor, 2011; Causse et al., 2013)

3. The resilient Ego at work in mastering the unforeseen situations: data and findings

From the theoretical background to the research data

As we have already said, it has been found that at least 70% of all air accidents involve human error. Causes of error include fatigue, workload size, and fear as well as cognitive overload, poor interpersonal communication, imperfect processing of information, and flawed decision making. (Ciniglio Appiani, 2013; Weidlich, Ugarriza, 2015)

On the other hand, we consider Mental Workload affects human performance in terms of the amount of mental capacity, both emotional and cognitive, required to perform a given task. Recent works take in account the relationship between the concepts of MW, situation awareness and operative performance that can support the success of operations even in the case of variations due to circumstances beyond an individual's control. (Borghini et al, 2014)

Although the quantitative methods remain privileged in this field (Bakker&al., 2007; Herbst&al., 2014; Estes, 2015), we will analyze the resilience as the subjective mental resource in facing unexpected events and Mental Workload from a qualitative research perspective. This is the best way to describe the resilience of military pilots in terms of the subjective perception of unexpected tasks in the case of unforeseen events. Focusing on the mental content allow us to discover which subjective factors promote success in the management of unforeseen situations.

The pilot remains the undisputed protagonist of the analysis, evaluation and management of risks, both in the case of single operations, complex activities and interactions. The interactions need to be managed with: (a) machinery and equipment, (b) context and (c) social group (O'Connor et al., 2002). The effectiveness of the process of mastering unforeseen events depends on the HF activated in everyone's

[4] In the USA's civil field, it was United Airlines that started a specific training, which took the name of Crew Resource Management (CRM). Initially (1980), the program was called Cockpit Resource Management (also CRM). It was in fact focused on the Human Factor in the cockpit. Only later, it was extended to the entire crew. The reason for the acronym's change from *cockpit* to *crew* was the extension in training to flight attendants, mechanics and anyone responsible for flight safety, not only to pilots.

evaluation.

Methods. A qualitative analysis of subjective testimonies about the most representative, unforeseen events met during one's military career

Our research aims to lead a qualitative analysis of professional experience in terms of subjective perceptions. In their job, the pilots' strength of the Ego is very important for the success of operations and for their personal safety. We will therefore explore the kind of mental content that accompanies specialized performance during unforeseen situations.

Actually, «no one else is more prepared to provide an accurate judgment on workload experienced than oneself». (Pereira da Silva, 2014, p. 314) However, it can become difficult to discriminate between physical workload and Mental Workload. Indeed, the person may consider the demands of external tasks and the experienced mental effort as both sources of perception, and jointly quantify the invested mental effort.

Among researchers, not only psychologists, attention to foundational knowledge of qualitative research methods in social sciences is increasing in order to gain clearer access to human experience, attitudes and resources (Wertz et al., 2011; Cahill, Cullen, Gaynor, 2018)

Our research involved 40 Italian military pilots (aged between 29 and 54 years). Specifically, the pilots interviewed came from: The Air Force (17), the Land Army (10), the Carabinieri (3), the Navy (5), the Financial Guard (3), the Firefighters (2).[5]

The aircraft on board which they operate (generally for several years, since they are officers in SPE (Permanent Effective Service) are fixed-wing (10) or rotating-wing (30). Both planes and helicopters are used for different purposes: from transport to reconnaissance, training, rescue or combat. It is not possible for us to specifically refer here to the aircraft models, because they are currently in use by the Italian Armed Forces, which treat this information as confidential.

The pilots were interviewed during the periodical check-up visits they must undergo at the Institute of Aerospace Medicine *Angelo Mosso* in Milan. More precisely, our sample was composed of Official Pilots intercepted during their "ordinary visits", or "routine visits", i.e. without the object of the visit being the emergency of a symptomatology or anything else that could affect suitability.

Each of them were asked to consent to the collection of data, specifying that their anonymity would be preserved. In all cases, they gladly accepted the invitation and responded without any problems or hesitation to our question, which we report here:

«Please, could you describe briefly the most significant unexpected event that you have had to face during your career as a pilot? We would like to know what it was about and what resources you had to activate in order to overcome it. »

In the table on page 7 we have summarized the descriptive data extracted from the testimonies of the 40 pilots we met. The data provide an overview of the main factors that can be detected in the management of unforeseen events and that contribute to creating resilience. We briefly illustrate the main findings:

- *The kinds of unforeseen events* that pilots had to face can be divided in three categories, as they are related: 1) to the machine; 2) to the weather conditions or external conditions of the aircraft (birds, lack of visibility, high voltage wires, brown out): 3) to a human error (HE), their own or those of others (such as: wrong manoeuvre, uncoordinated manoeuvre, physical indisposition; forgetfulness of a passage of the procedure, avalanche induced by the blades, bang to the trolley, sheared cable, etc.)
- We have indicated *extra-workload* the presence of additional workloads in terms of perceived

[5] To be precise, since 1961, the Italian Firefighters has been organised into a civilian, non-military National Corps.

operating pressure as it is always possible to trace its link with the conditions of the flight during which the unexpected was experienced; the operating pressure is perceived in most cases;

- *The type of perceived surprise*, when the unexpected was detected, i.e. the subjective condition of situational awareness recorded by the pilot. We grouped these data in: surprise of type A if the shock perceived is immediately attributed to a mechanical cause or in any case to the machine («Oh, God! the engine went out!», or «Oh, s…t, the engine went out!»). Surprise of type B, when the shock perceived is immediately attributed to a human error («Oh, God, what have I done! »). Not only: A+B collects the cases in which the pilot perceives a shock that first seems attributable to a breakdown of the machine, and immediately later to an error of the pilot or flight partner. The perceived surprise can be traced back to three main affections: 1) a feeling of *impotence* due to an emergency resulting from technical failure or external conditions of the aircraft; 2) *remorse or guilt* in the case of unforeseen events caused by human error; 3) *anger* if the perceived danger seems due to mechanical damage but is caused by a human error.
- *The contribution of others* is perceived by the subject as *favourable* when others of the crew or on the ground have supported and favoured positive resolution of the critical event. It can be an obstacle (*adverse***)** if others are perceived as a problem, a factor that contributed to produce or aggravate the critical event. *Hierarchy* comes into play when the presence of the other, known for his institutional role, has helped to make the critical event more complex
- *The professional resource* used to deal with the unexpected:

1) we indicated with *expertise* the technical and procedural competence that pilots have learned and internalized through training and experience. This concept is linked to the concept of the four stages of competence, studied by authors in several fields, «with novices moving from unconscious incompetence, conscious incompetence, conscious competence and finally unconscious competence». (Byrne, 2017, p. 193)

2) we called *CRM* the resolution procedure in which the contribution of crew cooperation has assumed a predominant importance with respect to the inevitable technical competence.

LEGEND

1 Aircraft:

- FWA = Fixed Wing Aircraft (Combat Aircraft or Transport/Cargo/Multiengine Aircraft)
- RWA = Rotary Wing Aircraft (Helicopter)

2 Role, or position:

The specific role of the pilot during the mission in which the unexpected occurred: pilot, navigator, co-pilot, aircrew member.

3 Kind of Unforeseen Event:

Flight incident, Serious Flight Incident; Flight Accident; Serious Flight Accident

4 Flight Type:

Mission; Training; Search and Rescue (SAR) Flight; Combat Flight; Transport Flight.

5 Event:

The event can be due to: Aircraft malfunctioning; Weather conditions; Human Error

6 Extra-workload:

Operative Pressure

7 Kind of surprise

The surprise is different as it can be attributed to a mechanical cause (A) or to a Human error (B)

8 Other's Presence:

It can be perceived as favourable or adverse. Sometimes the subject does not make any reference to it

9 Professional Resource

It can be of two types: the expertise due to training and experience; the contribution of the Other (CRM or CC)

1 AIRCRAFT	2 POSITION	3 KIND UNFORESEEN	4 FLIGHT TYPE	5 EVENT	6 EXTRA-WORKLOAD	7 KIND OF SURPRISE	8 OTHER'S PRESENCE	9 PROFESS. RESOURCE
RWA	Aircrew	FI	Mission	Aircraft		A	Favourable	Expert.
RWA	Pilot	FI	Mission	Aircraft	Op. Press.	A		Expert.
RWA	Pilot	FI	Mission	Aircraft	Op. Press.	A	Hierarchy	Expert.
RWA	Pilot	FI	Combat	Aircraft	Op. Press.	B		Expert.
RWA	Pilot	FI	Combat	Aircraft	Op. Press.	A	Hierarchy	Expert.
RWA	Pilot	FI	Mission	Aircraft	Op. Press.	A		Expert.
RWA	Pilot	FI	Mission	Aircraft	Op. Press.	A		Expert.
RWA	Pilot	FI	Mission	Aircraft	Op. Press.	A	Favourable	Expert.
FWA	Pilot	FI	Training	H.E.		A + B	Adverse	Expert.
RWA	Pilot	FA	Mission	Weather	Op. Press.	A	Adverse	Expert.
RWA	Pilot	FI	SAR	H.E.	Op. Press.	A	Adverse	Expert.
RWA	Pilot	FI	Training	Weather	Op. Press.	A + B	Adverse	Expert.
RWA	Pilot	FI	Mission	H.E.	Op. Press.	B	Adverse	Expert.
FWA	Pilot	FI	Training	Aircraft		A		Expert.
RWA	Pilot	FA	Mission	Aircraft		A		Expert.
RWA	Pilot	FI	SAR	H.E.	Op. Press.	B	Adverse	Expert.
RWA	Copilot	FA	Training	H.E.	Op. Press.	B	Hierarchy	Expert.
RWA	Pilot	FI	Transport	Weather	Op. Press.	A	Favourable	Expert.
RWA	Copilot	FI	Mission	Aircraft	Op. Press.	A	Favourable	Expert.
RWA	Pilot	FI	Training	Aircraft	Op. Press.	A	Adverse	Expert.
RWA	Pilot	FI	Training	H.E.	Op. Press.	A+B	Hierarchy	CRM
RWA	Pilot	FI	Transport	H.E.	Op. Press.	A+B	Hierarchy	CRM
FWA	Pilot	FI	Mission	Weather	Op. Press.	B	Favourable	CRM
FWA	Navigator	FI	Training	Aircraft	Op. Press.	A	Favourable	CRM
FWA	Navigator	FI	Training	H.E.	Op. Press.	A		CRM
FWA	Navigator	FA	Mission	Aircraft	Op. Press.	A	Favourable	CRM
RWA	Copilot	FI	Mission	Aircraft	Op. Press.	A	Adverse	CRM
RWA	Pilot	FI	SAR	H.E.	Op. Press.	B	Adverse	CRM
FWA	Pilot	FI	Training	H.E.	Op. Press.	A	Favourable	CRM
FWA	Pilot	FI	Mission	Aircraft		A	Favourable	CRM
RWA	Pilot	FI	SAR	Aircraft	Op. Press.	A	Favourable	CRM
FWA	Pilot	FI	Mission	Weather		A	Favourable	CRM
RWA	Copilot	FI	Mission	Aircraft	Op. Press.	A+B	Hierarchy	CRM
RWA	Pilot	FI	Training	H.E.	Op. Press.	B	Favourable	CRM
RWA	Aircrew	FI	Training	H.E.	Op. Press.	B	Adverse	CRM
RWA	Pilot	FI	SAR	Weather	Op. Press.	A	Favourable	CRM
RWA	Pilot	FA	Training	H.E.	Op. Press.	B	Hierarchy	CRM
RWA	Pilot	FI	Transport	Weather	Op. Press.	A + B	Hierarchy	CRM
FWA	Pilot	Serious FA	Training	Aircraft	Op. Press.	A+B	Adverse	No Act
RWA	Pilot	FA	Training	H.E.	Op. Press.	B	Hierarchy	Luck

The *Individual Psychological Factor* (IPF) is the skill that brings together the three aspects analyzed above in terms of *extra-workload*, *perceived surprise* and *contribution of the Other (place of the Other)*, preparing the individual to be conducive to the success of the operation, despite the danger to one's life. The IPF is at the same time: 1) an analytical factor, as it allows the subject to distinguish in a matter of seconds what is happening, how the operating pressure is changing, if there is an error or a technical failure, and how the contribution of the other can be used; 2) a synthetic factor, because thanks to the processing of contingent data, the subject will develop the resolutive conduct. Since this resolutive conduct will have served not only to return to the ground safe and sound, but also to increase personal experience on the flight, we can call it resilient.
In the brief testimony, each pilot told the most representative, unexpected event of his career, giving voice to the technical-professional as well as psychological-personal work that allowed him to manage it successfully. It is very interesting that each of the short interviews, together with the unexpected event and its perceived causes, illustrate the moment of the surprise we commented on above.

4. Analysis of three emblematic cases

We will now present three examples: three cases chosen from the forty pilots' testimonies collected. The question we asked each of them was: *«Please, describe briefly the most significant unexpected event that you had to face during your career as a pilot: what it was about and what resources you had to activate in order to overcome it.»*

A) By surprise, a prompt resolution (Military Helicopter Pilot).
«Years ago, also because of my long absence for missions, I went through a difficult period with my family. I had been flying for many years. I remember that once I happened to carry a General, who had also been operating as a helicopter pilot years before, but on a different machine. At a certain point he asks me for the controls. I give it to him. But I realized that he was not flying well: there was no agreement between the main rotor and the tail rotor. So, I intervened in turn on the controls (they are identical, there is no device that makes one of the two leader: it depends on the muscle strength with which you operate). At that point, the General, not noticing anything, obviously made a small force, but I forced, that is, I played by surprise. He didn't even notice it, but the other crew members did! At the debriefing I didn't say anything: why complicate things and set myself against a superior? »
Analyses. The unexpected is not due to a malfunction of the machine, except as an effect of the inexperience of the superior, who in the past "had also been operational". Thanks to the training, the subject was able to correct the flight attitude and "trap the error", which had been committed by those who had asked him for the commands. The mistake of the Senior Officer is therefore due to the overestimation of his skills in piloting, but there is also a share of error due to the "weight" of the hierarchy in the narration. The general may not have noticed anything, and the pilot, in fact, preferred not to report anything for reasons certainly understandable, but questionable.

B) Get-Home-Itis Syndrome (Air Force Pilot)
«We were returning from a long mission abroad, when we were asked to make an additional operational stopover. I was the commander, and I immediately asked the crew: okay, we all agreed. But we were also all tired, plaffed, and had passed the sixteen consecutive hours of flight! Upon arrival in Pisa, the weather conditions were disastrous, the heavy rain, you could not see anything. I thought: "If I don't see the runway in a few minutes, I have to get up, and go to another airport! We had enough fuel: the problem was not that, but the tiredness of all of us. On board there was a silence of lead. Fortunately, at a certain moment I saw the runway and we landed. Everyone else, then, complimented me: big pats on the back, and go!»

Analysis. In this case, the unforeseen event comes from the request to make an additional stopover. There are two sources of the problem: the adverse weather conditions and the fatigue of the entire crew. The difficulties experienced by the pilot are: operating pressure and GH-itis syndrome. It is really good that the Commander immediately resorted to Crew-Cooperation, as is demonstrated by the fact that he then received compliments from the whole team.

C) Ground resonance (Financial Guard, Helicopter Pilot)
«I was still on the ground, ready for take-off. Suddenly, I noticed that the helicopter had resonated with the ground (Ground Resonance). It's a bit like what happens when a tenor makes a high note and the windows break. I knew in that case you must immediately move or raise the helicopter, even a little, or immediately increase the speed of the rotor, otherwise the device will break, literally. It's something we're prepared for, in the sense that we're taught what it is and what to do, but you can't simulate it, because it would be too dangerous. I was able to get up from the ground. It all lasted 15 seconds. »
Analysis. The unforeseen event is entirely attributable to the machine. The pilot identified it correctly, he knew the possible procedures to remedy it, he was aware that the occurrence of Ground Resonance could not be simulated in practice. He worked accordingly, resolving the problem without structural or personal damage.

Finally, we are reporting in note a fine example of storytelling taken from the magazine "Sicurezza del Volo" (Flight Safety), published by the Inspectorate of Flight Safety of the Italian Air Force.[6]

4. Conclusions

In this last paragraph we would like to briefly summarize the main findings obtained with this work and show how they are enlightened and interpretable through the Freudian psychoanalytic point of view.

The surprise (see column 7)
The Ego, in processing what happens in external reality, as well as in its own thoughts, uses not only perception or representation, but also affection and verbal memories. For Freud, «(…) becoming conscious is no mere act of perception, but (...) a further advance in psychical organization.» (Freud, 1915, SE XIV, p.194). And again: «All perceptions which are received from without (sense-

[6] «During the final portion of a training mission flown by two Flight Instructors (training recurrency mission) the crew decided to practice some OEI (OEI - one engine inoperative) approaches. The OEI training consists in simulating single engine flight emergency conditions (…) During the second simulated engine failure approach the "ENG mode selector" was set involuntary to "idle" instead of "OEI training" causing the helicopter to be in a real single engine emergency operation. (…) In fact, one of the two pilots misconfigured the "ENG mode selector" instead of the "OEI training selector" during the second approach and the subsequent take-off. The pilot monitoring, instead of running the simulator switch, performed the actions usually accomplished during normal ground power check. In addition, both pilots did not realize through the "onboard instruments" indications of the erroneous procedure in place: (…) The incident has highlighted the threat that can be concealed in routine errors during flying activity. In the specific case the incident was generated by the co-pilot performing routine actions rather than focusing on the one required by the procedure; the "ENG mode selector", instead of the "OEI training" selector, was actuated. It is worth remembering that slips and lapses are related to "skill-based errors" that is unplanned actions during the execution of routine acts (check list, step of a procedure, unintentional use of commands). Generally, they occur when there is excessive workload, task fixation, distraction from foreign elements, or excessive relying on automation. Therefore, in order to avoid similar mistakes, it is recommended to analyze similar events with the help of the flight simulator where possible, to carry out more accurate and thorough briefings that define in detail all the procedures to be performed during the flight, to take full advantage of the CRM this way avoiding to run a command that is not pertinent to the action required by the current procedure, and to use the challenge and response check list technique, especially when training for unrecurrent particular maneuvers.» (M. Boveri, *Anatomy of Flight Inconvenience UH-139*, Sicurezza del Volo, 332/2017).

perceptions) and from within – what we call sensations and feelings – are Cs. [conscious] from the start.» (Freud, 1923, SE XIX, p. 19) This process of thought underway in the experience of surprise is not only cognitive, but aimed at judgement, which consists in knowing immediately if what happened concerns me or if it concerns an external element (the weather or the aircraft, in this case).

The place of the Other (see column 8)
We have recorded a tripartition: a) the Other exercises a positive influence on the subject; b) the Other exercises an adverse one; c) the influence must be attributed to the hierarchy. This third case illuminates a part of the subject, which concerns itself with dealing with authority. This relationship with authority is always demanding and is configured in a different way than that of a relationship with a friend or with the enemy. In the well-known Freudian study on the group psychology, Freud describes the "artificial group" of the army: «The Commander-in-Chief is a father who loves all soldiers equally, and for that reason they are comrades among themselves (...) Every captain is, as it were, the Commander-in-Chief and the father of his company, and so is every non-commissioned officer of his section.» (Freud, 1921, SE XVIII, p. 94).

Professional resources (see column 9)
We asked each pilot what he had thought up to quickly find the solution at the time of the unexpected challenge. In the reply, some of the pilots testified to the support they received from the training (expertise). Others, on the other hand, recognised the important or even decisive role of the Other (CRM or CC). In both cases, this is emotional (affective) recognition.

Further research could lead to an investigation
1) Whether or not human error produces a change not only in the type of surprise, as we have seen, but also in the characteristics of resilience;
2) Are there significant differences between those (few) whose narrative does not include operational pressure and their (many) colleagues who have reported such pressure?
3) Why do some subjects not describe or mention the quality of the presence of the Other? In all cases (airplane or helicopter, training or operational mission or other) each pilot never acted in isolation, so the presence of the Other is to be considered ubiquitous.
4) Possible extensions of our research also in the medical field, where the use of error management strategies is equally useful and already studied. (Helmreich, 2000)

A comment on the method
We know that new quantitative methods of measuring the Mental Workload are developed every year, so it would have been useful to mention the prejudices that "a brief testimony and its analysis" represent (lack of memory, for example). However, even in the case of a memory deficit that had erased part of the experience while keeping only the reported data, this would have been sufficient to outline the thinking that worked in the face of the unexpected. We hope that the development of our study will contribute to a better understanding of the subjective fundamentals of MW in unforeseen situations and of why some people can cope better than others with unforeseen difficulties.

Surprise, Mental Workload, Resilience: a very articulated network
Recent researches (Cantoni, 2014, Fornette et al, 2015) include determination, challenge desire, obstinacy, auto-efficacy among psychological constituents of resilience. Our research confirms these constituents as factors of the resilient Ego, which helps to manage external reality with the aim of finding good solutions even in facing dangerous situations. The analysis of unforeseen events, and the pilots' managing of the outcomes, allowed us, on the one hand, to explore the specific weight of Operating Pressure (OP) and "Get-Home-Itis Syndrome" (GHS), both of which, if present, complicate the execution of the military task. On the other hand, we were able to observe the importance of the

interpersonal factors that are always involved. We noticed that CRM can be explored both as a resource and as a complication. The first -positive- case is represented by the perceived others' support during the air force operations. The second -negative or problematic- case can be represented by the other perceived as an obstacle during the operation or by the not always comfortable relationship with hierarchy.
Resilience can and must be considered the *royal road* that helps us learn from the unexpected. The psychoanalytical point of view is the one that most emphasizes the arduous and inevitable work of the resilient Ego. Our findings show that pilots master unforeseen events based on technical skills (expertise) and non-technical skills (CRM) composing the safety system. If it is true that error management is a concept to constantly update, we need to improve knowledge about individual perception and resources in order to highlight the Human Factor.

References

1. Academie de l'aire et de l'espace: Dealing with Unforeseen Situations in Flight. Air and Space Academy (2013).
2. Bakker, A., Hakanen, Y., Demerouti, E., Xanthopoulou, D.: Job Resources Boost Work Engagement, Particularly, when Job Demands are High, Journal of Educational Psychology, vol. 99(2), pp. 274-284 (2007).
3. Borghini, G., Astolfi, L, Vecchiato, G., Mattia, D., Babiloni, F.: Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci. Biobehav. Rev., 1-18 (2014).
4. Bourgy, M.: Cognitive adaptation and improvisation in dynamic environments: Integrating sensitive experience in psychological models of expertise. Université de Paris 8, saint Denis, France (2012).
5. Boveri, M.: Anatomy of Flight Inconvenience UH-139, Sicurezza del Volo/Flight Safety, Istituto Superiore per la Sicurezza del Volo, Aeronautica Militare, 332, pp. 12-17 (2017).
6. Brischetto S., Tore R: Elementi di sicurezza aeronautica, Esculapio (2019).
7. Byrne, A.: Mental Workload as an Outcome in Medical Education. In: Longo, L., M.C. Leva (Eds.), H. Mental Workload. Models and Applications, Springer International Publishing AG, pp. 187-197 (2017).
8. Cahill, J., Cullen, P., Gaynor, K.: Estimating the Impact of Work Related Stress on Pilot Wellbeing and Flight Safety. In: Longo, L., M.C. Leva (Eds.), Human Mental Workload. Models and Applications, Springer International Publishing AG, pp. 7-32 (2018).
9. Cain, B.: A Review of the Mental Workload Literature, RTO-TR- HFM- 121. Part. II. Pp. July, (2007).
10. Cantoni, F.: La resilienza come competenza dinamica e volitiva, Giappichelli (2018).
11. Casner S.M., Geven R.W., Williams K.T.: The Effectiveness of Airline Pilot Training for Abnormal Events, Human Factors, Vol. 55, No. 3, June 2013, pp. 477-485 (2013).
12. Causse M., Dehais, F., Péran P., Sabatini U., Pastor, J: The effects of emotion on pilot decision-making: A neuroergonomic approach to aviation safety, Elsevier (2013).
13. Causse, M., Dehais, F. Pastor, J.: Executive functions and pilot characteristics predict flight simulator performance in general aviation pilots. The International Journal of Aviation Psychology, 21 (3) pp. 217-234 (2011).
14. Ciniglio Appiani, G.: Argomenti di medicina aeronautica, (Ed), Rivista Aeronautica, Roma (2010).
15. Cullen, P., Cahill, J. & Gaynor, K. (2017). Pilot wellbeing. Paper presented at the Flight Safety Foundation 2017 – 70th International Air Safety Submit (IASS), Dublin, October 23-25 (2017).
16. Estes, S.: The Workload Curve: Subjective Mental Workload. Human Factors, vol. 57, n. 7, pp. 1174-1187 (2015).
17. Fornette M.P., Darses F., Bourgy M.: How to improve training programs for the management of complex and unforeseen situations? In D. de Waard, J. Sauer, S. Röttger, A. Kluge, D. Manzey, C. Weikert, A. Toffetti, R. Wiczorek, K. Brookhuis, and H. Hoonhout Eds (2015). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2014 Annual Conference. ISSN 2333-4959 (online). Available from:

hfes-europe.org
18. Freud, S. (1915): The Unconscious, Papers on Metapsychology, SE, vol. 14, p. 194.
19. Freud, S. (1921): Group Psychology and the Analysis of the Ego, SE, vol. 18, p. 94.
20. Freud, S. (1922): The Ego and The Id. SE, 19, p. 19.
21. Hancock, P.A., Whither Workload? Mapping a Path for Its Future Development. In: Longo, L., Leva M.C. (Eds.): Human Mental Workload. Models and Applications, Springer International Publishing AG, pp 3-17 (2017).
22. Hays, M.: The Training Of Military Pilots: Men, Machines And Methods, School Of Advanced Airpower Studies, Air University Maxwell Air Force Base, Alabama, (2002).
23. Helmreich, R.L.: Cockpit Management Attitudes, Human Factors, 26(5), 583-589, (1984).
24. Helmreich, R.L., Merritt, A.C., Wilhelm, J.A.: The evolution of Crew Resource Management training in commercial aviation. International Journal of Aviation Psychology, 9(1), 19-32 (1999)
25. Helmreich, R.L., Merritt, A.C. Safety and Error Management: The Role of Crew Resource Management. In B.J. Hayward & A.R. Lowe (Eds.), Aviation Resource Management (pp. 107-119). Aldershot, UK: Ashgate (2000).
26. Helmreich, R.L.: On Error Management: Lesson from Aviation, BMJ 2000;320:781 (2000).
27. Herbst, N.R., Huber, N., Kounev, S., Amrehn, E.: SelfAdaptive Workload Classification and Forecasting for Proactive Resource Provisioning, Concurrency Computat, Pract Exper, 26, 2053-2078 (2014).
28. Hooper B.J., O' Hare D.P.A.: Exploring Human Error in Military Aviation Flight Safety Events Using Post-Incident Classification Systems. Aviat Space Environ Med. 84:803-13 (2013).
29. Kanki, B., Helmreich, R., Anca, J.: Crew Resource management, Elsevier, 2nd Edition (2010).
30. Landman, A., Groen, E.L., van Paassen, M.M., Bronkhorst, A., Mulder, M.: The Influence of Surprise on Upset Recovery Performance in Airline Pilots, The International Journal of Aerospace Psychology, 27:1-2, 2-14 (2017).
31. Longo, L., Leva M.C. (Eds.): Human Mental Workload. Models and Applications, Springer International Publishing AG (2017).
32. Longo, L., Leva M.C. (Eds.): Human Mental Workload. Models and Applications, Springer International Publishing AG (2018).
33. Moustafa, K., Luz, S., Longo, L.: Assessment of Mental Workload: A Comparison of Machine Learning Methods and Subjective Assessment Techniques, Springer International Publishing AG 2017, pp 30-50 (2017).
34. O'Connor, P., Flin, R. Fletcher, G.: Methods used to Evaluate the Effectiveness of Flighterew CRM training in the UK aviation industry. Human Factors and Aerospace Safety, 2, 235-255 (2002)
35. Pediconi M.G., Genga G.M.: On the Earth as well as in the Sky. Facing the Risks Between Workload and Resilience: A Psychoanalytic Point of View. In: Longo, L. & Leva, C. (eds). (2018). H-Workload 2018, The 2nd. International Symposium on Human Mental Workload. Netherlands Aerospace Centre, Amsterdam, 20-21 September, 2018. isbn:9781900454643
36. Pereira da Silva, F.: Mental Workload, Task Demand and Driving Performance: What Relation? Social and Behavioral Sciences 162, 310 – 319 (2014).
37. Shappell, S.A., Wiegmann, D.A.: The Human Factors Analysis and Classification System (HFACS). U.S. Department of Transportation Federal Aviation Administration (2000).
38. Sicard, B., Taillemite, JP., Jouve, E., Blin, O.: Risk Propensity in Commercial and Military Pilots, Aviation, Space, and Enviromental Medicine, vol 74(8) (2003),
39. Tiberi, L.: Piloti. Storia dell'Aeronautica Italiana, DVD, 100m, Distr. Istituto Luce, Italy (2011).
40. Weidlich, C.P., Ugarriza, D.N.: A Pilot Study Examining the Impact of Care Provider Support Program on Resiliency, Coping, and Compassion Fatigue in Military Health Care Providers. Military Medicine, 180, 3:290 (2015).
41. Wertz, F.J., Charmaz K., McMullen, L., Josselson R., Anderson R., McSpadden, E.: Five Ways of Doing Qualitative Analysis Phenomenological Psychology, Grounded Theory, Discourse Analysis, Narrative Research, and Intuitive Inquiry. Routledge (2011).
42. Wickens, Ch.D.: Mental Workload: Assessment, Prediction and Consequences. In: Longo L., Leva M.C. (Eds.), H. Mental Workload. Models and Applications, Springer International Publishing AG, pp 18-29 (2017).

Mental workload during (un)familiar food tasting experiences

Ana C. Martinez-Levy[1,4], Nicolina Sciaraffa[4,5], Giulia Cartocci[2] , Patrizia Cherubino[2,4] , Alessia Vozzi[4], Vincenzo Ronca[4] and Fabio Babiloni[2,3,4]

[1]Dept. of Communication and Social Science, Sapienza University of Rome, Rome, Italy
[2]Dept. of Molecular Medicine, Sapienza University of Rome, Italy
[3]Dept. of Computer Science, Hangzhou Dianzi University, Xiasha Higher Education Zone, 310018, Hangzhou, China
[4]BrainSigns Srl, Rome, Italy
[5]Dept. Anatomical, Histological, Forensic and Orthopedic Sciences, Sapienza University of Rome, Rome, Italy

Abstract. Food product experiences have already been studied from different phases interaction and by different measures. However, the measurement of the mental workload during the interaction with food products in a tasting experience has not been deeply investigated in literature. The aim of this study is to investigate such reactions using the Electroencephalography (EEG): brain signals have been recorded with a 6-channel system (EEG frontal theta) in order to test the interaction across two foreign food products and two local ones. Furthermore, participants were asked to evaluate familiarity with the products at first sight and after having tasted it. The EEG was processed in order to obtain a mental workload index, while the familiarity index was obtained as an average value on the declared judgments. A higher mental effort and less familiar perception was found during the tasting interaction with foreign products than with local ones. Results could deepen the knowledge on the cognitive response to food products tasting experiences characterized by their different origin in terms of familiarity.

Keywords: Mental workload, tasting experience, food, familiarity, EEG.

1 Introduction

In recent years the food and beverage sectors have been taken advantage of neuroscientific techniques for consumers' studies. Researchers and companies apply those methods to the study of products on its extrinsic features such as packaging, price, colour or shape and on its intrinsic features, such as taste and aroma. In particular, the term *neurogastronomy* has grown up in the last years [1] [2]**,** and within it the interest in the cognitive processes related to the taste sense. The studies carried out so far in the perception of consumers towards food and beverages have determined the importance of the perception of products in their final choice and acceptance in the market[3]–[5]. The hedonic perception of taste can be modulated by diverse factors, including consumption habits or the subconscious associations of products. A factor that could influence on the perception of food is the familiarity with it. Product familiarity is defined as "*the evaluated judgment of consumers regarding their subjective knowledge about the product"* [6]. Unfamiliar foods generate less positive expectations towards the product [7] and their absence of previous taste experiences are linked to low hedonic consumer perception[8]. So, the familiarity with a product is important for cross-cultural researches as products that are consumed in one culture could not be accepted or easily to perceive for different cultures consumptions[9]. Generally traditional likings' ratings are used to measure how acceptable is a product in cross-cultural researches. However, as described above, rational responses may not represent consumer preferences totally. Thanks to neuroscientific studies these aspects can be deeply understood. Particularly, the gustatory system and its human brain processing information has been deeply examined using techniques like functional Magnetic Resonance Imaging (fMRI) and magnetoencephalography (MEG)[10]. The initial sensory processing of taste is associated with the insula [11], which is considered the primary taste area. Instead, the secondary taste area is associate with the orbitofrontal cortex and pre-frontal cortex, as they are related to the taste hedonics' recognition [12]. Several studies employing the Electroencephalography(EEG) focused on the Pre-Frontal brain areas confirmed the relationship between prefrontal brain activity and the taste processing information [13], [14]. The possibility of the application of a non-invasive technique like EEG allows to investigate brain processes not only in laboratories, but also during daily activities in life. Particularly in the food and beverage sector, several studies imply EEG technique to analyse the extrinsic products features [5][15] and also intrinsic ones[16], [17]. Most of these researches are focused on understanding brain processes when an experience is pleasant or unpleasant by the imputation of an Approach-Withdrawal Index[5], [18], calculated by means of motivational processes in terms of alpha band (8-12 Hz), towards the stimuli based on which an increasing left hemisphere activity is associated with approach attitude to the stimulus, while an increasing right hemisphere activity is associated with withdrawal attitude[19]. On the other hand, changes in the EEG spectral power over the frontal scalp areas in theta frequency band (4-7Hz) have been connected to higher levels of task difficulty[20], its increase has been observed when the required mental workload increases [21]. Particularly the term "*mental workload"* can be defined as the proportion of information processing capability used to perform a task[22][23] and it involves neurophysiologic, perceptual and cognitive processes[24]. A high level of mental workload reflects not only task specificities, but also performer features[25]. It is applied in different research fields: neuro-aesthetics[26], for the detection of the effort employed during avionic and car driving tasks [27][28], during different challenging listening conditions [29], during human–computer interaction studies[30]. It is considered a very relevant mental concept in cognitive neuroscience applied to those fields where human decision-making is crucial, such as neuroeconomics and neuromarketing because of its close relationship between human performance[31]. Despite this evidence, the mental workload has not yet been studied in taste research. Therefore, the aim of this paper is to investigate the cognitive reactions of a group of local (Italian) people to the cross-sensory interaction with an intrinsic feature(taste) of products belonging to different countries (foreign and local). We estimated such cognitive reactions by using the mental workload index mentioned above. We investigated the influence of the familiarity with the products on these brain processes in order to predict if the external factors such as the origin of the products can influence on its decoding information processes during its tasting experience. Results will shed light on business applications for food companies/marketers and in academic researches on the brain circuits during a taste experience. Based on the aforementioned literature, the following research hypothesis was posed:

H1: Foreign products which are unfamiliar for consumers before and after the taste experience have higher mental workload values than local products during the tasting experience.

2 Materials and methods

2.1 Experimental protocol

Eight healthy volunteers (four female) all of Italian nationality have been involved in the study. None of them consumed Chinese food in their daily routine. Informed consent was obtained from each participant after the explanation of the study, which conformed to the revised Declaration of Helsinki and was approved by the local institutional ethics committee. The experiment consisted in the comparison of two different typologies of food products: a foreign group with Chinese products and the local one with Italian products. The foreign group consists of four different products, where the two most unfamiliar ones were chosen for the study; and the local group of two different products. Products were randomized and the same portion of food was given to all participants. During the study, participants interacted with the products during three different phases:

1. Observation of an empty plate as baseline (30 s)
2. Product observation (30 s)
3. Product tasting of variable duration.

Participants were asked to evaluate their level of familiarity with the products on a scale from 0 to 10 before and after the taste of each one. The question was: "How familiar are you with the (aspect/taste) of this product?". Thanks to the interview it was possible to choose the two foreign products that were less familiar to consumers in order to be compared with the two local products, both in terms of aspect (*before taste* question) and taste (*after taste* question). Figure 1 shows the four products tested.

Local Product 1

Local Product 2

Foreign Product 1

Foreign Product 2

Figure 1. Foreign and local products tested.

2.2 Signal processing

The frontal brain activity has been recorded by means of 6 dry electrodes (Fpz, AF3, AF4, AFz, F3, F4) using the LiveAmp system (BrainProducts) with a sampling frequency of 250 Hz. All the electrodes were referred to both earlobes and their impedances were kept below 10 kΩ. The EEG signals were firstly band-pass filtered with a fifth-order Butterworth filter between 1 and 30 Hz and then segmented into epochs of 1 s. The Fpz signal has been used to correct eyes-blink artifacts from the EEG data by means of the Reblinca algorithm[32]. Each EEG epoch with amplitude higher than ±80 µV or the slope trend higher than 3 was removed in order to have an artifact-free EEG dataset.

2.3 Mental workload Computation

From the artifact-free EEG dataset, the Power Spectral Density (PSD) was calculated for each EEG epoch using a Hanning window of 2 seconds with a buffer of 125 ms. Then, the EEG frequency bands were defined accordingly with the Individual Alpha Frequency (IAF) value estimated for each subject. The alpha peak has been obtained before starting with the experiment asking to the subject to keep his eyes closed for one minute, because the alpha peak is maximum during this condition. In particular, the theta (IAF-6 ÷ IAF-2) band has been defined. The Mental workload has been computed as the average of the PSD in theta band over the frontal electrodes. The difference respect to the baseline has been considered.

2.4 Performed Analysis

Wilcoxon signed-rank test [33] has been performed to assess the difference between the total average of the familiarity of both Foreign and Local products, and to compare the experienced mental workload during the tasting of both Foreign and Local products.

Two different Pearson correlation analysis[34] has been performed between the average mental workload during the taste experience and the declared average familiarity towards the products before (during the products' observation) and after the tasting.

3 Results and discussion

3.1 Declared results

In figure 2 the results on the declared judgments of participants before the taste experience-during the observation- showed a significant less familiarity for foreign products than local ones (p=0,0156) in the sight perception.

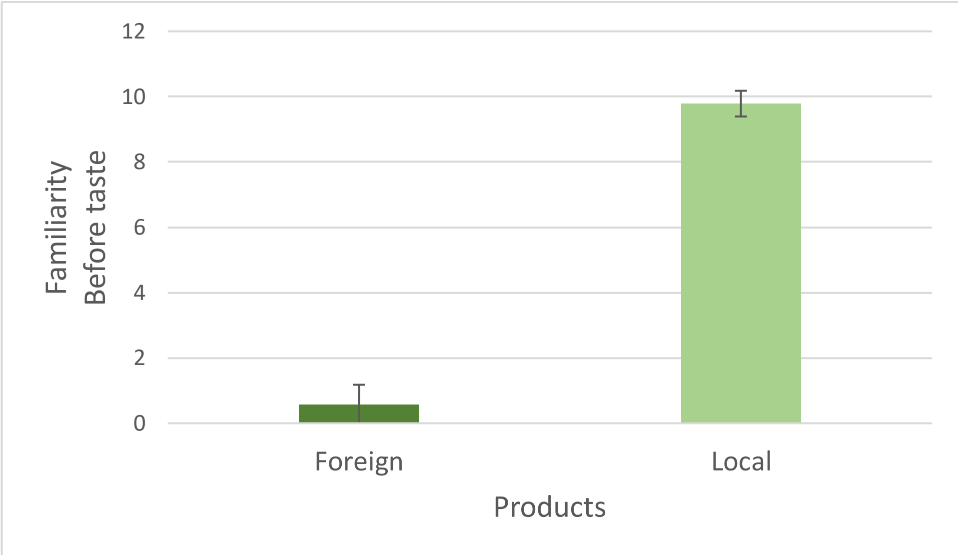


Figure 2. The graph shows the average declared familiarity values reported participants before the taste of the products. Error bars represent standard error.

Results on the declared judgments of participants after the taste experience in figure 3 showed a significant less familiarity for foreign products than local ones (p=0,0156) in the taste perception.

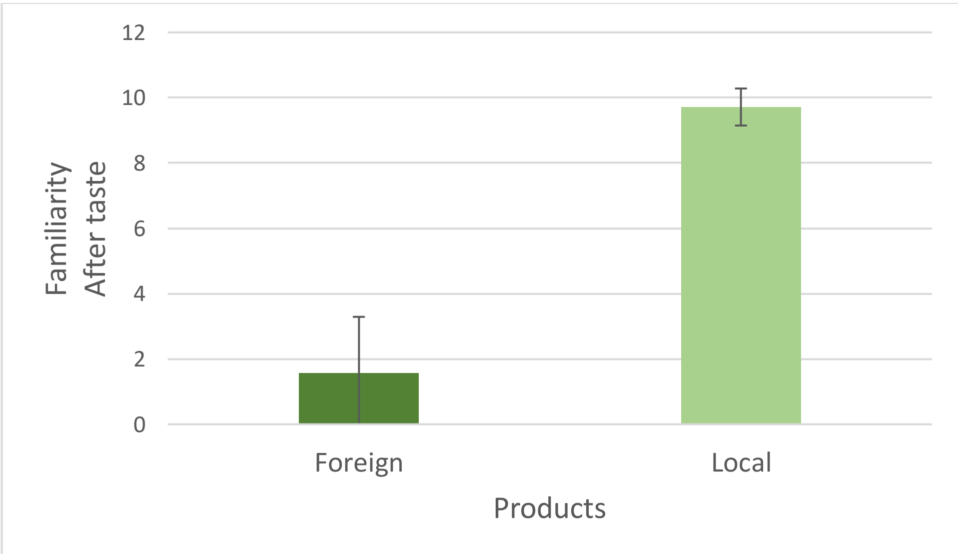


Figure 3. The graph shows the average declared familiarity values reported by the participants after the taste of the products. Error bars represent standard error.

3.2 Mental workload results

Results on the mental workload of participants during the taste of foreign products reported a significant higher mental workload than for the local ones (p=0,0156) (Fig. 4).

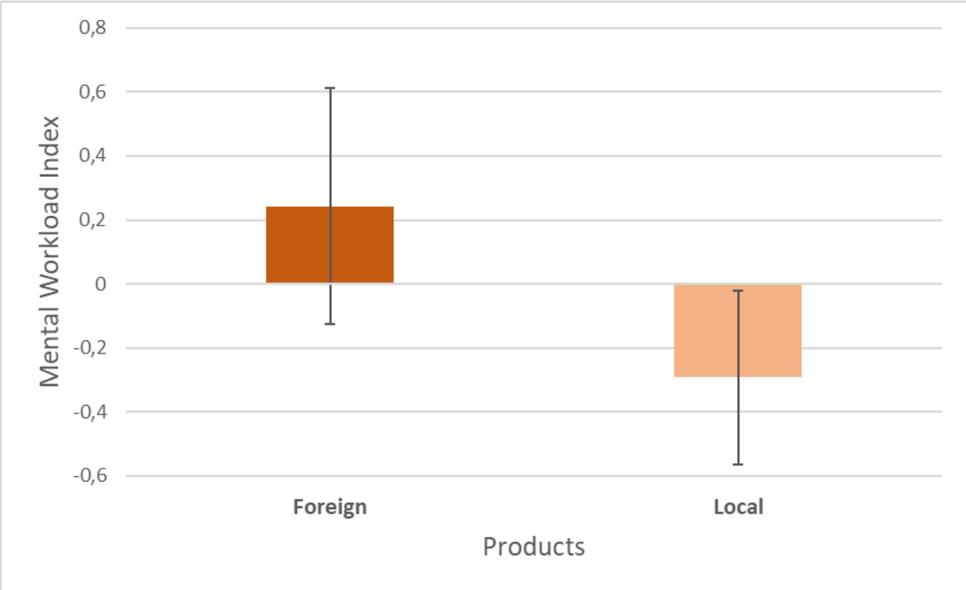


Figure 4. The graph shows the average mental workload during the taste of the products. Error bars represent standard error.

3.3 Mental workload and declared correlation results

The results showed a considerable negative correlation of the mental workload during the taste of products and their familiarity consumer perception before and after the taste. Figure 5 shows the significant negative correlation between the familiarity with the product before the taste (sight) and the mental workload during the taste (R= -0.6408; p= 0.0135). Figure 6 shows the significant negative correlation between the familiarity with the product after the taste (taste) and the mental workload during the taste (R= -0.5502; p=0.0415).

Figure 5. Correlation between Mental Workload and Familiarity before the taste.

Figure 6. Correlation between Mental Workload and Familiarity after the taste.

3.4 Discussion

The selected foreign products of the study had very different extrinsic features, such as color and shape, from the traditional Italy food. It enables consumers to recognize on a first contact with products their unfamiliarity, as the visual aspects were not recognized by previous models. This first impression could be considered as an expectative to the flavor. In fact, after the tasting experience, the difference between the declared familiarity of both groups of products (foreign and local) was also significative. This fact confirms that not only the sight but also the taste was not recognized. In the measurement of the brain activity during the taste interaction with the products, participants knew that after the taste they would be asked some questions. Therefore, during the taste experience they tried to recognize the flavor that they were tasting. On the one hand, the results of the declared familiarity after the taste show the unfamiliarity with foreign products. On the other hand, the results of the mental workload index show that this process requires higher frontal brain activity. These results confirm what previous literature says about the relationship of an unfamiliarity product with the consumer mental workload[5] and about the perception of the aesthetic experience (in this case considered as the observation before the taste): it is significantly modulated by the previous specific knowledge experienced by the participants[26]. The added value of this study lies in the innovation of the application of this cognitive index during a taste experience. Moreover, the correlation of the mental workload index and the familiarity shed light for food practitioners and different fields researchers. The insertion of products in new markets requires a high investment for companies, therefore a correct understanding of the consumers' brain processes against unfamiliar products could shed light on how to design the products based on the market where companies want to enter. Also, when chefs create new products, they should be aware that unfamiliar foods will elicit different brain responses in consumers. In the academic field, these results can be applied to different topics, such as the multisensory(sight-taste) interaction with products; the mental workload index application on a taste experience and the (un)familiar relation with mental workload. Finally, further research should be done with a group of participants that usually consume foreign products (considered as *experts*), in order to test if the workload index is still modulated as in this study with "non experts".

References

[1] R. S. Herz, "Birth of a Neurogastronomy Nation : The Inaugural Symposium of the International Society of Neurogastronomy," vol. 41, pp. 101–103, 2016.

[2] G. M. Shepherd, *Neurogastronomy: how the brain creates flavor and why it matters*. Columbia University Press, 2011.

[3] M. L. Kringelbach, "REVIEW FOOD FOR THOUGHT : HEDONIC EXPERIENCE BEYOND HOMEOSTASIS IN THE HUMAN BRAIN," vol. 126, pp. 807–819, 2004.

[4] C. Spence and X. Wan, "Beverage perception and consumption: The influence of the container on the perception of the contents," *Food Qual. Prefer.*, vol. 39, pp. 131–140, 2015.

[5] E. Modica *et al.*, "Neurophysiological Responses to Different Product Experiences," *Comput.*

Intell. Neurosci., vol. 2018, 2018.

[6] D. Giacalone and S. R. Jaeger, "Better the devil you know ? How product familiarity affects usage versatility of foods and beverages," *J. Econ. Psychol.*, vol. 55, pp. 120–138, 2016.

[7] H. Tuorila, A. Arvola, and L. Lähteenmäki, "Predicting the Intent to Purchase Unfamiliar and Familiar Cheeses: The Effects of Attitudes, Expected Liking and Food Neophobia," *Appetite*, vol. 32, no. 1, pp. 113–126, 1999.

[8] H. S. G. Tan, A. R. H. Fischer, P. Tinchan, M. Stieger, L. P. A. Steenbekkers, and H. C. M. van Trijp, "Insects as food: Exploring cultural exposure and individual experience as determinants of acceptance," *Food Qual. Prefer.*, vol. 42, pp. 78–89, 2015.

[9] J. Prescott, "Comparisons of taste perceptions and preferences of Japanese and Australian consumers: overview and implications for cross-cultural sensory research," *Food Qual. Prefer.*, vol. 9, no. 6, pp. 393–402, 1998.

[10] E. Iannilli, "Studying human gustation: A multidisciplinary approach," *J. Neurosci. Res.*, vol. 97, no. 3, pp. 228–229, 2019.

[11] F. Grabenhorst, E. T. Rolls, and A. Bilderbeck, "How cognition modulates affective responses to taste and flavor: top-down influences on the orbitofrontal and pregenual cingulate cortices," *Cereb. Cortex*, vol. 18, no. 7, pp. 1549–1559, 2007.

[12] M. L. Kringelbach, I. E. T. de Araujo, and E. T. Rolls, "Taste-related activity in the human dorsolateral prefrontal cortex," *Neuroimage*, vol. 21, no. 2, pp. 781–788, 2004.

[13] K. Ohla, N. A. Busch, and J. N. Lundström, "Time for taste—a review of the early cerebral processing of gustatory perception," *Chemosens. Percept.*, vol. 5, no. 1, pp. 87–99, 2012.

[14] I. E. T. De Araujo, E. T. Rolls, M. L. Kringelbach, F. McGlone, and N. Phillips, "Taste-olfactory convergence, and the representation of the pleasantness of flavour, in the human brain," *Eur. J. Neurosci.*, vol. 18, no. 7, pp. 2059–2068, 2003.

[15] L. N. van der Laan, D. T. D. de Ridder, M. A. Viergever, and P. A. M. Smeets, "Appearance matters: Neural correlates of food choice and packaging aesthetics," *PLoS One*, vol. 7, no. 7, 2012.

[16] P. Cherubino *et al.*, "Wine Tasting: How Much Is the Contribution of the Olfaction?," in *International Conference on Computational Methods in Experimental Economics*, 2017, pp. 199–209.

[17] M. Caratù, P. Cherubino, and A. Mattiacci, "APPLICATION OF NEURO-MARKETING TECHNIQUES TO THE WINE TASTING EXPERIENCE."

[18] G. Di Flumeri *et al.*, "EEG-based Approach-Withdrawal index for the pleasantness evaluation during taste experience in realistic settings," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, 2017, pp. 3228–3231.

[19] R. J. Davidson, P. Ekman, C. D. Saron, J. A. Senulis, and W. V Friesen, "Approach-withdrawal and cerebral asymmetry: Emotional expression and brain physiology: I.," *J. Pers. Soc. Psychol.*, vol. 58, no. 2, p. 330, 1990.

[20] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: A review and analysis," *Brain Research Reviews*, vol. 29, no. 2–3. pp. 169–195, 1999.

[21] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neurosci. Biobehav. Rev.*, vol. 44, pp. 58–75, 2014.

[22] K. A. Brookhuis and D. De Waard, "The use of psychophysiology to assess driver status," *Ergonomics*, vol. 36, no. 9, pp. 1099–1110, 1993.

[23] D. Kahneman, *Attention and effort. Englewood Cliffs, NJ.* 1973.

[24] C. L. Baldwin and J. T. Coyne, "Mental Workload as a Function of Traffic Density: Comparison of Physiological, Behavioral, and Subjective Indices," pp. 19–24, 2018.

[25] K. A. Brookhuis, C. J. G. van Driel, T. Hof, B. van Arem, and M. Hoedemaeker, "Driving with a congestion assistant; mental workload and acceptance," *Appl. Ergon.*, vol. 40, no. 6, pp. 1019–1025, 2009.

[26] G. Cartocci *et al.*, "The 'NeuroDante Project': Neurometric Measurements of Participant's Reaction to Literary Auditory Stimuli from Dante's 'Divina Commedia,'" in *International Workshop on Symbiotic Interaction*, 2016, pp. 52–64.

[27] G. Borghini *et al.*, "EEG-Based Cognitive Control Behaviour Assessment: an Ecological study with Professional Air Traffic Controllers," *Sci. Rep.*, vol. 7, no. 1, p. 547, Apr. 2017.

[28] N. Sciaraffa *et al.*, "Brain interaction during cooperation: Evaluating local properties of multiple-brain network," *Brain Sci.*, vol. 7, no. 7, p. 90, 2017.

[29] M. G. Wisniewski, E. R. Thompson, N. Iyer, J. R. Estepp, M. N. Goder-Reiser, and S. C. Sullivan, "Frontal midline θ power as an index of listening effort," *Neuroreport*, vol. 26, no. 2,

pp. 94–99, 2015.

[30] A. Gevins and M. E. Smith, "Neurophysiological measures of cognitive workload during human-computer interaction," *Theor. Issues Ergon. Sci.*, vol. 4, no. 1–2, pp. 113–131, 2003.

[31] F. P. da Silva, "Mental Workload, Task Demand and Driving Performance: What Relation?," *Procedia - Soc. Behav. Sci.*, vol. 162, no. Panam, pp. 310–319, 2014.

[32] G. Di Flumeri, P. Aricó, G. Borghini, A. Colosimo, and F. Babiloni, "A new regression-based method for the eye blinks artifacts correction in the EEG signal, without using any EOG channel," in *Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the*, 2016, pp. 3187–3190.

[33] R. F. Woolson, "Wilcoxon signed-rank test," *Wiley Encycl. Clin. trials*, pp. 1–3, 2007.

[34] J. Benesty, J. Chen, Y. Huang, and I. Cohen, "Pearson correlation coefficient," in *Noise reduction in speech processing*, Springer, 2009, pp. 1–4.

Graduated development training for driving license acquisition (Category B)

Manuel Picardi

[1] Dept. of Civil Engineering, Sapienza University of Rome, Italy
manuel.picardi@gmail.com

Abstract. Directive 2006/126/EC of the European Parliament and of the Council of the European Union details the minimum requirements for driving examiners and the test that needs to be taken before the issue of a driving license in member countries. However, there in no directive concerning the training of learner drivers or of driver trainers/instructors. Consequently, there are different systems of training carried out in every EU country and there is no scientific evidence to show that one training system is more effective than another. This means that training systems are judged depending on the number of lessons students need to pass their driving test in each country and the pass rate for tests in that country. A research project was devised in Italy involving 272 driving instructors. These instructors were provided with a bespoke training course aimed at reducing the pressure on their students and on themselves. The research demonstrated that the adoption of the specially designed training program was more effective in facilitating the learning of practical driving skills and was also less stressful for the driver trainer and the learner driver. The report below details this new and innovative training procedure and the results of this exercise.

Keywords: workload, simulators, driving school, driving training, progressive access training

1. Introduction

Car driving is considered a very complex activity, consisting of different concomitant tasks and subtasks that require a very high mental and physical coordination [1], [2] and moreover the surrounding environment is very unpredictable and out of control [3]. Within human factors research, a distinction has been made between *driver performance*, which reflects what a driver can do, based on his or her physical and mental capabilities, and *driver behavior*, which involves what a driver actually does and is influenced by social factors and self-motivations [4]. Despite the conceptual difference, it is often hard to distinguish driver behavior and performance, because of a strong inter-relation.

This conflict is already present while learning to drive. Most people who register for a driving school do so in order to obtain a driving license, not to learn how to drive properly and safely. The resulting situation, aiming exclusively at achieving that goal, generates a very high level of performance stress, because of a mismatching between students expectation and real actions.

The first hours of driving are, in my experience and according to anecdotal evidence, often the most complicated and frustrating for the students who are not used to the length of time required to learn a discipline as complex as driving. People could feel anxious before their driving lesson and sometimes they could also fail in their performance. This feeling people have before a trial is called “performance anxiety”, and it's a state of tension and fear felt when someone has to face an assessment situation of his capacity as a driving lesson or a driving test [5]. Starting from the hypothesis that emotions could impact cognitive processes [6] anxiety can have negative effects on people’s behavior during a demanding task, for instance profoundly affect the cognitive performance and so the learning process. This is the reason why performance anxiety is demonstrated to affect the driving behavior [7].

A progressive access training system used in the first few driving lessons should allow students to focus on each step of learning the different skills required individually, as often happens in many sports where motor coordination is required.

The first driving lessons are stressful for the driving instructors who have to manage the anxieties and sudden movements of the students. Thanks to the active use of the dual controls (in Italy clutch, brake and accelerator), the instructor manages the driving of the vehicle in a smoother way and focuses the attention on the dynamics of the traffic from the first few lessons. However, it is hard to check at the same time the quality of each movement performed by the students over the car commands. At the same time, the student has to learn and perform different movements and actions simultaneously, at a cost of a high workload, even resulting in a poor training. Experimental studies already highlighted how the execution of multiple instructional sequences appears to be much more complex and multilayered than what is suggested by the traditional description of instruction in terms of Initiation–Response–Evaluation (IRE) sequences [8].

In this context, the study aimed to evaluate how an innovative method of training based on a progressive access training protocol would improve driving training effectiveness reducing at the same time the cognitive workload level requested to the students and the instructors.

2. Materials and Methods

With the aim of reducing the cognitive workload, in accordance with experience gained in driver training in Switzerland [9], it was decided to use a progressive approach teaching method with the following characteristics

1. Create and consolidate the basic motor schemes;
2. Create and consolidate the coordination skills necessary for training
3. Optimize the ministerial time available to the candidate for training
4. Check the actual learning times of the exercises proposed in the individual driving lessons
5. Create a team of specialized instructors able to carry out specific lessons

6. Help the driving instructors to create the necessary skills for the students to be autonomous in the shortest possible time, without generating further workload.

2.1 The proposed method

The method used during the first few driving lessons consists of four separate training units. The application of the individual training units simplifies the workload for each student, since the student will face separately each single fundamental movement, i.e. observation and steering wheel control, right foot use for accelerator and brake pedals, right foot use for the gearbox.

The first training unit consists in learning the correct use of the steering wheel and the importance of all-round observation and control (Figure 1), considering the 70 % to be allocated for car direction, 15 % for infrastructure observation and the remaining 15 % for local scanning (check).

FIRST UNIT

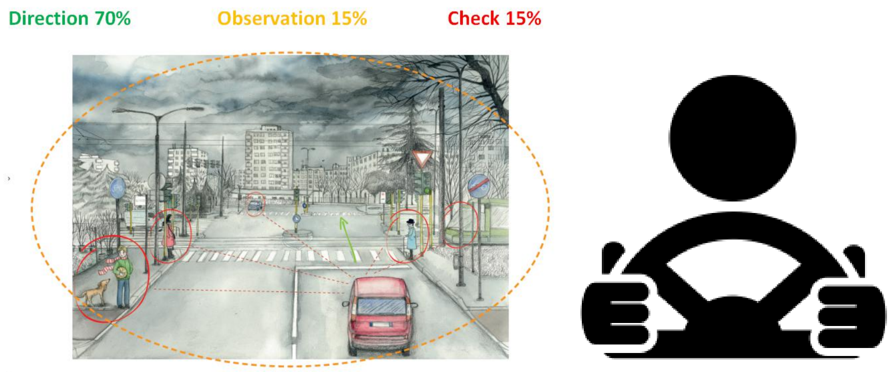


Figure 1. On the left, a graphical representation regarding how to distribute the all-round observation while driving. On the right, the symbol of the steering wheel, representing the fundamental control practiced during the first session.

In the second unit (Figure 2) the student, while putting into practice what they learned in the first unit, learns to use the right foot in the management of the accelerator and the brake.

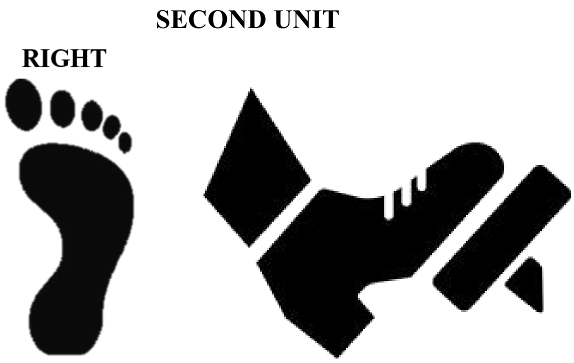


Figure 2. The symbol of the pedals managing through the right foot, representing the fundamental control practiced during the second session.

The third unit (Figure 3) introduces the use of the left foot and the right hand (gear lever) in addition to the commands previously learned.

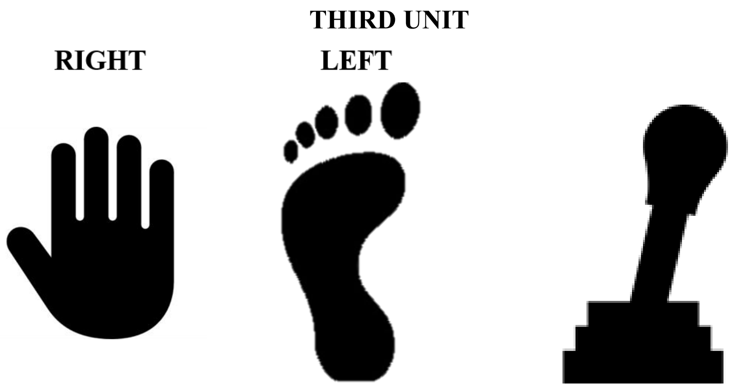


Figure 3. The symbol of the gear lever and the involved limbs, representing the fundamental control practiced during the third session.

In the last unit (Figure 4), the total coordination of the student is managed, and the commands are completely passed to him.

FOURTH UNIT

Figure 4. During the fourth session the student will have the full control of the car.

2.2 The study

From 2015 to 2017, voluntary training courses were held for about 300 driving instructors between 30 and 55 years of age, lasting 40 hours. In June 2019 the course participants were asked to answer a short questionnaire on the effectiveness of the progressive access training methodology during the first driving lessons.

The aim of the questionnaire was to understand the real effectiveness of the system in carrying out daily training activities, or if the application of the progressive access instruction methodology had:

- Reduced the cognitive workload of the students through the separation of the commands given by the Instructor;
- Reduced the time needed to assimilate the driving skills of the students, thus being able to dedicate more resources to focusing on the problems associated with circulation;
- Decreased the level of workload by the Instructor in the management of the first driving lessons.

In particular, the subjects were asked to indicate their level of agreement (Not at all – full disagreement -, Little, Enough, High) with the following 6 items by means of a 4-way Likert Scale [10]:

Question 1: Did the gradual explanation of the controls reduce the workload of the students during the first driving lessons?

Question 2: Did the gradual explanation of the commands speed up the learning of automatisms by the students during the first driving lessons?

Question 3: Has the gradual distribution of the controls contributed to raising awareness among students of road traffic issues since the first driving lessons?

Question 4: Do you apply the progressive access instruction methodology during the first driving lessons of your students?

Question 5: Has the use of the techniques provided by the progressive method reduced its workload as an Instructor during the first driving lessons?

Question 6: Overall, the use of the progressive methodology during the first driving lessons has changed the quality of your work?

3. Results

Below are the results obtained for each item:

Question 1: Did the gradual explanation of the controls reduce the workload of the students during the first driving lessons?

Not at all	0.4
Little	0.7
Enough	9.6
High	89.3

Question 2: Did the gradual explanation of the commands speed up the learning of automatisms by the students during the first driving lessons?

Not at all	0.4
Little	0.4
Enough	13.2
High	86

Question 3: Has the gradual distribution of the controls contributed to raising awareness among students of road traffic issues since the first driving lessons?

Not at all	0.7
Little	3.3
Enough	16.2
High	79.8

Question 4: Do you apply the progressive access instruction methodology during the first driving lessons of your students?

Not at all	0.7
Little	0.7
Enough	5.5
High	93

Question 5: Has the use of the techniques provided by the progressive method reduced its workload as an Instructor during the first driving lessons?

Not at all	0.4
Little	1.8
Enough	11.8
High	86

Question 6: Overall, the use of the progressive methodology during the first driving lessons has changed the quality of your work?

Not at all	1.1
Little	0.4
Enough	11.4
High	87.1

4. Discussions & Conclusion

In Europe, initial and periodic training for driver trainers is fragmented, in many cases un-regulated and is not always mandatory. A survey was carried out in 22 countries. Only in 14 there is an initial training course which can vary from 2 months to 2 years of attendance. For example, in Italy the course is mandatory and consists in 120 hours of theoretical and practical lesson [11].

The situation of periodic training is much more serious, in fact across 22 countries, only in 9 of them a periodic training course aimed to maintain skills is taken into consideration. The mandatory introduction of a progressive access training methodology through an EU Directive would be well received in the world of driving training. The results of the questionnaire show a very high approval rating for those who have voluntarily submitted to the course for improving their skills.

The professional driving training activity has undergone profound changes in recent years. The significant increase in the level of traffic in European metropolises [12], associated with a lower motivation of the population residing in large urban areas to obtain a driving licence [13], has led to greater difficulty in providing driving training. The introduction of a system that can reduce the cognitive workload for students, guaranteeing them less time and less effort to generate the motor coordination automatisms necessary to be able to drive safely, was highlighted by the results obtained from the proposed questionnaire.

The result associated with the reduction in the workload achieved by the students during the driving lessons is significant (89.3%), as is the result obtained in reducing the time required to complete the automation process of the basic controls, associated to motor coordination (86%). Although still exceptional, despite being lower than the other results obtained, the figure associated with raising awareness of the other problems associated with road traffic (79.8%). The results associated with the daily application frequency of the training system (93%) are also very interesting, undoubtedly due to the considerable reduction in workload for the driving instructor

(86%). Overall the use of a progressive approach methodology is judged more than positively (87%) by the driving instructors who participated in the study.

The result obtained from this work highlights the need for new research in the field of driving training, taking advantage of the theme of progressive learning to drive, so that further applicable solutions can be found, globally, able to further reduce the load of work for the people involved in the delicate educational path in question.

Acknowledgements

I would like to thank the researchers of the EU-funded SIMUSAFE Project for scientific support. I would like to thank "Scuola Nazionale Istruttori di Guida" (Italian Driving Instructors School) for data collection support. I also thank John Lepine, the former EFA president (European Federation of Driving Instructors) for reviewing the text.

References

[1] D. Parker and A. S. R. Manstead, "The social psychology of driver behaviour," in *Applied social psychology*, G. R. Semin and K. Fiedler, Eds. London: Sage, 1996, pp. 198–224.

[2] De Waard, "The measurement of drivers' mental workload," University of Groningen, Traffic Research Centre, Netherlands, 1996.

[3] Jian-you Z., Xiao-fen S., Liang Z., Shuang-xi Z., Xi-yang N. (2013) Driving Behavior Theory and Computer Simulation System of Driver's Risk Perception Based on 3D. Procedia Social and Behavioral Science, 96, pp. 1686-1695

[4] L. Evans, *Traffic Safety and the Driver*. Science Serving Society, 1991.

[5] Spielberger, C. D. (1972). Anxiety as an Emotional State, in C. D. Spielberger (ed.) Anxiety: Current Trends in Theory and Research. New York: Academic Press.

[6] Lazarus, R.S., (1991). Emotion and Adaptation. Oxford University Press, New York.

[7] Onyekuru, B.U., Ibegbunam, J.O. (2014). Relationships among test anxiety, locus of control and academic achievement among college students. European Scientific Journal. vol.10 (13)

[8] E. De Stefani and A.-D. Gazin, "Instructional sequences in driving lessons: Mobile participants and the temporal and sequential organization of actions," *J. Pragmat.*, vol. 65, pp. 63–79, May 2014.

[9] Maestro conducente (APF)/Maestra conducente (APF), orientamento.ch, 2019 downloaded: https://www.orientamento.ch/dyn/show/1900?id=773

[10] I. E. Allen and C. A. Seaman, "Likert Scales and Data Analyses," *Qual. Prog. Milwaukee*, vol. 40, no. 7, pp. 64–65, Jul. 2007.

[11] D.M. N°17 26.01.2011downloaded: www.mit.gov.it/mit/mop_all.php?p_id=9725

[12] ETSC Safer Roads, Safer Cities: How to improve urban road safety in the EU (PIN Flash 37), 2019, downloaded: https://etsc.eu/safer-roads-safer-cities-how-to-

improve-urban-road-safety-in-the-eu-pin-flash-37/
[13] Ministero delle Infrastrutture e dei Trasporti Direzione Generale per la Motorizzazione Centro Elaborazione Dati Gestione del Patrimonio Informativo ESITI ESAMI 2018 (2019) downloaded: http://www.mit.gov.it/documentazione/risultati-degli-esami-per-il-conseguimento-della-patente

Neurophysiological characterization of normal hearing and unilateral hearing loss children: a comparison among EEG-based indices for information processing and decision-making levels

Giulia Cartocci[1,2], Gianluca Borghini[1,2], Alessandro Scorpecci[3], Bianca M.S. Inguscio[1], Andrea Giorgi[1], Sara Giannantonio[3], Nicolina Sciaraffa[1,2], Alessia Vozzi[2], Vincenzo Ronca[2], Pasquale Marsella[3] and Fabio Babiloni[1,2,4]

[1] Dept. of Molecular Medicine, Sapienza University of Rome, Italy
[2] BrainSigns srl, Rome, Italy
[3] Department of Otorhinolaryngology, Audiology and Otosurgery Unit, Bambino Gesù Paediatric Hospital, Rome, Italy
[4] Department of Computer Science, Hangzhou Dianzi University, Hangzhou, China

Abstract. The identification of measurable indices of cerebral functions to be applied in clinical settings is ever more felt as necessary for a more thorough and objective evaluation of patients cognitive performance. In the present paper, the electroencephalographic-based indices of mental workload (WL = frontal θ/parietal α) and of mental engagement (ME = β/(α+θ)), calculated along the brain midline, have been employed to characterize the eventual specific patterns of cerebral activations during a speech in noise perception task in normal hearing (NH) and unilateral hearing loss (UHL) children. Results showed no differences between the groups for the frontal bilateral noise condition (in which both signal and noise were emitted by two loudspeakers placed +45° and -45° in relation to the participant), while in lateralized noise conditions the UHL group showed higher parietal ME values for the Noise to the Deaf Ear condition). Finally, the NH group showed a different distribution of ME values among frontal, central and parietal electrodes, with higher ME values in the central and parietal ones in correspondence of the Noise to the Left Ear condition. The WL index analysis did not provide any significant differences. Results suggest the relevance of including the analysis of the beta rhythm in the neurophysiological assessment of the neural processing of speech in noise stimuli in normal hearing and hearing impaired participants.

Keywords: workload, engagement, EEG, deafness, unilateral hearing, speech in noise perception

1. Introduction

The current need of the identification of objective measures of cerebral processing functions such as the listening effort in clinical (audiological) settings is witnessed by recent articles mentioning in the title words like "objective assessment", "objective measures” e.g. [1], [2]. In the scientific literature, well-known electroencephalographic (EEG) indices of cerebral processing have been successfully applied to various fields. In particular, in the present study two EEG-based indices have been selected for the neurophysiological characterization of normal hearing (NH) and unilateral hearing loss (UHL) children: the Mental Workload index (WL) and the Mental Engagement index (ME) [3], [4], [5]. The selected protocol for the application of such indices investigation was the *word-in-noise recognition task*, since it allows to elicit the effort necessary to discriminate the speech from the background noise [6]. The preliminary hypothesis is that this task should be characterized by a certain workload, but it is still debated if an eventual partial, i.e. unilateral, deafness would impact mental workload and engagement in performing the same task [28]. The ME index was developed in the framework of cognitive tasks assessment by Pope and colleagues [5]. The background for the development of such index was constituted by evidences that increases in beta activity would reflect a higher degree of alertness and greater engagement in the task, while increases in alpha and/or theta activity would reflect less alertness and decreased task engagement/information processing [5], [7], [8]. ME index was employed for the assessment of performance improvements during a vigilance task [9], or it has been applied to an educational setting for the assessment of the modulation exerted by emotions on learning, resulting to be predictive of the performances [10]. The WL index has been validated in several operative environments, such as in aircraft pilots, air traffic controllers, and car drivers [5], [11]-[13]. Such studies described that an increase of the frontal EEG power spectra in the theta band (4-7 Hz) and a simultaneous decrease in the parietal EEG power spectra in the alpha band (8-12 Hz) have been observed when the mental workload demand increases. Furthermore, WL index has been already calculated in adult deaf adult patients using a cochlear implant, for the comparison among sound processors and noise reduction filters use during a word in noise recognition task, evidencing that WL values were lower when the noise reduction filter function was employed [14], [15]. In addition, during the same word in noise recognition protocol as the one adopted in the present paper, WL index has been previously employed in hearing impaired children, but in patients presenting an asymmetric hearing loss in the two ears [16]. Expected results will show specific patterns describing the two experimental groups, possibly reflecting different neural strategies developed as a consequence of the neuroplasticity that coped with the naturally lateralized condition of the UHL group’s auditory system. In the light of such intrinsic lateralization, it has been chosen to investigate only midline electrodes, in order to avoid any potential cerebral activity lateralization influence.

2. Materials and Methods

2.1 Participants

Participants were divided in two groups: UHL children (n=13; 6f, 7m; age: 10 ± 2,2), characterized by profound deafness in one ear (average threshold for pure tone frequencies 250-4000 Hz ≥ 90 dB HL) and normal hearing in the controlateral ear (average threshold for pure tone frequencies 250-4000 Hz ≤ 20 dB HL), and NH children (n=12; 6f, 6m; age: 12 ± 2,5). Detailed information about the study were given to all participants and participants' parents were given and signed an informed consent. Participants were volunteers, who did not receive any compensation from taking part in the study. The experiment was performed in accordance to the principles outlined in the Declaration of Helsinki of 1975, as revised in 2000, and it was approved by the institutional Ethic Committee.

2.2 Protocol

The protocol has already been employed in previous studies concerning cochlear implant candidates [16] and cochlear implant users [17], [18] and consisted in a forced-choice word recognition task, with four conditions: Quiet, frontal Bilateral Noise (2 loudspeakers placed at +45° and -45° emitting both noise and signal), Noise emitted from a loudspeaker at +90° and frontal signal, Noise emitted from a loudspeaker at -90°. Stimuli were Italian disyllabic words from "Audiometria Vocale GNResound" [19], delivered free-field at 65 dB SPL. The background noise was emitted continuously with a signal-to-noise-ratio of 10 and was a babble noise. Before the beginning of the experimental tasks, participants were asked to look at a black screen for 60 seconds, and the corresponding EEG activity was used for the Individual Alpha Frequency (IAF) estimation [20]. Each experimental condition (Quiet and noise conditions) was constituted by 20 trials. Trials corresponded to 20 words randomly delivered, and after listening to each word (information processing phase), participants had to indicate the correct word-stimulus among four options (decision making phase), each of them appearing in a different box on the screen. The target word had 25% probability of appearing in one of the four positions on the screen (top left, bottom left, top right, bottom right). The information processing phase lasted about 2 seconds and the decision making phase lasted up to 5 seconds, depending on the individual response time, resulting in a trial of up to 8-second total length.

2.3 EEG recording and processing

A digital medical EEG system (Bemicro EBNeuro, Italy) was used to record the brain activity from 19 channels (Fp1, Fp2, Fz, F3, F4, F7, F8, Cz, C3, C4, T3, T4, Pz, P3, P4, T5, T6, O1, O2), with a sampling frequency of 256 (Hz). Impedances were kept below 10 kΩ, and a 50- Hz notch filter was applied to remove the power interference. A ground and a reference electrode were placed on the forehead. EEG recordings were filtered with a 4th order Butterworth band pass filter (1-40 Hz), so to reject continuous components as well as high-frequencies interferences, such as muscular artifacts. The Fp1 channel has been used to remove eye-blink contributions from each channel of the EEG signal by using the REBLINCA algorithm [21]. In order to remove other kinds of artifacts, specific procedures of the EEGLAB toolbox [22] have been employed. Firstly, the EEG signal was segmented into epochs of 2 seconds (Epoch length) shifted of 0.125 seconds (Shift). The present windowing was chosen so to obtain a high number of observations in comparison to the number of variables, and to respect the stationarity condition of the EEG signal [23]. EEG epochs with signal amplitude exceeding ±100 μV (Threshold criterion) were considered artifacts. After that, the EEG signal within each epoch was interpolated to estimate the slope within the considered epoch (Trend estimation). If such a slope was higher than 10 (μV/s), the considered epoch was marked as artifact. Then, the signal sample-to-sample difference (Sample-to-sample criterion) was analysed: wherever a difference in the absolute amplitude was higher than 25 (μV) (i.e. an abrupt non-physiological variation) the EEG epoch was considered an artifact. Finally, all the EEG epochs labeled as "artifact" were removed so to obtain an artifact-free EEG dataset from which estimate the parameters for the analysis. For each participant the IAF was computed on the 60-seconds-long Open Eyes segment [20], so to define the EEG bands of interest specifically for each participant. [4]. In particular, the considered EEG bands and their corresponding definition using the IAF were the following: theta [IAF-6 ÷ IAF-2 Hz], alpha [IAF-2 ÷ IAF+2 Hz] and beta [IAF+2 ÷ +IAF+16 Hz]. Finally, EEG recordings were segmented into trials, corresponding to each word of each experimental condition (Quiet, Bilateral Noise, Noise +90°, Noise -90°). The Power Spectrum Density (PSD) was calculated in correspondence of the different conditions with a frequency resolution of 0.5 Hz. For the purposes of the present study, only the brain midline has been considered (i.e. Fz, Cz and Pz EEG channels), and their corresponding PSDs normalized by using the z-score formula [24] by using the Quiet condition as reference condition.

Mental Workload index (WL). The WL index was defined as the ratio between the EEG PSD in theta band over the midline frontal electrode (Fz) and the EEG PSD in alpha band over the midline parietal electrode (Pz) (Equation 1) [5], [11].

$$WL = PSD\ \theta_{Fz} / PSD\ \alpha_{Pz} \quad (1)$$

Mental Engagement index (ME). The ME index has been defined as the ratio between the activity in the beta band and the sum of alpha and theta activity (Equation 2), as defined by Pope and colleagues [3].

$$ME = PSD\ \beta_x / (PSD\ \alpha_x + PSD\ \beta_x) \quad (2)$$

The ME was calculated for each x electrode, calculating the ratio between the PSD filtered in β band and the sum of the PSD filtered in α and θ band for each of the considered electrodes.

2.4 Statistical analysis

For both the EEG indices (WL and ME), non-parametric statistical analysis was performed on data since the absence of a normal distribution for all the data. For the comparison between groups (UHL and NH) it was used the Mann-Whitney U test;

while the Wilcoxon Matched pairs test was used for the comparison within each group, in order to compare the different lateralized noise conditions (respectively: Noise to the Deaf Ear and Noise to the Hearing Ear for the UHL group; Noise to the Right Ear and Noise to the Left Ear for the NH group). Finally, Friedman ANOVA was used in the comparison among electrodes (Fz, Cz, Pz) among noise conditions within each group.

3. Results

Behavioural results showed that both groups (UHL and NH) reported the highest percentages of correct responses in correspondence of the Bilateral Noise condition. Furthermore, the UHL group performed better in the Noise-to-the-Deaf-Ear condition than in the Noise-to-the-Hearing-Ear condition. In particular, in the latter condition they reached percentages below the commonly adopted clinical threshold of 80% words-recognition for attesting good performances. In contrast, NH group performed similarly in the two lateralized noise conditions (Fig.1).

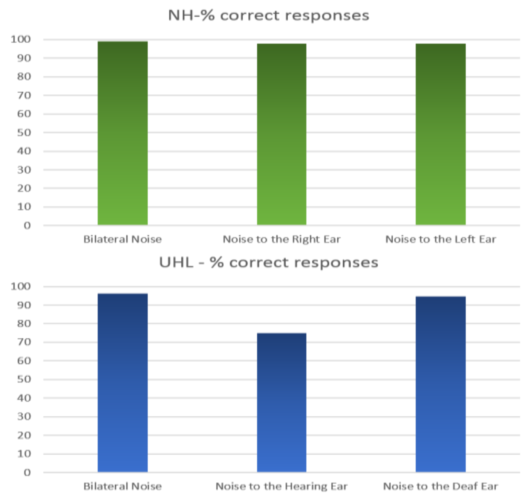


Figure 1. Graph showing the percentages of correct responses in the NH and UHL group in correspondence of the three background noise conditions (Bilateral noise, Noise emitted +90° from the participant and Noise emitted -90° from the participant).

Concerning the ME index, there were no statistically significant differences between the NH and UHL group in the Bilateral-Noise condition in none of the investigated electrodes (all $p>0.05$). Focusing separately on the two experimental groups, main statistical significances have been found in the parietal brain area. In particular, the UHL group showed higher levels of the ME index in the Noise-to-the-Deaf-Ear condition in comparison with the Noise-to-the-Hearing-Ear condition ($T=8.000$, $Z=2.621$, $p=0.009$), in the Pz channel (Fig.2). Moreover, concerning the NH group (Fig.3), the ME index in the Noise-to-the-Left-Ear condition resulted higher moving throughout the brain midline areas, that is from the frontal (Fz) to the central (Cz) and parietal (Pz) electrodes (ANOVA Chi Sqr. (N=12, df=2)=7.167; p=0.028). Concerning the WL index, it did not exhibit any significant difference neither between nor within the UHL and NH groups.

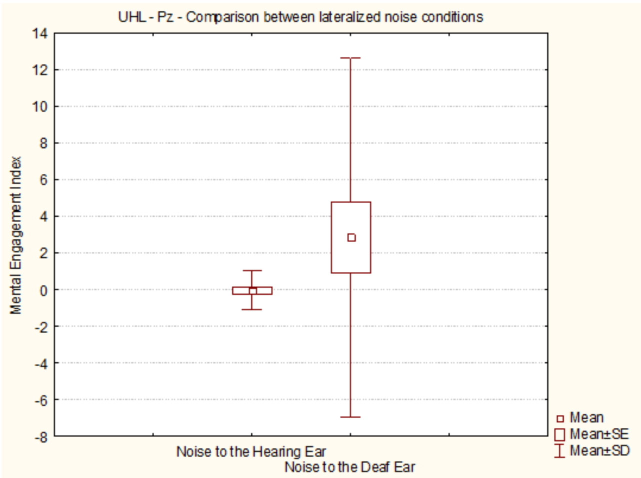


Figure 2. Graph showing the comparison between lateralized noise conditions in the UHL group in Pz electrode, evidencing higher average ME levels in the Noise to the Deaf Ear condition in comparison to the Noise to the Hearing Ear condition (p=0.009).

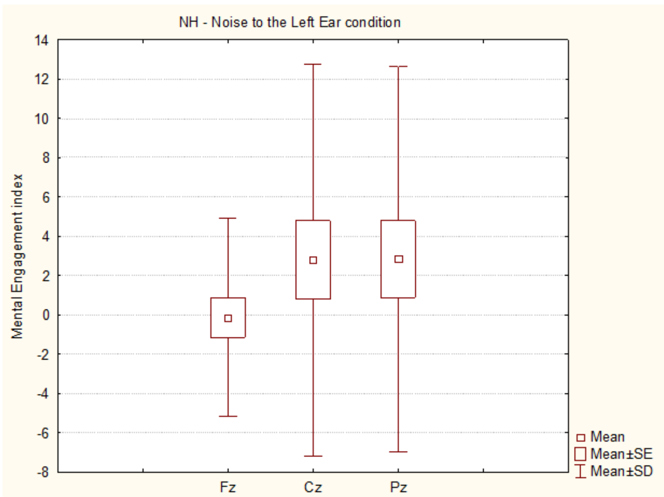


Figure 3. Graph showing the average ME index values among the midline electrodes (Fz, Cz, Pz) in the NH group, evidencing higher ME levels from the frontal toward the central and parietal electrodes (p=0.028).

4. Discussions

The present results suggest that ME index appears to be more sensitive than WL index to highlight differences of brain activations between the NH and the UHL group. It is interesting to note that according to previous studies [16], the hearing impaired group reported a higher level of cognitive processing, as indexed by the ME levels in Pz in the present paper, in the noise to the ear with the worse hearing capability ("Noise-to-the-Deaf-Ear" in the present study) condition than in the noise to the ear with the better hearing capability ("Noise-to-the-Hearing-Ear" in the present study) condition.

This result is also supported by a previous study involving asymmetrical hearing loss children, where the levels of parietal alpha power was employed as index of listening effort [24]. The sum of these evidences supports the hypothesis that such lateralized background noise condition is suitable to highlight different neural patterns adopted by asymmetrical hearing impaired patients (both UHL and asymmetric hearing loss children populations) producing different information processing and decision making indices levels in response to the varying background noise lateralization. In addition, it is worth to underline that between the lateralized noise conditions, the highest percentage of correct responses was obtained in the Noise-to-the-Deaf-Ear condition. This could be explained by the fact that in such condition a higher cognitive demand enabled better performances than the opposite condition. Concerning results obtained for the NH group, the higher ME levels estimated in the Cz and Pz electrodes in comparison to Fz, could be linked to the involvement of midline central and parietal areas in listening effort-related processes [25]. The Noise–to-the-Left-Ear condition, in which the ME levels difference was revealed, could be likely linked to the presence of the core language network mainly in the left brain hemisphere [26], and the presence of noise directed toward the ipsilateral ear to such network would affect it [27]. Finally, both NH and UHL children achieved the highest percentage of correct responses in the Bilateral Noise condition, showing no differences between the groups in terms of ME and WL in such condition.

5. Conclusion

Present results support the presence of specific neural strategies adopted by UHL population in order to cope with words-processing tasks, despite apparently next- to-normal hearing behavioral performances. Further investigations are needed for the validation of such hypothesis.

References

[1] P. J. Schäfer *et al.*, "Evaluation of an objective listening effort measure in a selective, multi-speaker listening task using different hearing aid settings," in

2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, pp. 4647–4650.

[2] Picou Erin M., Moore Travis M., and Ricketts Todd A., "The Effects of Directional Processing on Objective and Subjective Listening Effort," *J. Speech Lang. Hear. Res.*, vol. 60, no. 1, pp. 199–211, Jan. 2017.

[3] A. T. Pope, E. H. Bogart, and D. S. Bartolome, "Biocybernetic system evaluates indices of operator engagement in automated task," *Biol. Psychol.*, vol. 40, no. 1, pp. 187–195, May 1995.

[4] W. Klimesch, "EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis," *Brain Res. Rev.*, vol. 29, no. 2–3, pp. 169–195, Apr. 1999.

[5] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neurosci. Biobehav. Rev.*, vol. 44, pp. 58–75, Jul. 2014.

[6] M. A. Eckert, S. Teubner-Rhodes, and K. I. Vaden, "Is Listening in Noise Worth It? The Neurobiology of Speech Recognition in Challenging Listening Conditions:," *Ear Hear.*, vol. 37, pp. 101S-110S, 2016.

[7] J. F. Lubar, M. O. Swartwood, J. N. Swartwood, and P. H. O'Donnell, "Evaluation of the effectiveness of EEG neurofeedback training for ADHD in a clinical setting as measured by changes in T.O.V.A. scores, behavioral ratings, and WISC-R performance," *Biofeedback Self-Regul.*, vol. 20, no. 1, pp. 83–99, Mar. 1995.

[8] K. Offenloch and G. Zahner, "Computer Aided Physiological Assessment of the Functional State of Pilots during Simulated Flight," FRANKFURT UNIV (GERMANY F R), Sep. 1990.

[9] F. G. Freeman, P. J. Mikulka, L. J. Prinzel, and M. W. Scerbo, "Evaluation of an adaptive automation system using three EEG indices with a visual tracking task," *Biol. Psychol.*, vol. 50, no. 1, pp. 61–76, May 1999.

[10] M. Chaouachi, P. Chalfoun, I. Jraidi, & C. Frasson. (2010, May). Affect and mental engagement: towards adaptability for intelligent systems. In Proceedings of the 23rd International FLAIRS Conference. AAAI Press.

[11] A. Maglione *et al.*, "Evaluation of the workload and drowsiness during car driving by using high resolution EEG activity and neurophysiologic indices," *Conf. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. IEEE Eng. Med. Biol. Soc. Annu. Conf.*, vol. 2014, pp. 6238–6241, 2014.

[12] P. Aricò, G. Borghini, G. D. Flumeri, N. Sciaraffa, A. Colosimo, and F. Babiloni, "Passive BCI in Operational Environments: Insights, Recent Advances, and Future Trends," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1431–1436, Jul. 2017.

[13] G. Di Flumeri, G. Borghini, P. Aricò, A. Colosimo, S. Pozzi, S. Bonelli, A. Golfetti, W. Kong, F. Babiloni: On the use of cognitive neurometric indexes in aeronautic and airtraffic management environments. In Blankertz B., Jacucci G., Gamberini L., Spagnolli A.,& Freeman J. (Eds.) *Symbiotic Interaction* (pp. 45-56). Berlin: Springer International Publishing (2015).

[14] G. Cartocci, A.G. Maglione, D. Rossi, E. Modica, P. Malerba,G. Borghini, ... & F. Babiloni (2016). Applications in cochlear implants and avionic:

examples of how neurometric measurements of the human perception could help the choice of appropriate human-machine interaction solutions beyond behavioral data. PsychNol. J, 14(2-3), 117-133

[15] G. Cartocci, A.G. Maglione, E. Modica, D. Rossi, N. Quaranta, M. Zinfollino, A. Ardito, A. Vozzi, V. Ronca, F. Babiloni. Is the younger the less effortful? An electroencephalographic comparison among consecutive generations of cochlear implant sound processors. International Journal of Bioelectromagnetism Vol. 19, No. 1, pp. 11 - 17, 2017

[16] G. Cartocci *et al.*, "Mental workload estimations in unilateral deafened children," in *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*, 2015, pp. 1654–1657.

[17] G. Cartocci *et al.*, "Alpha and Theta EEG Variations as Indices of Listening Effort to Be Implemented in Neurofeedback Among Cochlear Implant Users," in *Symbiotic Interaction*, 2018, pp. 30–41.

[18] P. Marsella *et al.*, "EEG activity as an objective measure of cognitive load during effortful listening: A study on pediatric subjects with bilateral, asymmetric sensorineural hearing loss," *Int. J. Pediatr. Otorhinolaryngol.*, vol. 99, pp. 1–7, Aug. 2017.

[19] M. Turrini, F. Cutugno, P. Maturi, S. Prosser, F. A. Leoni, and E. Arslan, "[Bisyllabic words for speech audiometry: a new italian material]," *Acta Otorhinolaryngol. Ital. Organo Uff. Della Soc. Ital. Otorinolaringol. E Chir. Cerv.-facc.*, vol. 13, no. 1, pp. 63–77, Feb. 1993.

[20] A. Goljahani, C. D'Avanzo, S. Schiff, P. Amodio, P. Bisiacchi, and G. Sparacino, "A novel method for the determination of the EEG individual alpha frequency," *NeuroImage*, vol. 60, no. 1, pp. 774–786, Mar. 2012.

[21] G. D. Flumeri, P. Aricó, G. Borghini, A. Colosimo, and F. Babiloni, "A new regression-based method for the eye blinks artifacts correction in the EEG signal, without using any EOG channel," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 3187–3190.

[22] A. Delorme and S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.

[23] R. Elul, "Gaussian Behavior of the Electroencephalogram: Changes during Performance of Mental Task," *Science*, vol. 164, no. 3877, pp. 328–331, Apr. 1969.

[24] J.-H. Zhang, T. D. Y. Chung, and K. R. Oldenburg, "A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays," *J. Biomol. Screen.*, vol. 4, no. 2, pp. 67–73, Apr. 1999.

[25] M. G. Wisniewski, E. R. Thompson, and N. Iyer, "Theta- and alpha-power enhancements in the electroencephalogram as an auditory delayed match-to-sample task becomes impossibly difficult," *Psychophysiology*, p. n/a-n/a.

[26] J. Obleser, M. Wöstmann, N. Hellbernd, A. Wilsch, and B. Maess, "Adverse Listening Conditions and Memory Load Drive a Common Alpha Oscillatory Network," *J. Neurosci.*, vol. 32, no. 36, pp. 12376–12383, Sep. 2012.

[27] Y. Shtyrov, T. Kujala, R. J. Ilmoniemi, and R. Näätänen, "Noise affects speech-signal processing differently in the cerebral hemispheres," *Neuroreport*, vol. 10, no. 10, pp. 2189–2192, Jul. 1999.
[28] Alhanbali, S., Dawes, P., Lloyd, S., & Munro, K. J. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. Ear and Hearing, 38(1), e39-e48.

Reducing kitting errors in automotive industry through the reorganization of workstations layout: a mental workload approach

Lisa Jeanson[1 2], J. M. Christian Bastien[1], Alexandre Morais, Javier Barcenilla[1]

[1] Université de Lorraine, PErSEUs (EA 7312), Metz, F-57045, France
{christian.bastien, javier.barcenilla}@univ-lorraine.fr
[2] PSA 2-10 Bd de l'Europe, Poissy, 78092 Yvelines, Cedex09, France
lisa.jeanson@mpsa.com

Purpose and methodology

In 2004, the PSA Group, a French automotive manufacturer, developed the PSA Excellent System. This organizational system is based on the Lean Manufacturing principles and aims at optimizing vehicle production. One of the pillars of this system is the follow up of a "work standards" designed by the methods engineers [1]. In theory, work standards allow for the balancing of shifts, i.e. the organization of tasks that can be performed by operators within a given time period. It also contributes to the elimination of waste by limiting process variability and providing the operator with the right amount of time to perform his tasks and to reduce unnecessary resource consumption [1]. Another way of reducing the risk of assembly errors has been to develop a "kitting" activity in order to prepare in advance the parts to be assembled by the operators. Kitting is a process often used in mixed model assembly, in which different objects demanding different component are assembled, in alternative to continuous supply (also known as line stocking). In this process, the operator (called the 'picker') takes components in stores and sorts them in a 'kit' driven by an Automated Guided Vehicle (AGV). A kit is defined as 'a specific collection of components and/or subassemblies that together (i.e. in the same container) support one or more assembly operations for a given product or shop order' [6]. That is to say, one kit contains just the components that the operators need to assemble one object or, in our case, one part of an object. Each picker follows one AGV, which can drive several kits (in our case, 4 kits) and each part is put in a dedicated location on the kits. A light system, called 'pick to light' assists the operator in choosing the right components in the stores and storing this component in the right kit (kits are differentiated by colors). Even if this process seems simple and straightforward since it does not involve any complex task (to see the light and the color, to take the component, and to put the component in the corresponding kit), many errors occur. For example, from March to April 2018, 324 errors due to kitting were reported on a door assembly lines in Sochaux (France). Those errors create disruptions in assembly lines since operators often have to stop their tasks and leave their workstations to look for the right parts in the kitting area. This questions the major benefits put forward for the use of kitting over line stocking: time efficiency and low risk of error [7,8].

In order to understand these errors, we carried out a detailed analysis of two operators' activity in the kitting area associated to a door assembly line in Sochaux. This kitting area is characterized by 352 parts organized in a 75 meters long hallway in which 7 operators have to go back and forth to prepare 3 different vehicles. This organization is not specifically guided by any rule except from logistic and storage constraints. Each kit contains an average of 40 door components. Each AGV tracks 4 kits differing in terms of color which are only presented on the pick to light (meaning that operators have to know that the first kits on the AGV will receive parts identified by the green color on the store, the second parts indicated by the color purple, the third blue and the fourth yellow). The parts identified by the green and purple colors concern the front doors of the vehicles while the parts identified by the blue and yellow colors concern the back doors. AGV speed varies from 12 to 19 meters per minute. Thus, each operator prepares on the average 160 components in a round of 10 minutes. For the data collection, we combined several methodologies: hierarchical analysis of the prescribed and actual operators' tasks, operators were filmed for two hours and the video recordings were analyze using a behavioral coding system (The Observer XT[2,3])[4,5]. This approach has been used and described in a previous study [6,7].

Findings

The analyses of the data revealed discrepancies between work standards and actual tasks. As described previously, according to standards, pickers must follow the AGV and the pick to light which implies: (1) to identify the parts to be picked based on the state of the lights (switched on), (2) as well as the number of pieces (written on a digit located near the light), (3) to pair the color of the light with the corresponding kit (as described above), (3) to pick the part(s), (4) to switch the light off and (5) to place the part(s) on the kit. However, when the number of components to kit or the speed of the AGV increases, pickers tent to stop it. Similarly, when the number of components to kit or the speed of the AGV decreases, pickers tent to run in front of the AGV. Then, we noted that pickers often change the prescribed order of the components to kit, sorting them by colors (thus by kits) and types of components instead of following the order prescribed by the pick to light. At last, we observed that pickers tent to pick some components a few seconds before the lights are on. The strategy pickers choose aims at keeping ahead of the AGV so as to be able to cope with hazards, to keep up the pace and maintain performance.

These strategies are costly for the pickers because they involve operations, steps, reasoning, that are not taken into account in the design of the workstations. This adds time pressure and forces them to walk and kit faster, leading them to errors. This also requires them to constantly interrupt their work and memorize the color of the kits they have to put the parts in. In these situations, operators' workload increases as well as the risk to make errors. During our observations, we reported 6 situations in which the pickers forgot one or more components and had to stop the AGV to pick them up. We also counted 22 errors that had an impact on production such as the fall of an object, the omission of a task or any error representing a waste of time or a risk to damage the parts.

Differences between prescribed and real tasks may be accounted for by ‘anticipated regulations’, or strategies used collectively by operators to cope with production constraints, i.e. to deliver parts on the assembly lines on time. As we demonstrated in a previous study [6], these regulations seem to appear to compensate flaws in workstations design. In other words, pickers do not kit the way engineers would like them to kit. More precisely, considering time pressure and the level of efficiency expected on workstations, pickers’ action have to be automated to gain time and lower cognitive load [5]. This could explain how they are able to select some parts before their light goes on. That’s also probably why they tent to ritualize the order of the components they kit by sorting them by colors and types. However, to do so, they need a stable and repetitive environment. Yet, we find that the color of the kit, the number of components to kit and AGV speed change constantly and randomly making task intensity impossible to predict. Furthermore, we identified other interruptions in pickers’ work caused by: empty containers (they have to remove), other pickers (when one of them make a mistake, everyone is disturbed) or the pick to light (when the lights don’t switch on or off fast enough).

In other words, each time pickers’ work pace tends to change or each time they are disturbed, they find strategies to cope with the situation. Thus, we make the assumptions that reducing the sources of work variability will allow the operators to perform their tasks automatically and, at the same time, increase their performance and reduce the need for regulations. Consequently, it would confirm that regulations are a symptom of constraints (among which we find mental workload) that are not taken into account during the design of workstations. We identified several sources of variation that we could lower while reorganizing workstations layout: (1) the number of kits to prepare simultaneously, (2) the number of parts to kit by area and (3) the speed of the AGV.

Intervention

Thanks to these analyses, we were able to prescribe a new way of storing components so as to stabilize pickers’ activity. First, we proposed to store parts according to the color of the kit they were to be put in. Hence, when kitters go up to the stores on the first part of the round, they only have to pick the parts that have to be put in the first two kits. Then, when they go down, they only have to pick the parts that have to be put in the last two kits. This way, they have less information to process at one time. In addition, we sorted components according to their ‘type’ (for example, all locks are put next to each other), their ‘product type’ (locks for one type of vehicle are put next to each other) and their characteristics (right locks for one type of vehicle are put next to each other). This way, we were able to predict and balance the number of parts pickers had to pick in one specific area. Finally, since we were able to lower the variations of task intensity and we set AGV speed according to the numbers of components to kit and reduced the speed range.

Validation

We conducted the same analyses than previously several months after our intervention. Results showed a decrease in errors that had an impact on the production (from 22 to 10) and we recorded only one situation where pickers had to stop their AGV (versus 6 previously). Similarly, the teamwork we studied reported only 47 errors between February and March 2019 (versus 324 before from March to April 2018). Furthermore, we recorded less situations where pickers walked in front of and behind their AGV, suggesting that anticipated regulations on those workstations had become less essential for pickers to succeed in their tasks. At last, we asked every picker which layout they preferred and 21 over 24 chose the new layout.

Conclusion

Observations on workstations and intervention allowed us to demonstrate that the workstation design was responsible for the differences between prescribed tasks and real actions as well as operators' strategies. These differences should be interpreted as symptoms of workload that can be reduced by redesigning the work situation. In this way, we attempt to help engineers creating environments where operator's mental workload is optimized, so that operators can automatically accomplish their tasks without putting their health and performance at risk.

Keywords: Mental Workload; Cognitive processes; Variations of activity; Automated actions; Performance; Prevention of error.

References

1. Morais, A., Aubineau, R.: Articulation entre l'ergonomie et lean manufacturing chez PSA. Activites **9**(2) (2012).
2. Noldus, L.P.: The Observer: a software system for collection and analysis of observational data. Behavior Research Methods, Instruments, & Computers **23**(3), 415-429 (1991).
3. Noldus, L.P., Trienes, R.J., Hendriksen, A.H., Jansen, H., Jansen, R.G.: The Observer Video-Pro: New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. Behavior Research methods **32**(1), 197-206 (2000).
4. Jeanson, L., Bastien, J.C., Morais, A., Barcenilla, J.: From Prescription to Regulation: What Workers' Behavior Analyses Tell Us About Work Models. In: Congress of the International Ergonomics Association 2018, pp. 193-202. Springer
5. Jeanson, L., Morais, A., Barcenilla, J., Bastien, J.C.: N'oubliez pas de brancher les connecteurs: automatismes et traitement de l'information. In: 53ème Congrès de la SELF - Présent et Futur de l'Ergonomie, Toulouse, 20-22 Septembre 2017 2017
6. Jeanson, L., Christian Bastien, J.M., Morais, A., Barcenilla, J.: From Prescription to Regulation: What Workers' Behavior Analyses Tell Us About Work Models. In: Bagnara, S., Tartaglia, R., Albolino, S., Alexander, T., Fujita, Y. (eds.) Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018), Cham, 2019// 2019, pp. 193-202. Springer International Publishing
7. Jeanson, L., Bastien, J.M.C., Morais, A., Barcenilla, J.: Tell me how you kit, I'll tell you how you think: the use of behavior analyses to optimize parts supplies in automotive industry. Paper presented at the Human Workload Conference, Amsterdam,

Facing Mental Workload in AI-Transformed Working Environments

Nina Gusenleitner[1], Sandra Siedl[1], Gernot Stübl[2], Axel Polleres[3], Gabor Recski[4], Roland Sommer[5], Maria Chiara Leva[6], Mario Pichler[7], Theodorich Kopetzky[7] and Bernhard A. Moser[7]

[1]Institut für Arbeitsforschung und Arbeitspolitik at the Johannes Kepler University Linz, [2]PROFACTOR, Steyr, [3]Vienna University of Economics and Business, [4]APOLLO.AI GmbH, Linz, [5]Verein Industrie 4.0 Österreich, [6]School of Environmental Health, Technological University Dublin, [7]SCCH, Hagenberg, Austria

{nina.gusenleitner,sandra.siedl@arbeitsforschung.at; gernot.stuebl@profactor.at; axel.polleres@wu.ac.at; gabor@apollo.ai; roland.sommer@plattformindustrie40.at; chiara.leva@dit.ie; mario.pichler,theodorich.kopetzky,bernhard.moser@scch.at}

Abstract. We are focusing on working environments with an increasing role of AI for solving problems involving detection and classification of patterns and events, optimization, predictions or providing other services that generally are associated with human intelligence such as reasoning and learning. Such AI transformed working environments raise new requirements and challenges in terms of mental workload. We discuss these challenges in the context of Human-AI teaming.

Keywords: Artificial Intelligence, Human Centered AI, Machine Learning, Knowledge Graph, Human-AI Teaming

1 Introduction

Recent advances in AI, above all machine and deep learning, have brought about unprecedented possibilities in automation, prediction and problem solving with impact on our way of working. Among recent research on the topic of AI, there is unanimity that this new technology increases the efficiency, flexibility, and productivity of operations in the industrial and service sector but, at the same time, there is justified scepticism towards its implementation due to its unforeseen consequences at various levels, see [2]. Clearly, companies face the challenge of integrating AI into their operations and have to take measures in order to increase the acceptance for AI accordingly, see [1]. However, scientific research remains silent when it comes to fostering the acceptance of AI within organizations. So far, research is centred around increasing consumer's acceptance of AI, see [2, 4], or studies that quantitatively examine the consumer's acceptance of AI in consideration of different fields of application [3, 5]. Therefore, it is about now time to consider the user's side from an employees' point of view in order to foster AI in a human-technology relationship.

1.1 Example "Smart Maintenance"

Smart maintenance relies on accurate predictive analytic models for both prediction and optimization. While most machine learning techniques analyze (streams of) numeric sensor data, process mining analyzes streams of events. These events can originate directly from system logs, but can also be generated by machine learning algorithms analyzing the sensor data streams. Although sensor data allow predicting some basic events, more complex events are often the result of a complex process of heterogeneous events, which necessitates human input to the AI system for refining the event analysis capabilities of the AI system. Additionally, the algorithms need to be able to cope with the ever-changing dynamics of the systems and the environment, and therefore be self-adaptive. This process usually takes place in iterative loops including steps of interpretable hypothesis generation from the AI system and the integration of human observations and context knowledge in combination with support by application domain experts.

1.2 Example "Chat-bot based Customer Service"

AI-based dialogue systems, commonly referred to as chat-bots, are now an integral part of customer relations in many industries, including, but not limited to, online retail, banking, insurance, recruiting, etc. With AI handling the most trivial types of interactions, such as mechanically answering the most common types of questions or following a script to record customer data, humans should be able to focus fully on the more demanding aspects of customer relationships. To make this possible it is important to take into account additional context information such as existing business processes and user intention as starting point for designing new user experiences that are best suited for deploying conversational AI that enhances, rather than disrupts, day-to-day customer relation activities.

2 Discussion of Mental Workload Challenges

Our research interest is motivated by questions related to the acceptance problem of AI in the world of working with focus on operational users such as technicians, operators or implementers in the industrial and the customer service sector.

What influences human-system interaction and how can individuals' expectations be shaped to foster AI acceptance?

- Are there cognitive stress factors: How easy understandable and unambiguous is the AI system's output? How relevant? What is about the statistical nature of the output?
- Personalized relevance, granularity of information etc.
- Are there additional stress factors: ambient noise, pressure to perform?
- How much critical questioning is required?

- What is about responsibility aspects: duty of care versus information overload?
- Psychological aspects: Are there any imagined or perceived threats, e.g., in terms of violated privacy (“big-brother-is-watching you”-feeling)?

Which factors are specifically relevant in different work contexts?

- Stability of working and process conditions, e.g., shift in data characteristics due to changing or specific process or working conditions;
- Type and intensity of dependency of human actions and decisions on AI system;
- Degree of integration and system complexity;
- Impact of potential incorrect decisions: compensation of wrong decisions and possibilities of discharge;

How can user acceptance of AI be fostered and what are the implications for AI systems?

- Performance measures of the AI system: What about the accuracy and generalization capabilities?
- Trustworthiness: Are there stability guarantees? What is about security against manipulations and data integrity? Is a recommendation provided by the AI system underpinned by sufficiently enough evidence based on training data or is it the result of an interpolation or extrapolation? Is the AI system operating in the designed working range? What is about changing conditions?
- Meaningfulness: How understandable and relevant are the recommendations provided by the AI system?
- Transparency: Are there conclusive arguments for a provided recommendation?

2.1 Challenges for Human-AI Interface

At the core of these questions, we identify human-AI interface problems and challenges as follows:

AI-to-human (“the AI system’s output is meaningful and understandable”):

- Tackle the information overload problem**,** e.g., by exploiting attention mechanisms;
- Tackle the trust problem, e.g., by providing additional meta information in terms of confidence measures;
- Tackle the comprehensibility problem by increasing the AI system’s capability to explain the system’s output in terms of the user’s language (that is extracted from log files, comments etc.);

- Improve the (psychological) user experience by taking the user's qualification and social aspects into account, e.g., by adopting the level of abstraction, and if the set-up allows, by empowering the AI to give precise visual feedback, e.g., by utilizing Spatial Augmented Reality techniques;

Human-to-AI ("human artifacts are meaningful to the AI system"):

- Make the AI system's input channels and formats more flexible, e.g., by exploiting weakly supervised learning approaches in combination with knowledge graphs;
- Extend the AI system's interaction possibilities from standard mouse and keyboard use to natural interaction techniques, e.g. by integrating vision-based human gaze and pose detection;
- Enhance the AI system's situation understanding capabilities e.g. by Visual Grounding of natural language queries;

3. Towards integrating mental workload context in Human-AI Teaming Systems Challenges

Our envisioned technological conception relies on an innovative integration of various building blocks including mental workload models for designing a Human-AI Teaming framework. This view leads to an enhanced learning loop taking operational as well as mental workload context into account as illustrated by Fig. 1.

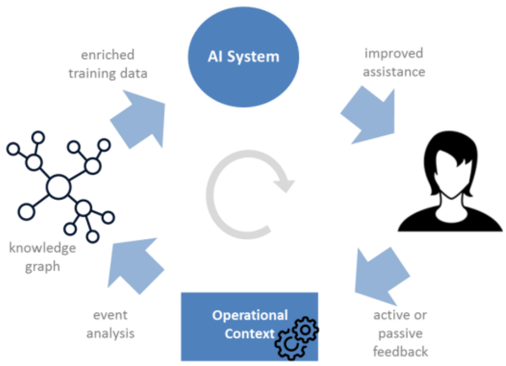


Fig. 1: Enhanced learning loop integrating mental workload context

At the core of our approach there is the idea for a "learning by listening and observation" concept that allows the AI system to adopt its (detection, classification etc.) capabilities in terms of expressiveness, explainability and relevance. The "listening and observation" part refers to extracting context-tags for an enhanced event analysis based on parsing system logs, interpretation of human interactions with the machine (resp. process, system) and additional workload as well as operational context information.

This approach is tackled by the recently granted Austrian exploratory project "AI@Work: Human Centered AI in Digitized Working Environments".

Acknowledgments. This work is supported by the Austrian FFG project nr. 25122825.

References

1. Daugherty, P. R. and H. J. Wilson. Human + Machine: Reimagining Work in the Age of AI. Boston: Harvard Business Review Press (2018).
2. M. Hengstler, E. Enkel, and S. Duelli: Applied artificial intelligence and trust; The case of autonomous vehicles and medical assistance devices," Technological Forecasting and Social Change, vol. 105, pp. 105–120 (2016).
3. D. Küpper, M. Lorenz, K. Kuhlmann, O. Bouffault, L.Y. Heng, J.V. Wyck, S. Köcher, and J. Schlageter: AI in the Factory of the Future: The Ghost in the Machine, Boston Consulting Group (2018).
4. H. Nitto, D. Taniyama, and H. Inagaki: Social Acceptance and Impact of Robots and Artificial Intelligence, Nomura Research Institute papers, vol. 211, pp. 1–15 (2017).
5. PwC, Bot.Me: A revolutionary partnership. How AI is pushing man and machine closer together, www.pwc.com/consumerintelligenceseries (2017).

Innovations in Dutch train cabs: the impact on workload, attention allocation and hazards

Vera Verstappen and Elise Pikaar
Vera.Verstappen@ns.nl, Elise.Pikaar@ns.nl

Nederlandse Spoorwegen, The Netherlands

Netherlands Railways has developed several DAS-applications (route context information and coasting information) to support the train driver in his task. Before implementing these innovations, we want to assess what the impact is on workload, attention allocation and risk level of hazards. To assess the impact of innovations a quantitative analysis based on eye tracker, simulator data and workload surveys is used. Overall the findings suggest that experienced workload isn't statistically significant different when driving with the DAS-applications (route context information application and route context and coasting information application), compared to driving with a basic paper time table. Also the attention allocation of train drivers doesn't statistically significant differ in the three conditions. Finally, for several identified hazards (i.e. SPADs) we also found no statistical significant difference in risk levels for the three conditions. The assessed innovations don't have a negative impact on workload, attention allocation and risk level of hazards.

WorkingAge: Smart Working Environments for all Ages.

Vera Rick[1], Julia N. Czerniak[1], Alexander Mertens[1], Vincenzo Ronca[2], Alessia Vozzi[2], Pietro Aricò[2,3], Gianluca Di Flumeri[2,3], Nicolina Sciaraffa[3], Carlos Alberto Catalina Ortega[4], Marteyn van Gasteren[4], Maite Cobo[4], Verena Nitsch[1] Fabio Babiloni[2,3] and Gianluca Borghini[2,3],

[1] IAW – Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Bergdriesch 27, Aachen, Germany
{v.rick, j.czerniak, a.mertens, v.nitsch }@iaw.rwth-aachen.de

[2] Brainsigns srl,
Lungotevere Michelangelo 9, 00192 Rome, Italy
{vincenzo.ronca, alessia.vozzi, pietro.arico, gianluca.diflumeri, fabio.babiloni, gianluca.borghini}@brainsigns.com

[3] Sapienza University of Rome,
Piazzale Aldo Moro, 5 00185 Rome, Italy
{pietro.arico, gianluca.diflumeri, nicolina.sciaraffa, fabio.babiloni, gianluca.borghini}@uniroma1.it

[4] Instituto Tecnológico de Castilla y León,
López Bravo 70, E-09001 Burgos, Spain
{maite.cobo, marteyn.vangasteren, carlos.catalina}@itcl.es

Abstract. WorkingAge (WA) will use innovative HCI methods to measure emotional, cognitive and physical strain of users. At the same time, with the use of Internet of Things (IoTs), sensors will be able to detect environmental conditions. The purpose is to promote healthy habits of users in their working environment and daily living activities in order to improve their working and living conditions. By studying the profile of elderly workers and the working place requirements in three different working environments (Office, Driving and Manufacturing), both profiles (user and environment) will be considered. Information obtained will be used for the creation of interventions that will lead to healthy aging inside and outside the working environment. WA will test and validate an integrated solution that learns the user's behaviour, health data and preferences and naturally interacts with the user through continuous data collection and analysis, with data protection always being a first concern.

Keywords: Emotive, Health, Mental States, Human-Machine-Interaction, Mental Workload, Multi-modal approach, Ontology, Strain, Stress, Worker

Understanding the Effect of Mental and Physiological Stressors on Cognitive Performance

Charelle Bottenheft[1], Anne-Marie Brouwer[1], Eric Groen[1] and Jan van Erp[1]

[1] TNO, Human Factors, Kampweg 55,
3679DE Soesterberg, the Netherlands, charelle.bottenheft@tno.nl

Abstract. Performing cognitive tasks is becoming important for an increasing part of workers who, at times, have to perform tasks under physical and mental load, or combinations of these. Effects of different stressors on performance have been studied in isolation, but little is known about the combined effects of two or more stressors. This study examined how a combination of two stressors influence working memory, reported effort and stress, as well as physiological indicators of mental state (arousal, attention and load). Skipping a meal is used as physiological stressor and noise burst as a mental stressor. 21 participants came to the lab twice, once after skipping a meal and once after eating a meal in the morning. They performed blocks of 2-back tasks, which were alternately presented with and without noise bursts. We found no main effect of skipping a meal on any of the variables. While noise did not affect reaction time and accuracy, it appeared to generate arousal and overall increased attention (higher EDA and P300) that was experienced as higher load and stress. Our results illustrate that physiological variables may help to reveal and understand the effects of stressors on individuals, besides measures of performance and reported experience.

Keywords: cognition; workload; noise; skipping a meal; psychophysiology

1 Introduction

Performing cognitive tasks is becoming important for an increasing part of workers who, at times, have to perform these tasks when under physical and mental load or a combination of these [1]. The effects of different stressors on performance have been studied in isolation [4] [5], but little is known about the combined effects of two or more stressors. This study examined the effects of a combination of two stressors on working memory performance, with skipping a meal as physiological stressor and noise bursts as a mental stressor. Given the flexibility of humans to adapt to changing circumstances, effects of stressors may not always be obvious when examining performance. For instance, additional invested effort may keep performance unaffected [2]. To get insight in the effect of the stressors on the required effort we used a wide range of measures and markers, including subjective effort and stress, and physiological indicators of mental state (arousal, attention and load).

2 Method

Twenty-one participants took part in this study. They were recruited through the TNO participant pool or acquainted with one of the test leaders. The present study was approved by the TNO Internal Review Board. All participants gave written informed consent. Participants were aged between 18 and 55 years (mean age 39.9 years) and 11 of the participants were male. Participants came to the lab twice, once after skipping a meal, and once after eating a meal in the morning. They performed a series of 2-minute blocks of a 2-back task. This task required participants to watch a sequence of letters on a computer monitor, where they had to indicate whether or not a letter was the same as the letter presented two positions earlier. Blocks were alternately presented with and without white noise bursts at random center frequencies (85 decibel), played through speakers. After each 2-minute block the participants rated their subjective mental effort on the Rating Scale of Mental Effort (RSME [3]) and their self-reported stress on a visual analogue scale (VAS, 0 = not at all, 100= extremely). During the 2-back task electrodermal activity (EDA), electroencephalography (EEG) and electrocardiography (ECG) were collected. After pre-processing the data, variables were tested with two-way repeated measures analysis of variance (SPSS: General Linear Model/Repeated Measures).

3 Results & Discussion

We found no effect of skipping a meal on any of the dependent variables. There was also no effect of white noise on the reaction time and accuracy of the 2-back task. However, the white noise stressor did significantly affect the number of missed trials (more missed trials in the no-noise); the reported mental effort (more effort in the noise condition); and the reported stress (more stress in the noise condition). Heart rate and heart rate variability were not affected by the stressors. However, EDA (both slow and fast response) was higher for noise than no-noise, indicating higher arousal for noisy conditions. In addition, EEG P300 brain response following letters in the noise condition indicated higher general attention, though it differentiated less well between target letters and non-target letters. Finally, an interaction effect between noise and meal on EEG alpha activity suggested that under no-noise conditions, when a meal was skipped, participants were investing little effort.

Our results illustrate that physiological variables may help to reveal and understand the effects of stressors on individuals, besides measures of performance and reported experience. Specifically, while noise did not affect reaction time and accuracy, it appeared to generate arousal and overall increased attention (higher EDA and P300) that was experienced as higher load and stress (subjective reports). Besides this, a more aroused state (high EDA) caused by noise can have beneficial effects (less skipped answers). The alpha results indicate that this may especially be important when participants skipped a meal.

References

1. Paulus, M. P., Potterat, E. G., Taylor, M. K., Van Orden, K. F., Bauman, J., Momen, N., Padilla, G.A. & Swain, J. L.: A neuroscience approach to optimizing brain resources for human performance in extreme environments. Neuroscience & Biobehavioral Reviews, 33(7), 1080-1088.3 (2009)
2. Veltman, J. A., Jansen, C., Hockey, G. R. J., Gaillard, A. W. K., & Burov, O.: Differentiation of Mental Effort Measures: consequences for adaptive automation. NATO SCIENCE SERIES SUB SERIES I LIFE AND BEHAVIOURAL SCIENCES, 355, 249-259 (2003).
3. Zijlstra, F., & van Doorn, L.: The construction of a scale to measure subjective effort. In Technical Report. Delft University of Technology (1985).
4. Adolphus, K., Lawton, C. L., Champ, C. L., & Dye, L.: The effects of breakfast and breakfast composition on cognition in children and adolescents: a systematic review. Advances in Nutrition, 7(3), 590S-612S (2016).
5. Szalma, J. L., & Hancock, P. A.: Noise effects on human performance: a meta-analytic synthesis. *Psychological bulletin*, *137*(4), 682 (2011).

Study of the mental workload and stress generated using digital technology at the workplace

José Castillo[1]✉, Edith Galy[1]✉, Pierre Thérouanne[1] and Raoul Do Nascimento[2],

[1] Laboratoire d'Anthropologie et de Psychologie Cliniques, Cognitives et Sociales (LAPCOS), Université Côte d'Azur, 25 avenue François Mitterrand, 06357 Nice, France
[2] NPxLAB,1047, Route des Dolines, 06560 Valbonne, France

Jose-manuel@npxlab.io

Edith.galy@univ-cotedazur.fr

Abstract. The introduction of technology into the work context has been and is still one of the most studied topics in the field of Human factors, Computer sciences and Psychology, due to the main advantages and disadvantages that may appear after its implementation and use. The current research aims to understand what and how important is the impact of digital technology on the mental health of employees. This study will be based on two main axes: The first one is oriented to evaluate and understand how (and to what extent) the mental workload of employees is influenced by the use of digital technology at work and the second one is focused on understanding how this technology impacts the employee's stress level. This article will describe in a synthetic way the purpose of the study, the research questions and the planned contributions to carry out this study.

Keywords: Human Mental Workload, Technostress, Human–Computer interaction, Human Factors.

1 Introduction

The introduction of technology into the work context has been and is still one of the widely studied topics by different disciplines such as Human Factors or Cognitive Psychology. This is reflected in the fact that more and more researchers are interested in the study of the relationship between the digital interfaces or tools and the assessment of concepts like Human mental workload [4] or Technostress [1,5]. Nowadays, Digital Technologies are becoming increasingly important in all types of organizations, examples of this are the Sentiment analysis with Text Mining and the Turnover prediction with Neural Networks. Both are new tendencies that Human Resources are adopting thanks to the massive application of artificial intelligence [6] or the utilization for employees of smartphones to access online education and training materials anytime from their own companies, all of that provided by ubiquitous computing [2].

The concept of technostress was triggered by the high labor demands, as well as the lack of technological or social resources related to ICTs. The possible consequences of technostress are psychosomatic complaints; sleep, headaches, muscle aches, gastrointestinal disorders, as well as organizational damage; absenteeism and reduced performance due to non-use or misuse of ICTs in the workplace [5]. It is for this reason, that it is important to continue the study and evaluation of this disorder that affects more and more collaborators.

The present study will follow two axes: The first one, focusing on assessing the influence of digital technologies on the mental workload and the second one, on assessing how these technologies are impacting the employee's stress. We will also study the impact of the mental workload of the employees on their stress level. Both studies will be carried out after an understanding of the context of the work and the use of new digital technologies.

This research will be based on the IWA (Individual - Workload - Activity) model [3], because this model fosters an integrative approach to understanding the cognitive evaluation of employees. The IWA model consists in taking into consideration the three components to understand the mental workload borne by workers: The first corresponds to individual characteristics, the second corresponds to the activity and the third component represents the mental workload with three dimensions (intrinsic load, extraneous and germane load). The latter was taken and adapted from Sweller's theory of cognitive load [7,8]. It is expected to identify which are the factors that are directly related to technologies such as utility and usability levels, individual factors such as skill level or level of computer literacy, and, finally the contextual factors of the use of digital technologies such as constraints related to use or frequency of use.

2 Research Questions

- What are the main characteristics of digital technologies, applied in the context of work, that can influence (increase or decrease) the levels of workers' mental load and stress?
- How can the experience of a mental overload in the work context, caused by the use of digital technology, impact the worker's stress?

3 Expected Contributions of the Research

- This research project will ultimately make more recommendations for the improvement and design of interfaces than those that exist today, considering the context and the individual;
- Propose individualized action plans within companies to facilitate the day-to-day work of managers and their teams in an increasingly digital context;
- This project can serve as a basis for human resources departments to set up specific training actions for individuals most sensitive to stress.

References

1. Ayyagari, R., Grover, V., & Purvis, R.: Technostress: Technological Antecedents and Implications. MIS Quarterly, 35 (4), 831--858 (2011)

2. Cascio, W. F., Montealegre, R.: How technology is changing work and organizations. Annual Review of Organizational Psychology and Organizational Behavior, 3, 349--375 (2016)

3. Galy, E.: A multidimensional scale of mental workload evaluation based on Individual – Workload – Activity (IWA) model: validation and relationships with job satisfaction. Tutorials in quantitative Method in Psychology (in press)

4. Longo, L.: Subjective usability, mental workload assessments and their impact on objective human performance. Proceedings Part II, 16th IFIP TC 13 International Conference, 25-29 September, Mumbai, India (2017)

5. Salanova, M., Llorens, S., Cifre, E., Nogareda, C.: El Tecnoestrés: Concepto, Medida e Intervención Psicosocial. [Technostress: Concept, Measurement and Prevention]. Madrid: Nota Técnica de Prevención. Instituto Nacional de Seguridad e Higiene en el Trabajo (2007)

6. Strohmeier, S., Piazza, F.: Artificial Intelligence Techniques in Human Resource Management---A Conceptual Exploration. In: Kahraman, C., Onar, SC. (eds) Intelligent Techniques in Engineering Management: Theory and Applications 2015. IRSL, vol 87, pp. 149--172. Springer, Cham (2015)

7. Sweller, J.: Cognitive Load During Problem Solving: Effects on Learning. Cognitive Science. 12, 257--285 (1988)

8. Sweller, J.: Cognitive load theory, learning difficulty and instructional design. Learning and Instruction. 4, 295--312 (1994)

Workload considerations in Urban Air Mobility

Tamsyn Edwards[1], Savita Verma[2], Jillian Keeler[2]

[1] San Jose State at NASA Ames Research Center, Moffett Field, CA 94035, USA
[2] NASA Ames Research Center, Moffett Field, CA 94035, USA
{tamsyn.e.edwards, savita.a.verma, jillian.n.keeler}@nasa.gov

1 Introduction

Urban air mobility (UAM) is receiving increased attention in the aviation literature as a traffic management system for the operation of passenger and cargo-carrying new entrants in urban airspace [1, 2]. UAM has been defined by the National Aeronautics and Space Administration (NASA) ATM-X project as "a safe and efficient system for air passenger and cargo transportation within an urban area" [3, p.3366). The roles and responsibilities of ATCOs and other human operators in relation to UAM traffic management remain undefined. Exploration of the human operator role in UAM is an essential element of the progression of UAM concept development. Identifying and exploring human factors issues such as task demand, associated workload, and performance, during an early stage of concept development, affords the opportunity to identify capabilities, as well as potential risks and associated mitigations, of human operator roles. The research reported in this paper aimed to contribute further understanding of human factors considerations, specifically workload, for near-term UAM operations.

2 Method

A human in the loop simulation was conducted, centered on low-altitude tower control sectors in the North Texas Metroplex area. Three within-measures variables were utilized. Task demand was manipulated to create three simulation scenarios, consisting of *low, medium* and *high* density UAM traffic. Two forms of communication procedure were utilized as the second variable, specifically, *current day* communication procedures and *reduced verbal* communications procedure implemented via a letter- of-agreement (LOA). Finally, the routes available to UAM consisted of two levels – the use of *current day helicopter routes* and *modified routes* that were optimized for UAM vehicles. Participants were six recently-retired controllers who had previously worked in tower control. Self-reported workload was measured throughout each simulation at 4-minute intervals using a modified uni-dimensional Instantaneous Self-Assessment scale (ISA). Each simulation session lasted for 40 minutes. A total of six retired controllers took part in the simulation, consisting of 4 males and 2 females.

3 Results

Subjective ratings of workload were considered in relation to UAM vehicle density. Inferential statistics were conducted to explore whether differences in average workload ratings between traffic densities were significant. A significant main effect of UAM traffic density was found on self-reported workload ($F(2,10) = 4.65$, $p<0.05$). A significant main effect of UAM traffic density on self-reported workload was identified ($F(2,10) = 9.31$, $p<0.01$). Pairwise comparisons revealed that average workload ratings were significantly lower in low density traffic compared to medium density traffic ($p=0.01$) and high-density traffic ($p<0.05$).

When considering the low UAM density condition, there appear to be differences between C (current day route, no LOA) CL (current day routes, LOA) and M (modified routes, LOA) conditions. No significant differences were found between average workload ratings in C, CL, and M conditions for the low, medium or high UAM traffic scenario.

4 Discussion and conclusion

It is acknowledged that the findings presented in this paper are provisional and need to be interpreted with caution. The small number of participants in the study may have influenced inferential statistics, resulting in the possibility of a type II error. The findings presented in this paper suggest that under current day operating procedures, using current day routes, UAM operations would be significantly restricted if actively controlled by human operators. Reduction of workload using a LOA did appear to support an increase in capacity of controlled traffic, as did optimization of routes. However, even with these adjustments, the scalability of UAM operations would remain restricted relative to the envisaged mid and far-term operations. The critical focus of future research therefore moves to supporting the development of an UAM system that is scalable, whilst maintaining, and even improving, the exceptional aviation safety standards.

References

1.Thipphavong, D.P., Apaza, R., Barmore, B., Battiste, V., Burian, B., Dao, Q., Feary, M., Go, S., Goodrich, K.H., Homola, J. and Idris, H.R., 2018. Urban Air Mobility Airspace Integration Concepts and Considerations. In *2018 Aviation Technology, Integration, and Operations Conference* (p. 3676).
2. Courtin, C., Burton, M.J., Yu, A., Butler, P., Vascik, P.D. and Hansman, R.J., 2018. Feasibility Study of Short Takeoff and Landing Urban Air Mobility Vehicles using Geometric Programming. In *2018 Aviation Technology, Integration, and Operations Conference* (p. 4151).
3. Chan, W.N., Barmore, B., Kibler, J., Lee, P.U., O'Connor, N., Palopo, K., Thipphavong, D.P. and Zelinski, S., 2018. Overview of NASA's ATM-X Project. In *2018 Aviation Technology, Integration, and Operations Conference* (p. 3363).

Organising committee

- General chairs, editors and programme chairs
  - Dr. Luca Longo
  - Dr. Chiara Leva
- Local chairs
  - Dr. Gianluca Di Flumeri
  - Dr. Gianluca Borghini
  - Dr. Pietro Aricò
- Finance chairs
  - Bridget Masterson
- Marketing, Graphics and publicity chairs
  - Andrea Gervasini
  - Chiara Brambilla

Programme committee

- Andy A. Smith Psychology Cardiff University, UK
- Mobyed Ahmed Mälardalen University, Sweden
- Julie Albentosa French Armed Forces Biomedical Research Institute,France
- Pietro Aricò Sapienza University of Rome, Italy
- Bargiotas Paris Saclay Paris, France
- Emanuele Bellini Khalifa University, UAE
- Gianluca Borghini Sapienza University of Rome, Italy
- Bethany Bracken Charles River Analytics, USA
- Suzy Broadbent BAE Systems, UK
- Karel Brookhuis The University of Groningen, The Netherlands
- Aidan Byrne Swansea University, UK
- Brad Cain CAE Defence and Security, Canada
- Jose Cañas University of Granada, Spain
- Giulia Cartocci Sapienza University of Rome, Italy

- Martin Castor Group of Effectiveness, Interaction, Simulation, Technology and Training, Sweden
- Loredana Cerrato Nuance Communications, USA
- Lorenzo Comberti Politecnico di Torino, Italy
- Ferdinand Coster Yokogawa Europe Solutions B.V., UK
- Katie Crowley Trinity College Dublin, Ireland
- Dick De Waard The University of Groningen, The Netherlands
- Micaela Demichela Politecnico di Torino, Italy
- Gianluca Diflumeri Sapienza University of Rome, Italy
- Tamsyn Edwards San Jose State University, NASA AMES, USA
- Enda Fallon NUI Galway, Ireland
- Jialin Fan Cardiff University, UK
- Pedro Ferreira The Resilience Engineering Association, France
- Matjaz Galicic Mott MacDonald, Ireland
- Glauco M. Genga Società Amici del Pensiero Sigmund Freud, Italy I
- van Gligorijevic mBrainTrain, Serbia
- Dave Golightly University of Nottingham, UK
- Birgit Greiner University College Cork, Ireland
- Lisa Jeanson mpsa.com, France
- Tara K. Smith HFE Solutions, UK
- Bridget Kane Karlstad University, Sweden
- Fiona Kenvyn Metro Trains Melbourne, Australia
- Patricia Lopez de Frutos CRIDA ? ATM R&D + Innovation Reference Centre, Spain
- Ivan Macuzic University of Kragujevac, Serbia
- Anneloes Maij Netherlands National Aerospace Laboratory (NLR), The Netherlands
- Alice Marascu Nokia Bell Labs, Ireland
- Lynne Martin NASA, USA
- Pavle Mijovic mbraintrain, Serbia

- Ann Mills Rail Safety and Standards Board, UK
- Enrique Muñoz de Escalona Fernández University of Granada, Spain
- Fedja Netjasov University of Belgrade, Serbia
- Leonard O?Sullivan University of Limerick, Ireland
- Maria Gabriella Pediconi University of Urbino, Italy
- Melinda Ratkai University of Malaga, Spain
- Philippe Rauffet Université Bretagne-Sud, France
- Bujar Raufi South East European University, R. Macedonia
- Audrey Reinert Purdue University, USA
- Vincenzo Ronca BrainSigns, Italy
- Alessandra Sala Nokia Bell Labs, Ireland
- David Strobhar Beville Engineering, USA
- Thea Raduentz Federal Institute for Occupational Safety and Health, Germany
- Alessia Vozzi BrainSigns, Italy
- Mark Young Loughborough University, UK
- Rolf Zon Netherlands National Aerospace Laboratory (NLR), The Netherlands

Acknowledgments

We wish to thank all the people who helped in the Organizing Committee for the Third International Symposium on Mental Workload, Models and Applications (H-WORKLOAD 2019). In particular the local chairs, Dr. Gianluca Di Flumeri, Dr. Gianluca Borghini, Dr. Pietro Aricò, and the members of the scientific committee listed in this book for their contribution toward the revision process and their constructive feedback to the scientific community around it. We would also like want to include in our thanks the main sponsors of the event, the Sapienza University of Rome, the Irish Ergonomics Society, and the Technological University Dublin. Without the support of this extended community the conference and the book would not have been realized. A special thank you goes to the researchers and practitioners who submitted their work and committed to attend the event and turn it into an opportunity to meet and share our experiences on this fascinating topic.

H-WORKLOAD
models & applications

November 14th and 15th

Rome 2019

www.hworkload.org