Towards a method for the objective assessment of cognitive workload

A pilot study in Vessel Traffic Service (VTS) of maritime domain

Andrea Malagoli[1], Matteo Corradini[1], Paola Corradini[1], Todd Shuett[2] and Sergio Fonda[1]
[1]*University of Modena and Reggio Emilia, Italy* -[2]*Kongsberg Norcontrol AS, Horten, Norway*

***Abstract*—The complexity of traffic control systems, as well as the growing volume of traffic, interconnected missions types and mission demands on the operators, indicates that critical attention should still be paid to the problem of operator's cognitive workload (WL). On the other hand, the development of traffic control towards on-line measurement of cooperative aspects between humans and machines, is part of the more general need to measure the human agents "situation awareness" in industrial environment.**

The University of Modena and Reggio Emilia was partner of the European Artemis project "Designing Dynamic Distributed Cooperative Human-Machine Systems" (D3CoS 2011-2014) [1] to define affordable methods, techniques and tools addressing the specification, development and evaluation of cooperative systems where human and machine agents are in charge of common tasks, assigned to the system as a whole. One of the basic keys to reach an optimal human-machine cooperation is the measure of the human operator workload.

In order to setup a possible method for the objective evaluation of cognitive workload we had to investigate aspects of the functional status of human operators interacting with a simulator in maritime domains.

We recorded objective psycho-physiological measures: eye blinks, respiration rate and amplitude, electrodermal activity, heart rate variability, and blood pressure. They were analyzed and correlated with subjective self-assessed responses from two questionnaires: NASA-TLX and Rating Scale Mental Effort (RSME), with the aim to realize a mathematical model for classifying the operators' mental workload.

The purpose of this paper is to present the methods, applied on a pilot study, that we carried out to discriminate the WL intensity, based on psycho-physiological signals alone.

***Keywords*—mental workload, EMD, psycho-physiological signals, R cran**

I. Introduction

Mental workload (WL) is a multidimensional and complex construct for which there is no clearly defined and universally accepted definition in literature: there are inconsistencies related to its sources and its mechanisms that give rise to a considerable variety of methods of evaluation. [2] In general, it can be considered as a "quantitative function" of the relationship between the mental demands required by a task and the available resources of the human operator. In order to gather a valid and reliable assessment of WL it is necessary to employ a set of measures according to: (1) psycho-physiological measures and (2) self-assessed measures (rating scales). For (1) we decided to acquire information by autonomic nervous system measures with:

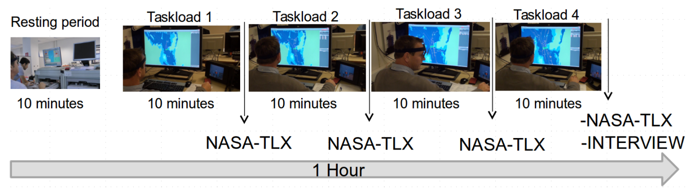

Figure 1. Timeline schema of the experiment. At the beginning, we need time to instrumental setup, to present the experiment to the subject and to pose probes on him. A resting period was used to record baseline signals before starting with the four WL phase. At the end of each task-load the subject filled a NASA-TLX and a RSME questionnaire.

1) Electro Dermal Activity (EDA) which is linearly correlated to arousal and measures stress and frustrations too [3].
2) Ocular activity: eye activity reflects central nervous system activity and indicates also mental WL [4] [5] . In particular the eye blink rate yields information about task demands, level of fatigue, memory and response demands.
3) Cardiovascular measures: they have been reported to be sensitive to WL [6], and emotional activation which was evaluated by means of self-assessed measures. ECG signal, Heart Rate (HR), Heart Rate Variability (HRV), Low Frequency to High Frequency ratio (LFHF) of HRV spectrum and Photoplethysmography (PPG) were measured directly. Blood Pressure (BP) was indirectly calculated from ECG R-peak and PPG peak delay with single calibration.
4) Respiration: respiration rate, inspiration and expiration time, complete cycle time, volume and flow rate was measured. For the (2) we used two different questionnaires: 5. a NASA-TLX (overall WL score) and RSME (Rating Scale Mental Effort) [7] [8]

II. Methods

A typical situation found in Human Machine Interaction (HMI) is testing an operator's behavior to different tasks. We developed a set of methods based on the comparison between objective physiological measurements and subjective responses from questionnaires to assess the workload status of a subject during his job in front of a monitor interface simulating.

We performed an experiment in which we tested our methodologies and tools on five different Vessel Traffic Service (VTS) operators working at the VTS centres in Horten and Brevik, Norway. In this paper we focus on the methodologies and, we report just data and results of only one single VTS operator. The subject was a Canadian male volunteer, aged 34, right-handed, with normal visual acuity, normal hearing and good health. He reported experience of 3 years in his job. No compensation was granted for his participation; after a briefing and an interview to gather preliminary data, he signed the consent form to be the experiment's subject.

According with our partners, we defined four different workload scale scenarios (underload, medium, high and overload task). The workload intensity was varied increasing the number of the following elements: the vessels in the monitored area, the vessel reporting, software alerts and communications between VTS operator and vessels. An experienced instructor (T.S.), hidden at the sight of the subject, simulated vessel radio communication.

The duration for each scenario was about 10 minutes (6min for signal recording, 4min for questionnaires filling). A complete timeline schema of the experiment is represented in Fig1.

All the psycho-physiological measures were recorded by BioSemi Mod. Active Two and sampled at 2048Hz, except the ocular activity that was recorded by Jazz Novo – Ober Consulting and sampled at 1000Hz. The scenario simulation software, provided by Kongsberg, was C-Scope Operator Client 4.7. [11] At the end of each scenario a NASA-TLX and RSME questionnaires were compiled by the operator. Raw signals were acquired and saved as files in EDF/ASCII format. These signals were converted to text using EDF Browser software before analysis in R (version 3.1) [12] with RStudio (v 0.98) [13].

In addition to conventional digital filtering techniques, Empirical Mode Decomposition (EMD) [9] [14] was applied to all recorded signals, in order to identify and separate signal information from noise. EMD is helpful to analyze composite, nonlinear and non-stationary signals. One of the relevant advantages of EMD is that the basis functions are derived from the signal itself; infact, compared to Fourier analysis, whose basis functions are fixed sine and cosine waves, EMD is adaptive. The fundamental idea of EMD method is an iterative sifting process that decomposes the signal into a sum of Intrinsic Mode Functions (IMF), considered as basic building blocks of the data time series (e.g., noise is one of these "building blocks"). An IMF is a signal which must satisfy two criteria: extremes and zero crossings points are in equal number or differ at most by one and the mean of upper and lower envelopes of IMF is zero. EMD can be applied if the signal has at least two extremes, one maximum and one minimum, so enabling a successful decomposition into IMFs; this property is quite common in bio-medical signals. The software used to perform the EMD analysis was the package EMD of the R language (R CRAN repository) [9]. Apart the usual noise cleaning, the main EMD preprocessing purpose was to identify peaks and extract signal components with physiological significance. The analysis has been performed

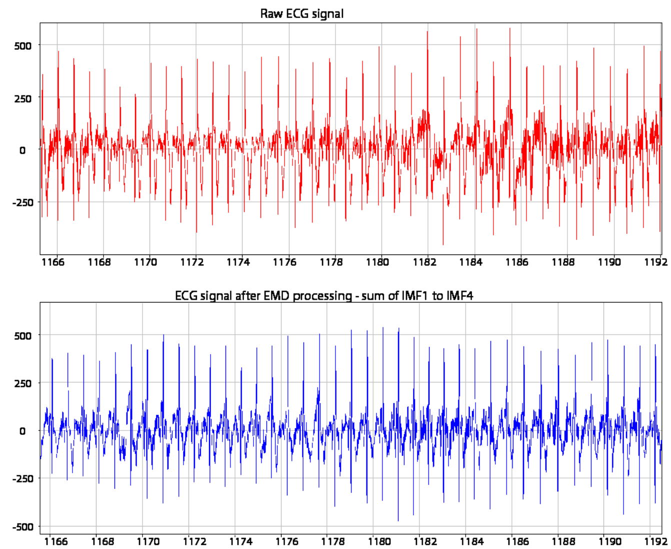

Figure 2. Raw ECG signal (red) and ECG cleaned signal by EMD technique

for every parameter and separately for each workload phase. The IMF number for our signals was typically ranging from 12 to 15. For each signal we defined a different procedure to remove noise and clean data.

Heart Rate: Heart Rate (HR), expressed as beat per minute, was evaluated from ECG signal, detecting the time interval between two consecutive R peaks. The large ECG drift, still present after digital band pass filtering (0.1-10Hz) and represented by an intrinsic mode function (IMF), was easily removed so enabling R-R time interval series identification. Fig 2.

LF/HF ratio: Starting from the R peaks series, the heart rate variability is analyzed from tachogram estimating its spectrum behavior in low (LF, 0.04-0.15Hz) and high frequency (HF, 0.15-0.4Hz) band, focusing into LF/HF ratio., All the heart rate variability analysis was carried on by means of RHRV package [10].

Blood Pressure: Systolic blood pressure was derived from a linear model [8] that evaluated the systolic blood pressure as a function of the delay between R peak of ECG and blood "pulse wave transit time" (PWTT) peak, detected from photoplethysmogram (PPG) peak, inside every R-R cycle.

$P_{syst} = -0.4148 * PWTT + c$ where the calibration $c = P_{syst_0} + 0.4148 * PWTT_0$ constant is evaluated by P_{syst_0} measured directly by sphygmomanometer in mmHg at the beginning of the experiment; $PWTT_0$ is obtained as the average value among six values of the pulse wave transit time: three before the direct measure and three after. PWTT peaks were detected from PPG signal in the same way as R peaks: low pass band filtering (0.1-4Hz) and PPG decomposition by EMD. The sum of three IMFs (IMF6+ IMF7+ IMF8) proceed the final PPG signal profile Fig 3. For both R peaks than PPG peaks we used a custom peak detection function to eliminate false peaks.

Electrodermal activity: Electrodermal activity (EDA), recorded by Biosemi equipment, has been divided in tonic and phasic component. For the tonic component we evaluate

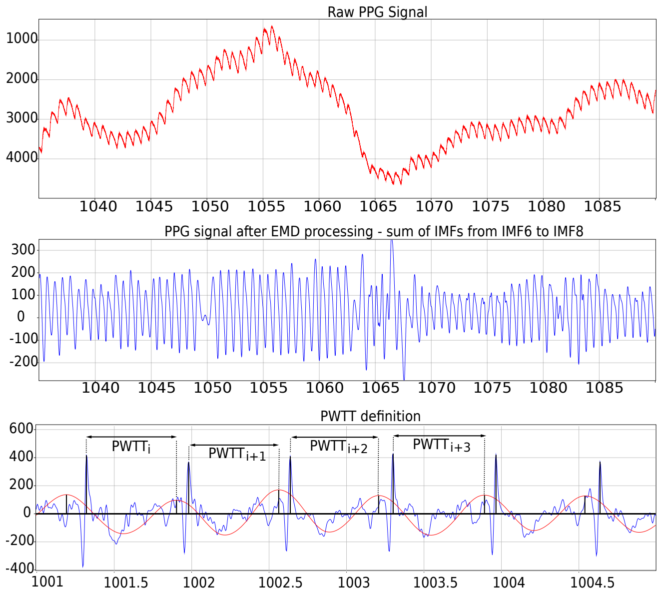

Figure 3. Processing of raw PPG Signal. Processed PPG (middle graph), obtained by the sum of the three IMFs, IMF6+ IMF7+ IMF8. ECG and PPG signals (bottom graph), after EMD processing and in the same time interval, with the corresponding R and systolic peaks. The PWTT intervals are marked as PWTTi. Horizontal and vertical scale of the top and middle plots are "s" and "au (arbitrary units)" respectively.

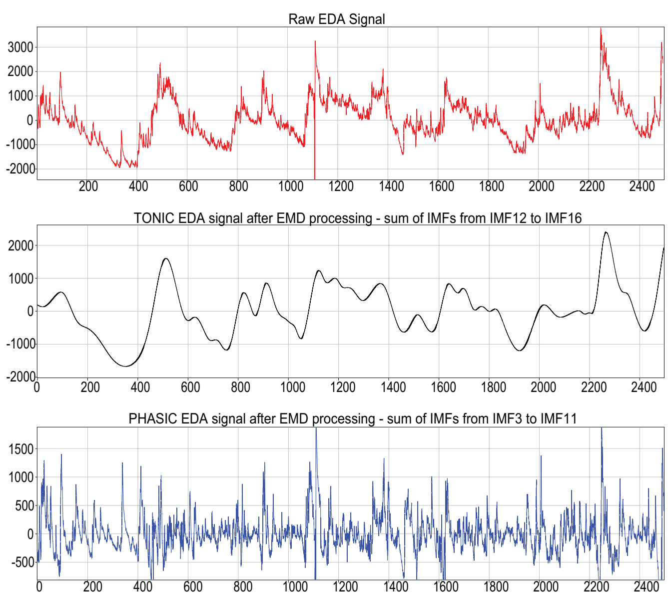

Figure 4. Processing of EDA raw signal. Tonic (slow) EDA signal estimated as sum of IMFs from 12 to 16. Phasic (fast) EDA signal using IMFs from 3 to 11.

the mean value, while for the phasic we reported the mean rate value as shown in Table 1. The raw data acquired was decomposed with EMD technique: the IMFs(12-16) were used to estimated the tonic component, while the IMFs(3-11) were used for phasic component Fig 4.

Respiration: Respiration signal was recorded by Biosemi chest belt. It was decomposed with EMD technique in 14 IMFs. We used the slow component of the IMFs (from 9th to 13th) to detect peaks – the end of the inspiration phase – and the valley – the end of expiration phase. Starting

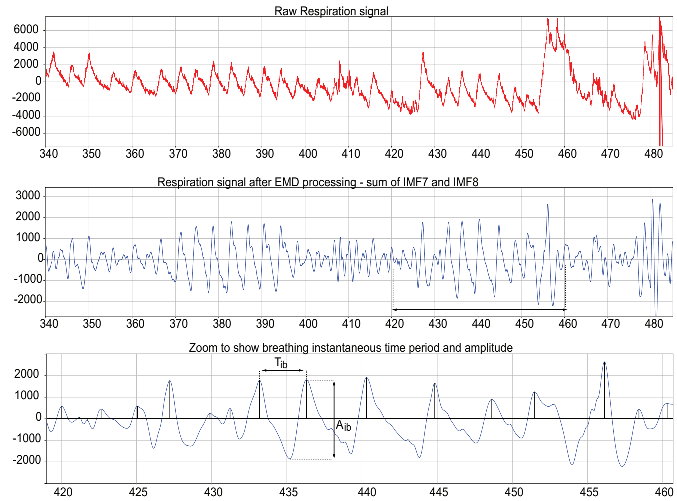

Figure 5. Respiration raw signal. Processed respiration signal as sum of IMFs from 9 to 13 (middle graph). An example of breathing period and amplitude (bottom graph).

from these two points the respiration rate was evaluated from time difference between peaks, and respiration amplitude from amplitude difference between peaks and valleys.

Eye blinks rate: Eye movements were recorded by Jazz Novo instrument. It acquires information of eyes movement in XY plane. The blinking events were identified from characteristic signal profile assumed in both first and second derivative of X and Y trajectories.

TLX – RSME: The NASA-TLX produces a numeric index called Overall UnWeighted Score (OUWS), as result of a sum of six values corresponding to the following scales (each within a 0-100 interval): Mental demand (MD), Physical demand (PD), Temporal demand (TD), Effort (Ef), Performance (Pe), Frustration (Fr). Filling RSME the VTS operator produces a single score in the range of 0-150 [8].

For every WL phase all the eight psycho-physiological recorded data were split into 30s intervals in order to get many sets of equivalent measures and associated with those obtained in NASA-TLX and RSME questionnaires; and analyzed with MANOVA and Discriminant analysis with R Studio. The duration of 30s has been chosen because it is the minimum time value to give a reliable valuation for many parameters (HRV, EDA, and eye blinks). The four workload phases may be considered as factors in the experimental design. The questionnaires data were collected at the end of each of the four workload phases. As a consequence, we have only one record of responses from both NASA-TLX and RSME associated to the number of twelve records collected for every objective parameter; in this condition, a statistical analysis to clarify the relationship between objective and subjective measurements is not possible. To accomplish the statistical analysis, the TLX and RSME results must be extended over the workload phases to cope the corresponding number of values got from objective measurements.

The replication of the same number must be excluded because a null variance would be associated to the new series. We applied a method based on "jittering" which, in our case,

Table I
LIST OF PHYSIOLOGICAL SIGNALS

Parameter	Definition
mHR	Mean value of heart rate [bpm]
LF_HF	Low frequency to high frequency ratio, from spectral analysis of heart rate variability (HRV) [number]
mSyst	Mean value of systolic blood pressure, based on pulse wave transit time R-PPG peak [mmHg]
mEDAtonic	Mean value of tonic component of EDA [mSiemens]
mEDAphasic	Mean rate value of phasic component of EDA [mSiemens]
mBreathrate	Mean respiration rate [Bpm]
mBreathamp	Mean respiration "peak-valley" amplitude [arbitrary units]
nBlinks	Mean value of the eye blinking per minute rate [EBpm]

means applying a small normal distribution of noise to the numerical indexes. In practice the test index is considered as the global average value obtained in the workload phase; a series of twelve random values with mean equal to the test index and a standard deviation in the range $1\% - 5\%$ of the mean is generated and it is considered as the subjective experimental result produced every 30s into the current workload phase. In this way we have produced a reliable table of subjective and objective measurement data. Of course, if a workload phase has a longer duration, the series length increases accordingly. The jittering calculation has been provided by the function jitter() in R. A sample of the new assessment of results of the TLX and RSME tests is reported in Fig 6.

A preliminary multivariate analysis was performed with MANOVA to check if some variability of the eight parameters arises as function of workload levels. Its positive result (significant F value, $p < 0.01$) indicated that the eight groups differ on the set of workload measures with a p-value < 0.05. Linear Discrimination Analysis (LDA) was then applied to predict workload phase by physiological parameters values. We used R MASS package applying the linear discrimination analysis with the lda() function in two ways, training and cross validation.

Furthermore a Canonical Correlation Analysis (CCA) was used to measure the association among workload subjective estimations (TLX and RSME) and objective psychophysiological measures. (Fig 7).

III. RESULTS

Univariate Analysis

The graphs of mean values series are reported in Fig.8. The data have been fitted using the LOESS method; it performs a local, non parametric, regression at the point x, by fitting a regression function to the data points within a chosen neighborhood of the point x. LOESS method is suitable when a robust fitting method is necessary to include outliers. In Fig 8 the fitting is carried out taking the workload phases as continuous series.

OUWS	RSME	mHRV	mSystP	mEDAtonic	mEDAphasic	mBREATHampl	mBREATHrate	nBlinks	LF_HF	WorkLoad
1671	1093	8841	14070	1569632	12	179530	6	0	0.014	W1
1706	1085	8747	14078	1627169	11	362599	6	0	0.032	W1
1715	1089	8776	14962	1670867	9	284556	6	0	0.028	W1
...continueW1...										
5308	5109	8701	14618	1842042	20	312896	5	3	1.501	W2
5199	5215	9183	14581	1884021	6	348357	5	5	2.415	W2
5268	5150	8882	14770	1852063	8	175391	6	4	1.054	W2
...continueW2...										
6485	7155	9348	13808	1803716	14	311044	6	1	0.008	W3
6513	7351	9191	13860	1716207	9	39180	7	6	0.047	W3
6383	7161	9093	14775	1706926	5	32102	7	6	0.007	W3
...continueW3...										
9089	11402	9201	14941	130679	18	632626	5	4	1.224	W4
9273	11331	9608	15028	121314	9	360641	7	11	0.781	W4
9193	11216	9414	14868	119673	16	1047772	6	4	3.057	W4
..continue W4..										

a

TLX-NASA								RSME
Workload	MD	PD	TD	Ef	Pe	Fr	OUWS	
W1	2	2	2	2	7	2	17	11
W2	9	4	8	11	12	9	53	52
W3	12	4	13	11	11	13	64	73
W4	19	4	19	18	14	17	91	112

b

Figure 6. a) A sample of the database resulted by averaging signals in periods of 30 seconds and b) table of subjective scores recorded from questionnaires at the end of each scenario.

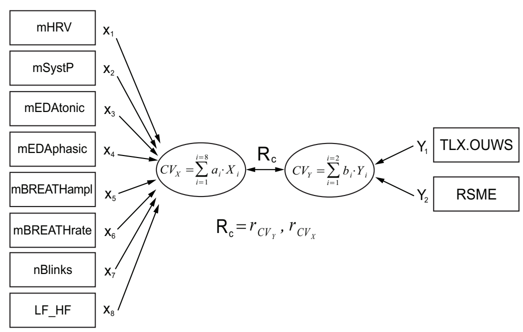

Figure 7. CCA . On the left the eight objective physiological signals and on the right the subjective responses from questionnaires.

Hearth Rate and LFHF: The heart rate signal, in bpm units, is reported in Fig 8. It presents a trend increasing from W1 to W4 phase, much steeper in W1 and W2, less sharp in W3 and nearly constant in W4. The maximum variance values are obtained in W2 and W4 phases. The signal is further processed to analyze the heart rate variability spectral characteristics, and reported in the same figure.

Electrodermal activity: The signal is decomposed in its tonic and phasic components. The mEDAtonic trace in Fig 8 corresponds to the mS value of the EDA baseline, called as the slow activity, got from EMD method for every 30s interval. The mEDAphasic trace represents the number of fast responses to endogenous or external events/stimulations accumulated for each 30s interval. The red lines show the LOESS interpolation for the two parameters as a function of workload magnitude. The tonic component, which shows a rather stable amplitude in the first three workload phases,

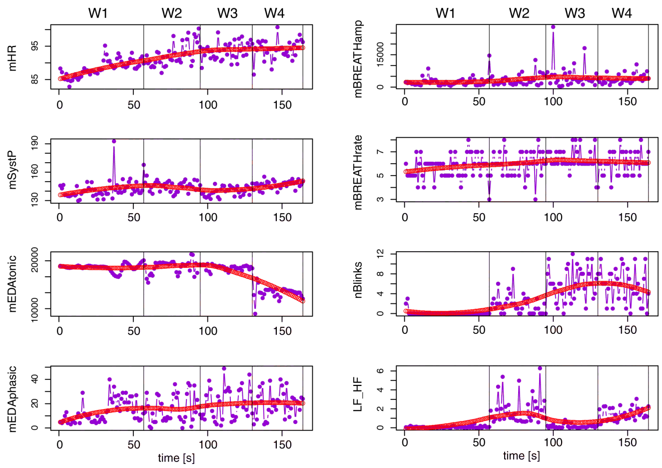

Figure 8. Parameters obtained from all the eight objective signals from the four different WL. The red line is the LOESS fitting

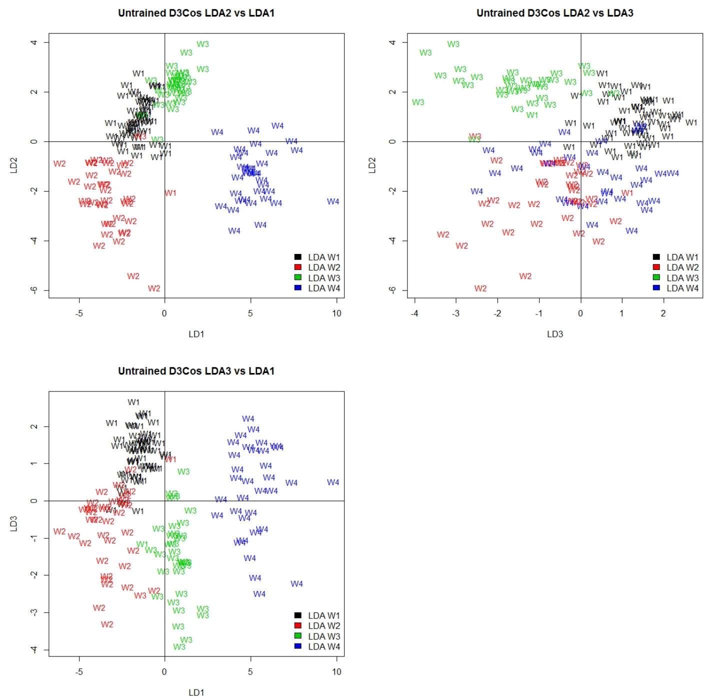

Figure 9. The LDA function values are computed without any previous training reference, including all the original data. Top left - LD2 vs LD1: WL1 and WL4 are well classified; Top right - LD2 vs LD3: distinguishes between WL2 and WL3; Bottom left - LD3 vs LD1: strong compactness for WL2 and good separation/classification among others. Every WL point in the plots is originally marked with a color and a label; a mis-categorized point (put WL in different group from original one) is plotted with a color changed.

decreases markedly in the overload phase of workload. The rate of fast phasic components shows an oscillating behavior; however the LOESS line presents a slowly increasing curve from low to high workload.

Eye blinks rate: The Fig 8 presents the rate of eye blinks along the four workload phases. Despite its oscillating pattern, the LOESS interpolation shows a growing rate with a maximum value reached in W3 and a damping trend in W4. Respiration rate and amplitude The LOESS interpolation signal of the breathing rate (Fig 8) is rather constant in W2-W4 phases interval, after a slight increment from 8 breaths/m to 9 breaths/m during W1 phase. The maximum variance is reached in W3.

The breathing amplitude: LOESS signal, reported in arbitrary units in Fig 8 is constant in W1 phase, increases during W2 and remains constant in W3 and W4 phases at the maximum value reached in W2. The maximum variance level is reached in W4.

TLX – RSME: NASA-TLX and RSME evaluated at the end of each scenario were consistent with the increase in the WL.

Multivariate Analysis

The last step is based on application of Linear Discriminant Analysis (LDA) and Canonical Correlation Analysis (CCA). Our Linear Discriminant Analysis deals with four WL levels (groups) and eight physiological variables (predictors). This goal is achieved constructing three discriminant equations $D_i(i = 1, 2, 3)$ which are linear combinations of the predictor variables, such that the different workload groups differ as much as possible on D. We get three discriminant functions because the number N of possible discriminant functions arises from the rule: $N = min(p, q) - 1$ where p and q are the number of groups and of variables respectively. We have 4 workload groups and 8 physiological variables, so $N = 4 - 1 = 3$ discriminant functions called LD1, LD2 and LD3 in this case. They are found developing a script in R language using the R package MASS. In Fig 9 is presented a typical result obtained with the three LDAs; in which the number of misclassifications is very small (error rate 3%), so reflecting the correct choice of workload scaling in the experimental design.

Canonical correlation analysis is used to identify and measure the associations among two sets of variables, in our case between the subjective (metric independent variables) and objective physiological variables (multiple dependent measures). The theoretical aspect in this case is grounded over the definition of two linear combinations (canonical variates) of subjective and objective variables respectively and finding the correlation coefficient between all the possible “variates” according to the scheme presented in Fig 7.

The CCA may be divided in two parts: the preliminary one related to the analysis of the usual correlation between all the variables, subjective and objective independently. This part is useful to acquire knowledge about basic behavior and possible relationships in the data frame. The results about this part are reported in Fig 10 (top right). The CCA method provides a set of new orthogonal dimensions to “variables of observe” the relative correlations between variables, which are represented in Fig 10 (top left). The method’s advantage is the possibility to correlate variables different in their nature as in our case. The graph reported in Fig 10 (bottom left) and Fig 10 (bottom right) shows a good classification of workloads reached by the physiological variables, “supported” also by the presence of subjective variables in the canonical variates correlation calculus.

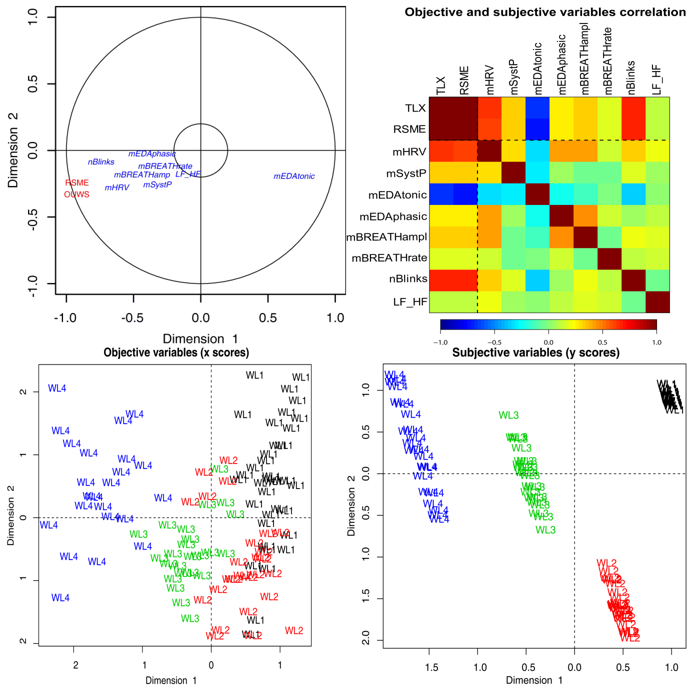

Figure 10. Top left - Shows the canonical correlation coefficients for all variables. The larger circle has radius=1, the smaller radius=0.5. The smaller the distance from the center, the lesser the correlation expressed by the variable. Variables positioned in the same direction are correlated with amplitude proportional to their relative distance. Top right - Represents the correlation values between variables couples, according to the color bar code. Bottom left - The first two canonical variates (Dim1 and Dim2) are the horizontal and vertical plot axes. Every objective variable is represented in its canonical coordinates and its position in the graph is related also to its correlation with subjective variables. Bottom right - First two canonical variates for subjective variables (NASA-TLS and RSME). The plot shows no miss-classification for the WL levels.

IV. Conclusion

We realized a methodology that combine in a new way various methods: from signal processing by empirical mode decomposition to advanced statistical analysis, in a single procedure to classify the WL based on both self-assessed measures and psycho-physiological parameters. In fact, using filtering and EMD technique it is possible to evaluate a useful signal to perform the analysis.

In this pilot study, each single physiological signals is correlated to workload status and to subjective auto reported measurements. Furthermore both LDA and CCA can discriminate workload status in a single subject. We may conclude that it is possible to infer about the workload quantitatively using a number of eight physiological variables, easy to be recorded, and the developed method seems to be adequate to support the goal.

Next steps are moving on two different direction: first of all we need to test our methodology on more subjects to confirm our results; secondly we will consider the possibility to reduce the number of recorded signals to facilitate both the experiment setup and the calculation in order to perform real-time evaluations of the WL.

Acknowledgment

This work was part of the European project D3CoS (www.d3cos.org) founded by Artemis, SP-8, GA No:269336. Our laboratory ElaSTyC would like to thank Kongsberg for the partnership.

References

[1] Lüdtke, A.; Javaux, D.; Tango, F.; Heers, R.; Bengler, K. and Ronfle-Nadaud, C. (2012). Designing dynamic distributed cooperative Human-Machine Systems., Work (Reading, Mass.) 41 Suppl 1 : 4250-4257.
[2] Fallahi, M.; Motamedzade, M.; Heidarimoghadam, R.; Soltanian, A. R. and Miyake, S. (2016). Effects of mental workload on physiological and subjective responses during traffic density monitoring: A field study., Applied ergonomics 52 : 95-103.
[3] W. Boucsein. Electrodermal Activity. Springer-Verlag GmbH (2011).
[4] Riggs, L. A.; Kelly, J. P.; Manning, K. A. and Moore, R. K. (1987). Blink-related eye movements., Investigative ophthalmology & visual science 28 : 334-342.
[5] Cardona, G. and Quevedo, N. (2014). Blinking and driving: the influence of saccades and cognitive workload, Current eye research 39 : 239-244.
[6] Fallahi, M.; Motamedzade, M.; Heidarimoghadam, R.; Soltanian, A. R. and Miyake, S. (2016). Assessment of operators' mental workload using physiological and subjective measures in cement, city traffic and power plant control centers., Health promotion perspectives 6 : 96-103.
[7] Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later, 50 : 904-908.
[8] Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research, Advances in psychology 52 : 139-183.
[9] Kim, D. and Oh, H.-S. (2013). EMD: Empirical mode decomposition and Hilbert spectral analysis, R package version 1.
[10] Rodriguez-Liñares, L.; Méndez, A. J.; Lado, M. J.; Olivieri, D. N.; Vila, X. A. and Gómez-Conde, I. (2011). An open source tool for heart rate variability spectral analysis, Computer methods and programs in biomedicine 103 : 39-50.
[11] Schuett, T. (2014). Reducing VTS operator stress, Port Technology, 2014, 1-2.
[12] R Core Team (2013). R: A Language and Environment for Statistical Computing, .
[13] Team, R. (2015). RStudio: Integrated development environment for R. Boston, MA, .
[14] Kim, D. and Oh, H.-S. (2009). EMD: a package for empirical mode decomposition and Hilbert spectrum, The R Journal 1 : 40-46.