Paper 8, Wednesday 25th May 2016

Tracking Stress and Workload in the Maritime/Tugboat World

Dimitri van Heel, Sr Project Manager, Teamleader Nautical Operations, MARIN (Maritime Research Institute Netherlands), The Netherlands

Kerstin Klinkenberg, Partner, K + S Projects, Germany

Prof Benjamin Blankertz, Professor, Technische Universität Berlin, Germany

Daniel Miklody, Master of Science, Technische Universität Berlin, Germany

Wendie Uitterhoeve, Project Manager Human factors, MARIN (Maritime Research Institute Netherlands), The Netherlands

Martin Ouwerkerk, Senior scientist, Philips Group Innovation, The Netherlands

SYNOPSIS

The Dutch research institute MARIN, the Technical University of Berlin, Philips and k+s projects would like to present the findings of a pilot study into stress/workload measurements during training in a shiphandling simulator. Working on a tug has unique demands, made more challenging by fatigue and a high workload, and influenced by special demands such as noise, the intense mixture of private and worklife and what often prove to be extreme environmental challenges.

The objective of this pilot study is to determine the most suitable tool to determine the workload a person experiences when executing complex tug manoeuvres. Scenarios are selected strictly according to real life on the bridge, focussing on tugboat reality. Measurements include heartbeat rate and skin conductivity, plus an EEG, but also include simulator signals. The human body is a very sophisticated control circuit, and to choose different ways of looking at the way a human expresses himself is like joining puzzle elements to a complete picture. Together they pinpoint what is most challenging for the test person. The findings will be evaluated to assess which signals are most suitable to obtain a reliable workload indicator. This tool can be used to study the impact of job procedures, modern bridge design and time/work shift systems, but also to measure the effectiveness of training programs. The pilot study is the follow up of the demonstration of "Training meets Science“ during the maritime conference and exhibition ITS 2014 in Hamburg, initiated and organized by k+s projects, University of Applied Sciences Bremen and University of Technology Berlin.

INTRODUCTION

It was in 2014, in Hamburg, where the last ITS took place, when delegates could see "fancy hat fashion" in front of a ship handling simulator in the foyer between the ITS exhibition and the conference hall. These were the first steps, made of tracing workload and visualizing brain activity in a maritime simulator. Students of the Hochschule Bremen where confronted with a tugboat scenario, escorting a passenger vessel upstream in an estuary into a lock. Not only students but also tugexperts, delegates from your companies literally put their heads in our hands, helping to promote the idea of putting more focus on the human factor, which also implicates the appreciation of the work done in the wheelhouses and ships on the oceans. In Hamburg, only EEG measurements were taken.

Picture#1 ITS 2014, k+s booth

The promising results of the EEG measurements demonstrated on the ITS in 2014, led to the initiative to go for the next step; the development of a workload indicator. The workload indicator is a number that gives an indication of the workload an individual perceives in a certain situation. It provides insight in how much effort is needed to deal with the situation.

A reliable workload indicator can contribute to better results in many fields: The effectiveness of a training can be better assessed, since it can tell you how the mental efforts of a trainee develops when an exercise is repeated over time. It can also be helpful when looking at the design of a new terminal or port. In this case the workload indicator can be used to find the lay-out which causes the pilot least effort. Or when you are designing your operation or work process, the indicator can help you identify the most critical phases of the operation. Furthermore, it can be a useful tool when designing bridge lay-outs or man machine interfaces, where it can be used to determine in which way data is presented to a seafarer in the most effective way. Many more examples can be given in which an objective assessment of the experienced workload can be useful.

In a cooperative project k+s projects, the University of Berlin, MARIN and Philips worked together to further develop the workload indicator. The general idea behind this project can be described in one sentence:

Perform bridge simulations with different levels of workload and use separate tools to estimate the experienced workload in order to come to a reliable workload indicator.

In 2015 a number of bridge simulator experiments were executed in which tug masters performed easy and difficult tasks. During these experiments multiple physiological parameters were measured. The objective of these efforts was to determine which kind of measurements, or which combination of measurements can serve as an indicator for mental workload in a nautical environment.

The simulations were executed on a tug simulator and included electroencephalography, heart rate frequency, breath frequency and skin conductivity measurements. Also simulator signals were collected. The measurements were evaluated independently and combined to determine if they can contribute to a reliable workload indicator.”

First we will give a definition of workload and then we will look at how workload translates into physiological features. After that we explain the set-up of the experiments and guide you through our findings:

WHAT IS WORKLOAD AND WHAT IS ITS EFFECT ON THE HUMAN BODY?

A simplistic definition of workload is that it is a demand placed on humans [1]. However, the workload a person experiences is not only task-specific, it is also person-specific [2]. It is therefore better to distinguish between demand and workload. The task demand is the goal that has to be attained by means of task performance and is independent of the individual. The demand leads to an amount of information processing capacity that is used for the task. This is the so called workload. The workload a person experiences depends on individual restrictions. The same task demand will lead to different workloads for different individuals. When a task becomes more difficult the workload increases. The difficulty of a task is dependent upon context, state, capacity and strategy that is used to perform the task. A task may be relatively easy for a well trained person and very hard on a novice. After a sleepless night a task will be more difficult, regardless experience.

The reaction to task demand expresses itself in physiological reactions. The cognitive information processing takes place in the brains. Measuring brain activity provides information about this cognitive process and is a direct measure for the mental workload.

The autonomous nerve system regulates physiological reactions like heart rate, blood pressure, breath frequency, skin conductivity, adrenaline and other hormones. At moments with increased workload, the sympathetic part of the autonomous nerve system activates the body to prepare for the so called fight of flight status. More adrenaline for example will increase heart rate at these moments. On the other hand, after the stressor has gone, the parasympathetic part starts to rebalance. More relaxation is expressed in a lower heart rate for example.

Physiological measurement relies on evidence that increased mental demands lead to increased physical response from the body. Measuring physiological changes is therefore an objective, but indirect, measure for the experienced workload. It is the most exact and objective way to find workload because it does not require a response/opinion from the person. Sometimes the body tells another story than the person, whose oral response may be influenced by unknown or hidden motivations. Literature describes different methods to measure workload. As there is no single physiological measure that indicates workload with 100% guarantee, the fidelity of a workload indicator increases when multiple measures are combined.

HOW TO MEASURE MENTAL WORKLOAD?

In this study we focused on brain activity, cardiac activity, respiratory activity and skin conductivity. The way in which these activities are measured and what the measurements tell us about workload is explained in the following sections.

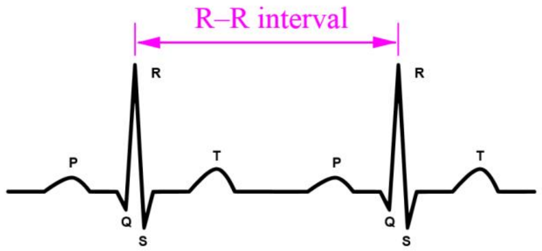
Measuring brain activity:

The brain activity was measured with an electroencephalograph, better known as EEG. It uses electrodes placed on the scalp to detect the electrical activity on the surface of the scalp. We used a 64 channel high quality EEG system with active electrolyte gel contacted electrodes. The EEG signals are the changes of voltage measured at the scalp which result from neuronal activity in the brain. A very basic and simple abstraction could be, that changes of voltage and frequency are measured of a huge pile of electric wires. And it is not only the workload that leaves traces in this pile of wires. The acquired signals are a mixture of the true brain activity and other 'noise sources' such as muscle activity (in particular of facial and neck muscles), eye movements and blinking and electrical devices. It is easy to understand that it needs quite some experience to filter brain activity patterns and determine “footprints” of mental workload.

With a different degree of success wave patterns can be correlated to a mental workload. Brain activity in the theta frequency range (4 to 7 Hz) in frontal brain regions have been found to positively correlate with the level of workload, see e.g. [5, 6, 7]. With respect to the more prominent alpha frequency band, most studies report a negative correlation of cognitive workload and alpha power at parito-occipital scalp locations, see e.g. [8, 6]. However, these studies used tasks in the visual modality to induce workload, such that one can only derive the implication of alpha reduction for workload in visual resources. In general, the functional role of alpha band oscillations is not yet conclusive. Some studies using auditory stimulation even found an increase of alpha activity with increasing workload ([9, 10, 11, 12]).

Measuring cardiac activity:

With the same amplifier system we recorded an Electrocardiogram (ECG) of the heart in a standard 3 electrode montage optimized for R-peak detection. During the simulation runs a complete ECG is recorded. A schematic representation of an ECG is given in the figure below. From this ECG R-peaks are determined and the interval time between two following R-peaks is calculated.



Picture #2, RR interval

Time traces with these RR intervals are input for analysis. Time domain results like mean and standard deviation of RR intervals, as well as variability of the RR intervals are interpreted as measures for experiencing mental workload. Literature reports a consistent pattern of cardiovascular activity from laboratory and field studies; heart rate increases and heart rate variability (HRV) decreases as a function of increases in cognitive workload [13].

Using frequency domain analysis methods, a spectral distribution for the RR series is calculated. From this analysis, the relative power within the low and high frequency bands and the ratio between the LF and HF band powers are used as indicators for experiencing workload. The theoretical reaction of cardiac measures to workload are given in the table below.

Measure	Reaction to higher workload
Mean RR	Decrease
Std RR	Decrease
Mean HF	Increase
Std HF	Decrease
RMSSD	Decrease
LF/HF	Increase
Power LF	Increase
Power HF	Decrease

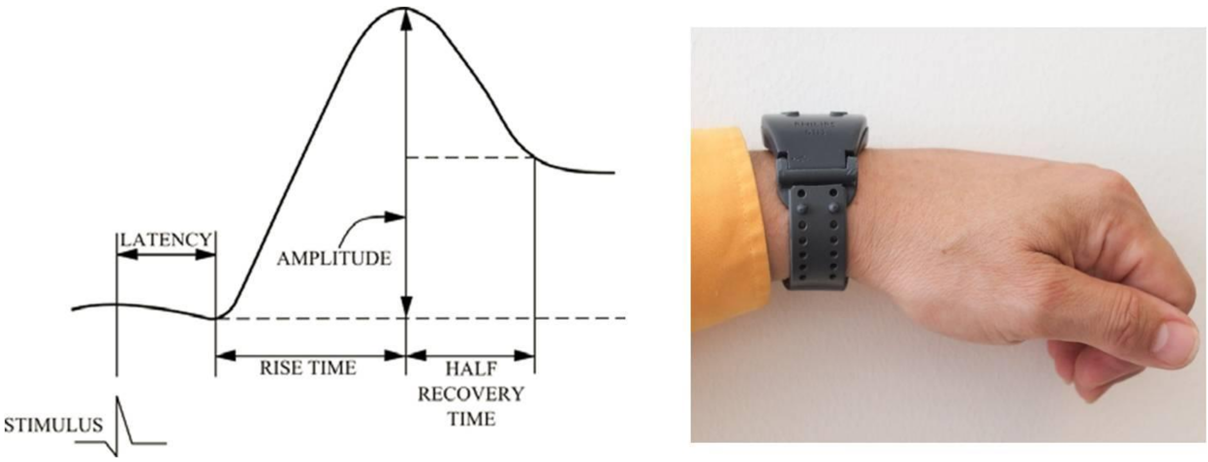
Table #1, Relation between cardiac activity and workload

Measuring breathing activity:

The respiration rate increases under stressful attention conditions [14] and as a result of increased memory load or increased temporal demands [15]. The breath frequency can be obtained from the ECG signal. While breathing the ECG electrodes move relatively to the heart causing small potential changes. These potential changes are used to compute the respiration frequency.

Measuring skin conductivity:

The participants wore two sensor bracelets, which measured the skin conductivity. The conductivity is directly related to the amount of sweat produced by the sweat glands. Therefore, the conductivity is a good indicator for the arousal/stress state of a subject. The conductivity will react to a stimulus with a certain latency. After an initial rise in conductivity, it will recover again. A value that can be used to express the arousal is the increase of conductivity per minute.



Picture #3, Skin conductivity signal and bracelet

SETUP OF THE EXPERIMENTS

In this study the workload of tug masters in a realistic setting was studied. Tug assist tasks were performed on a tug simulator, where a representative 60 tonnes bollard pull ASD tug was simulated. The experiment was executed on one of MARIN’s tug simulators in the Netherlands.

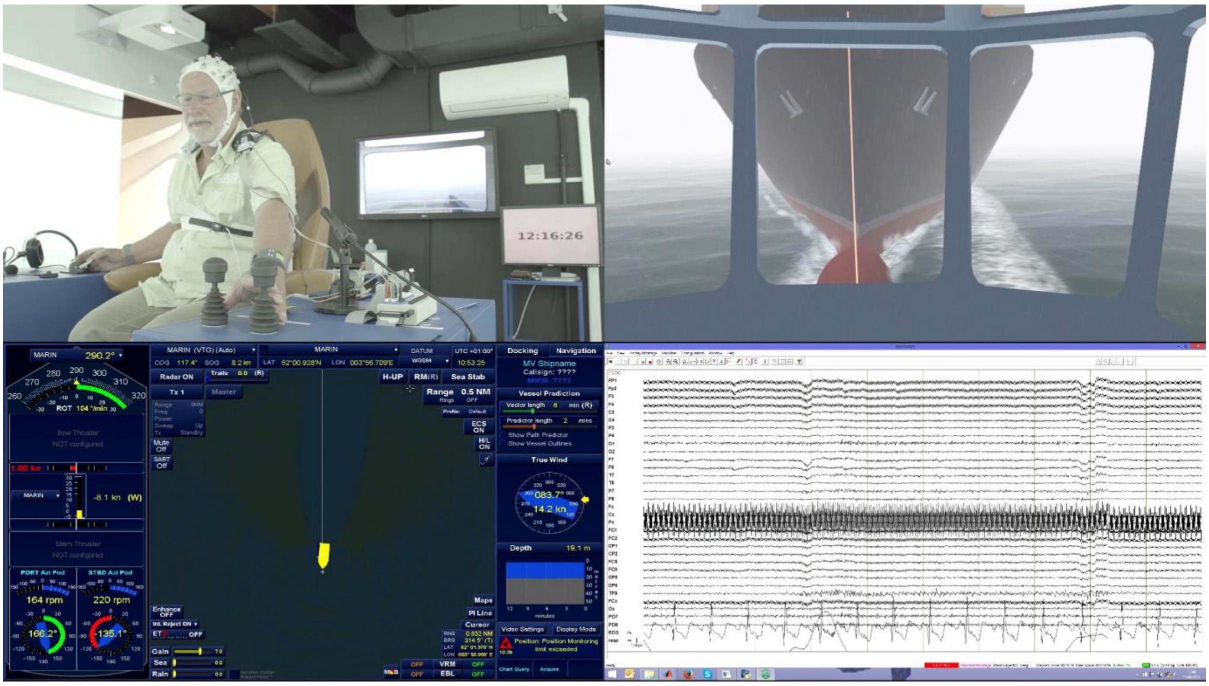
The tug simulator has a 270° visual image and an additional monitor provides the view astern. The simulator is equipped with controls for the thrusters and tow winch and displays with indicators showing ship information like: the engine revolutions, thruster angles and parameters such as the speed, rate of turn, force in the tow line, wind speed and direction, etc. Also a radar screen with electronic chart and a conning display were fitted .The tug captains were instructed via VHF orders from the instructor.

Picture#4, Simulator setup

Physiological measures were obtained in the ways described in the previous section. EEG, ECG, external marker signals and simulator data are synchronised with LabstreamingLayer [16]. The skin conductor bracelets used a local data storage. Prior to and at the end of the tests, the bracelets had to be synchronised with the other data recordings.

Prior to the experiment, the candidates were asked to fill in a perceived stress scale. This rating indicates the amount of stress the candidates experienced during the previous month. After each run, the candidate quantified the perceived mental effort using the Rating Scale of Mental Effort (RSME).

Synchronised video recording of the candidate in the simulator, the captain’s view, physiological measurements and the radar screen were taken to capture an overall picture of the run.



Picture#5, Synchronised video recording

10 tug captains from MARIN's network participated in this experiment. All men, in age between 30 and 65 years old, with different levels of experience in ASD tug driving. They participated voluntarily in this experiment and received a compensation for their participation. They received participant information in advance and went through a short briefing at the beginning of the experiment in which objectives, procedures, protocol and their privacy rights were addressed. They signed an informed consent.

PILOT EXPERIMENT

The study consisted of two distinct phases. In the first phase the technical set-up of the experiment was tested and different scenarios were evaluated. The insights gained in this phase were used to optimize the scenarios and test protocol for the second phase. The pilot experiment was executed in three days. Three tug captains participated.

In the first phase, or pilot experiment, three scenarios were tested:

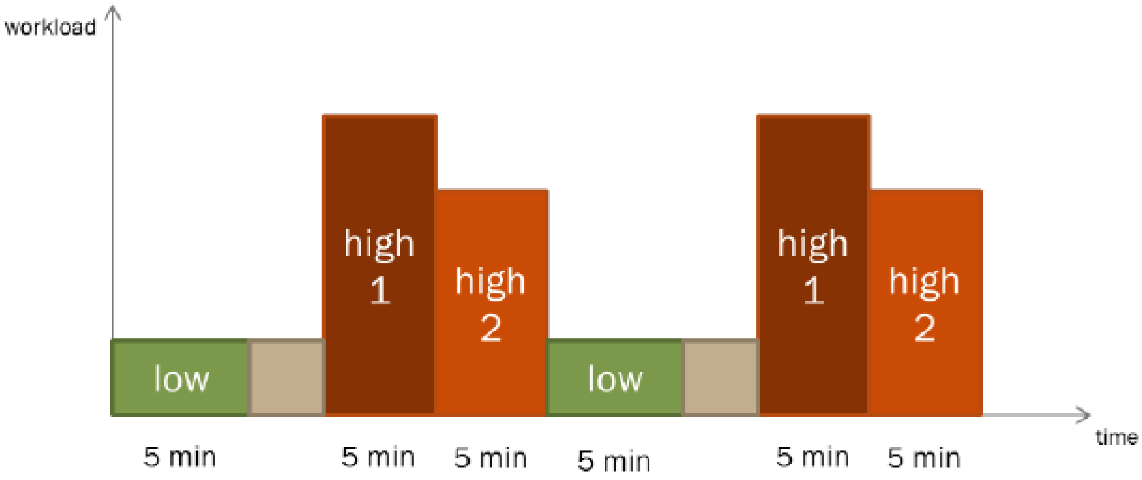
***Scenario 1, bow-to-bow operation**:*
In this scenario the tug was escorting a large container vessel en-route to the port of Rotterdam. This scenario was divided in several periods of low and high workload:

Condition1 – free sailing (low workload): The tug master was instructed to follow the container vessel and keep station at the starboard quarter. After 5 minutes the tug was ordered to reposition to the bow and wait there for further instructions.

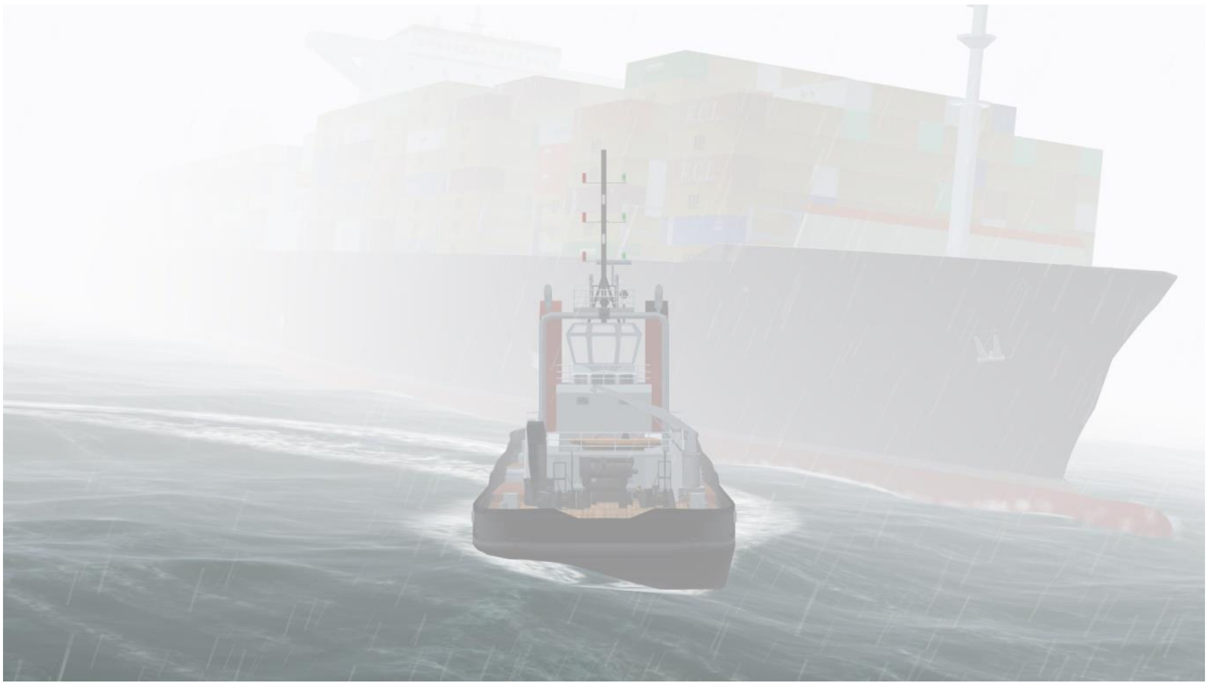
Condition 2 – connecting (high workload): The tug was ordered to approach the bow of the ship to receive the messenger line. The bow-to-bow position was maintained for 5 minutes. The tug was typically 15-30 m in front of the container vessel, sailing astern at 7 kts. This condition is referred to as high-1.

Condition 3 – pulling (high workload): This condition started with the connection of the towline. The tug master was ordered to keep a constant tow force and direction. After five minutes the towline was disconnected and the tug was ordered to reposition to the starboard quarter of the ship. This condition is referred to as high-2.

When the tug reached the original start position, the sequence was repeated.



Picture #6 , Experiment design bow-to-bow pulling scenario



Picture #7, bow-to-bow pulling scenario

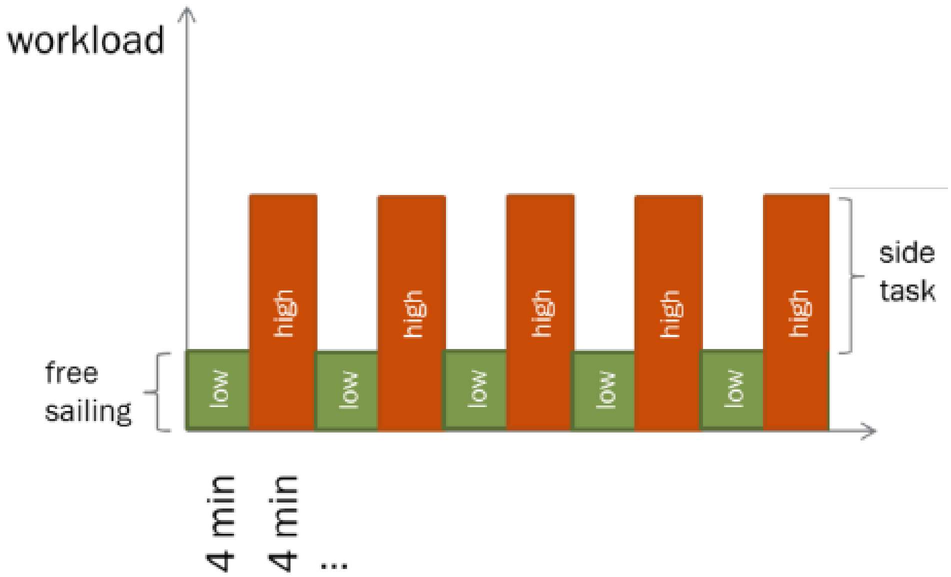
Scenario 2, n-back task:
The n-back task is commonly used in neuroscientific research as a manipulation tool for cognitive workload. N is typically chosen between 0 and 3 in order to induce different levels of workload. We used this to have a condition comparable to common research and to see how much our bow-to-bow

scenario corresponds to the neural patterns of this commonly known task. We used an auditory 2-back task, where the subject had to follow a stream of spoken numbers. If the last number heard corresponded with the digit 2 back, he had to press a button. The digits 1-9 were used with 3 s interleave randomly (75%) and forced 2-back repetition (25%) to get a reasonable amount of repetitions. The 2-back was played auditorily to keep a realistic behavioural scheme of the captain. There were two conditions, 4 mins each:

Condition 1: Free Sailing Condition (low workload): In this condition, the same low workload task of the bow-to-bow scenario is induced for comparison.

Condition 2: Free Sailing with 2-Back Condition (high workload): The 2-back task was used additionally to the Free Sailing to induce a higher workload while keeping the primary task constant.

Both conditions were repeated five times resulting in a total duration of 40 minutes for the whole phase.



Picture #8, Experiment design, N-back task

Scenario 3, pull back operations:
In this scenario a typical offloading operations from a Single Point Mooring was simulated. In this kind of operations an oil tanker is connected to a buoy with a hawser to receive oil from an oil producing platform. A tug is used to provide a pullback force to prevent the tanker from moving forward and come in contact with the buoy. In constant conditions this is a low workload scenario. In our scenario the high workload was introduced by a sudden increase and rotation of the wind, which led to a sudden rotation of the tanker. The tug master had to keep a constant pull and had to maintain a 6 o'clock position with respect to the tanker

RESULTS OF THE PILOT EXPERIMENT

The pilot experiment showed that the physiological measurements can be executed in a realistic set-up without too much degradation of the signals. Although some interference occurred, signal quality of ECG and EEG measurements remained sufficiently high. The tug masters provided good feed-back to improve the scenarios. The N-back task was graded as the most demanding task, requiring considerable mental effort. In general the bow-to-bow operations were not regarded as very difficult. Even stationing close to the bow was regarded as a normal operation, requiring little effort. However, this changed when the wave height was increased and visibility reduced. Under these conditions the bow-to-bow operations were graded equally demanding as the strenuous N-back task.

Scenario 3, the pull-back operation was regarded as quite difficult. However, the tug masters acted very differently on this task. While one captain really struggled to achieve the task at hand, others decided it was too difficult and continued the run without really trying to achieve the requested position and pull back force.

Overall the first experiment showed that it was possible to measure good data and with this data it was possible to determine EEG classifiers for predicting high and low workload.

It became clear that the time needed for briefing, preparation, executing the three different scenario's and debriefing took more than 4 hours. Our aim to test two candidate's per day and to collect sufficient repetitions of the same conditions per candidate were in conflict with this. Based on the results of the pilot in combination with the above remarks, it was decided not to repeat scenario 3 in the second experiment, but instead focus on bow-to-bow operations. The N-back task was considered a valuable scenario for comparison.

FINAL EXPERIMENT

The second simulation session was executed in a five day period. In this period ten tug captains participated, who all performed scenario 1, the bow-to-bow operations twice, and scenario 2, the N-back task, once. At the end of the week a demonstration session was given to interested parties. During the experiment a lot of data and movie material was collected which were analysed by different parties. TU Berlin was responsible for the EEG analysis, Philips analysed the results of the skin conductivity measurements, and MARIN did an analysis of the ECG and simulator data. From the 10 participants one of the candidates became sick during the simulations. Therefore his results were not valid for further use.

SUBJECTIVE RESULTS

After each run, the candidates rated the perceived mental effort during that run. The results are presented in the table below. The overall average shows that the candidates experienced the second bow-to-bow scenario as less demanding than the first and the n-back task. This can be due to the learning effect. It is also seen that the rating of the N-back task is either much higher or much lower compared to the bow-to-bow scenario. It is possible that candidates rated only the demand needed for manoeuvring. Although the N-back task is a validated tool to increase task demand, it is possible that differences in individual strategies contribute to the amount of workload experienced by the candidate. The score reflects the evaluation of an entire run and does not distinguish between low and high workload phases within the run.

Candidate	RSME Bow-bow-1	RSME N-back	RSME Bow-bow-2
1	5	5	15
2	40	70	60
3	40	70	30
4	40	10	30
5	71	105	38

6	38	73	47
7	40	25	15
9	71	26	38
10	60	40	60
Average	45	47	37

Table #2, RSME scores per simulation run

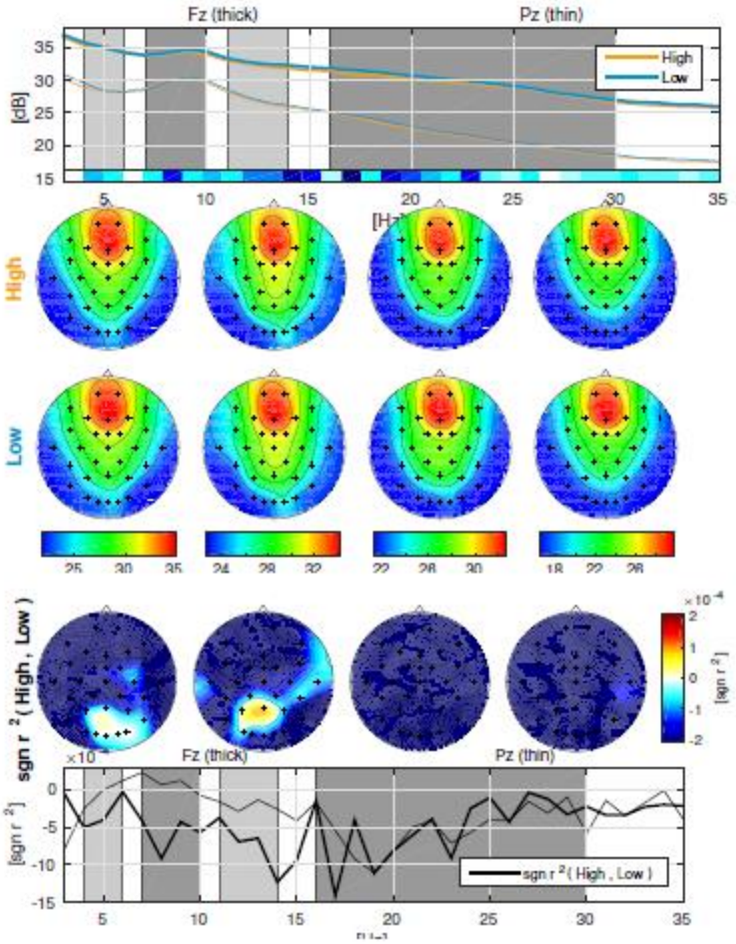
EEG ANALYSES AND RESULTS

With spectral analyses the occurrence of specific wave patterns over a longer period of time can be determined. The aim of the analyses is to determine so called classifiers. A classifier can be described as a wave pattern that is indicative for high or low workload. In other words: a pattern of activity in a specific frequency band in a certain area of the brain is linked to assumed periods of high and low workload.

We have seen in the spectral analysis, that the data is strongly affected by artifacts. Apart from noise of the technical devices, there are artifacts from muscle activity as well as from eye movements. Furthermore, there may be motion artifacts due to the motion of the electrode cables induced by head and trunk movements. The muscular and ocular artifacts are indicative of the workload condition for a number of participants and could in principle be used for the workload classifier. However, we focused on finding a classification method that doesn't use those artifacts.

To find the classifiers, first the data was filtered and both automatic and manual artifact reduction was applied. This resulted in a cleaned EEG signal. For each participant a spectral analysis was made in which the density over the frequency range was determined for each of the electrodes. A grand average was determined showing the average density distribution over the scalp.

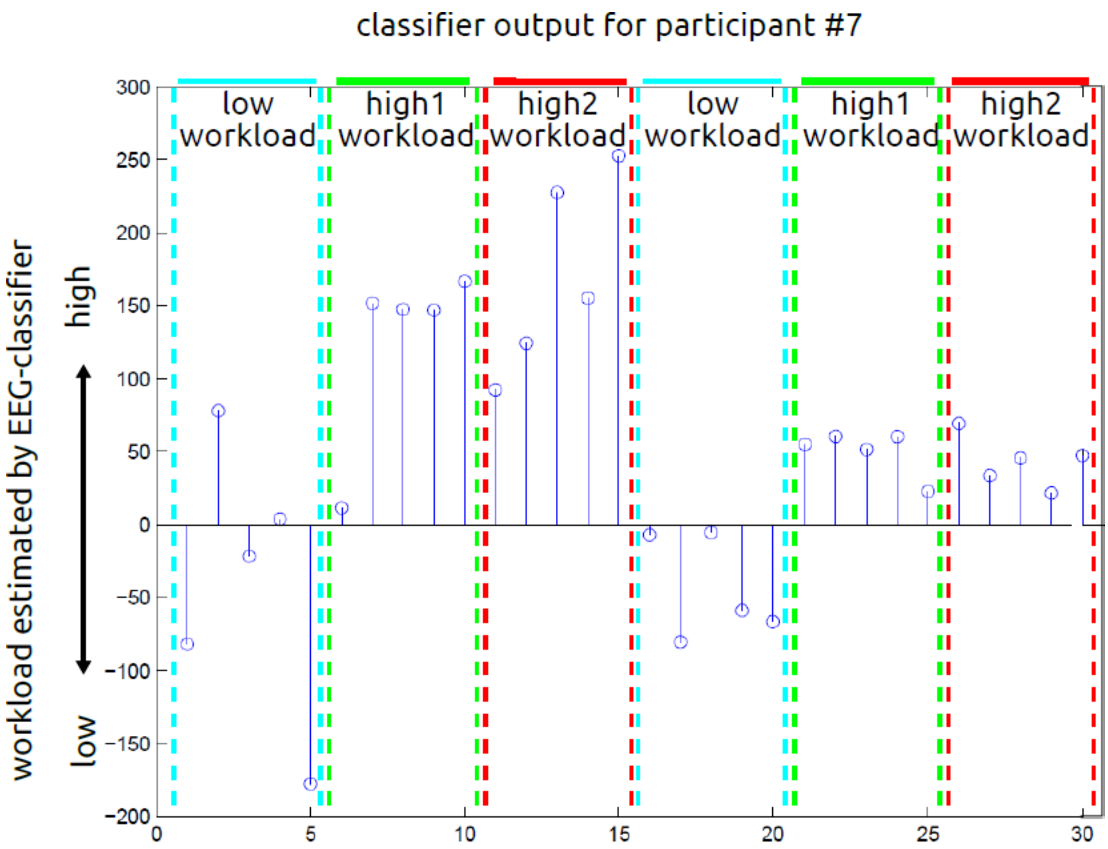
The grand average of the bow-to-bow task is shown in the figure below. The most informative plots are the scalp maps of the r2 scaled difference of high minus low workload condition, which are at the bottom of the figure. The four maps correspond to the four frequency intervals that are shaded in the plot of the power spectral density. There is one for the theta range, two for alpha and one for beta. In general the differences are very small.



Picture #8, Grand average of the spectral analysis of the bow-to-bow task

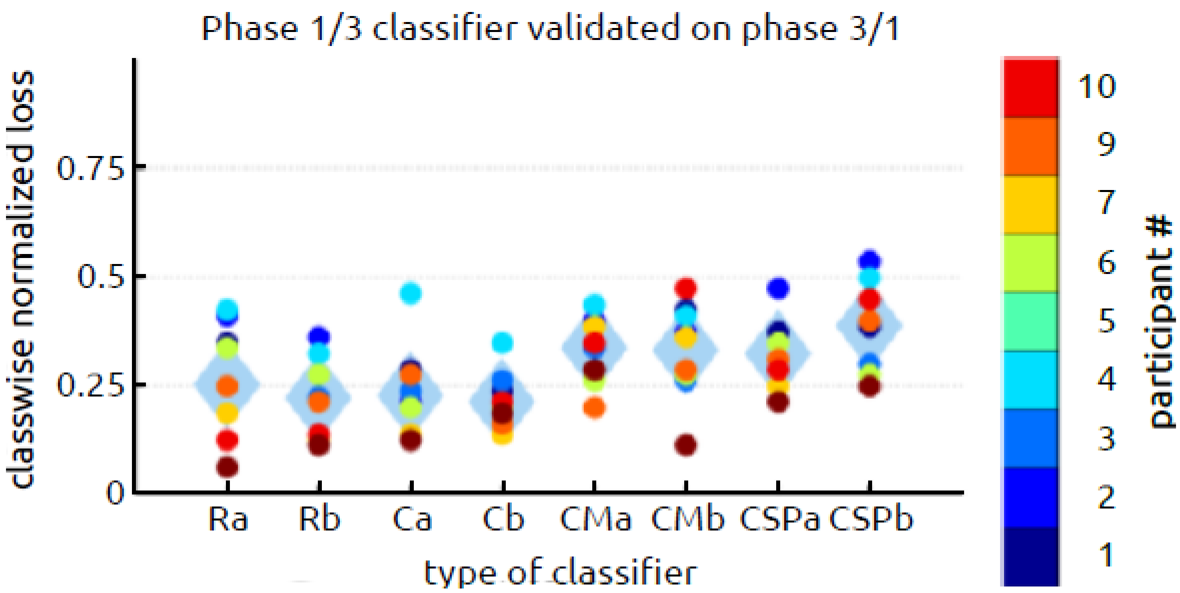
The classifiers were determined by a machine learning tool that finds linear combinations of features that characterizes or separates two events (in our case these events are periods of high workload and periods of low workload).

The classifiers were determined in the first part of each scenario and tested in the second part. It proofed easier to determine accurate classifiers for phase 2 (the N-back task). Transferring the classifier between the two phases 1 and 3 does not degrade the performance appreciably. The results of the classifiers were plotted for each of the simulations done. These plots show the 'linear classifier output' which is a dimensionless value connected to how sure we are about the class (low or high workload). The figure below shows the linear classifier output for candidate seven. The values refer to one minute windows. Negative values correspond to low workload, positive values correspond to high workload. The periods of low workload and periods high-1 and high-2 are indicated in this plot as well. It can be seen that the candidate experienced the first period of high workload more demanding than the second. A possible explanation is the learning effect: He already knows what to do in the second repetition. According to the EEG, five candidates experienced the high-1 condition as more demanding, while four candidates experience the high-2 condition as more demanding.



Picture #9, Linear classifier output candidate 7

The accuracy of the classifiers is expressed in the normalized loss. Different pre-processing methods were used which resulted in different classifiers. Method Ca, which uses MARA artifact removal [17] in a band from 1-20 Hz, worked best, with lowest normalized losses for the different candidates. It should be noted that for each candidate a candidate specific classifier was determined.



Picture #10, Classifier validation for different (pre)processing methods

ECG ANALYSIS AND RESULTS

The ECG measurements were filtered and RR time traces were extracted. The RR time traces were analysed in both time domain and frequency domain. The time domain analysis provides inter beat rates and heart rate frequencies (averages and standard deviations). Furthermore the RMSSD value is given. This is the root mean square of the sequential differences for successive inter beat periods. With spectral analysis the power distribution in both the low frequency (LF) and high frequency (HF) band could be determined. Statistics are given for the absolute power in the LF and HF power band, the LF/HF ratio and the LF and HF normalized power. A comparison was made for different phases of the experiment. Table #3 shows the results per candidate for each of the time blocks: low, high-1 and high-2. Each of these time blocks lasts around five minutes and the values expressed are average values.

#	condition	mean_RR	std_RR	mean_HF	std_HF	RMSSD	LF/HF	LF power	HF power	LF n.u.	HF n.u.	EDR
1	low	0.83	0.031	72.48	2.80	0.013	2.64	1.77E+02	6.75E+01	72.45	27.53	0.182
	high 1	0.82	0.029	73.14	2.61	0.013	6.17	3.80E+02	6.15E+01	86.03	13.94	0.229
	high 2	0.86	0.029	69.86	2.37	0.014	3.87	3.01E+02	7.79E+01	79.40	20.53	0.223
2	low	0.75	0.026	80.18	2.78	0.017	1.91	2.96E+02	1.61E+02	63.56	36.34	0.244
	high 1	0.74	0.027	81.23	2.90	0.016	2.41	2.98E+02	1.27E+02	69.44	30.54	0.234
	high 2	0.75	0.028	80.41	2.98	0.016	3.09	3.73E+02	1.13E+02	72.12	27.85	0.226
3	low	0.71	0.027	84.96	3.24	0.010	15.01	4.09E+02	2.72E+01	93.59	6.40	0.125
	high 1	0.66	0.032	90.91	4.28	0.011	34.53	7.58E+02	2.16E+01	96.77	3.22	0.125
	high 2	0.70	0.030	85.92	3.64	0.010	19.98	4.50E+02	2.67E+01	94.60	5.39	0.125
4	low	0.71	0.040	85.87	5.06	0.012	7.66	3.62E+02	4.93E+01	85.07	14.91	0.133
	high 1	0.70	0.032	86.68	4.10	0.010	4.16	1.97E+02	4.78E+01	80.48	19.51	0.125
	high 2	0.73	0.048	83.32	5.79	0.016	8.31	6.12E+02	7.54E+01	88.11	11.85	0.123
5	low	0.76	0.066	79.72	6.81	0.024	13.16	2.15E+03	1.54E+02	91.73	8.25	0.215
	high 1	0.74	0.044	82.71	4.73	0.017	12.92	7.10E+02	8.93E+01	89.56	10.40	0.193
	high 2	0.82	0.067	73.61	6.14	0.026	8.78	1.49E+03	1.78E+02	88.30	11.68	0.183
6	low	0.82	0.016	73.52	1.46	0.006	7.61	4.83E+01	1.14E+01	81.89	18.07	0.263
	high 1	0.79	0.012	76.45	1.15	0.005	5.09	2.00E+01	4.94E+00	81.88	18.08	0.289
	high 2	0.81	0.016	74.50	1.43	0.006	4.56	4.24E+01	1.16E+01	80.34	19.63	0.256
7	low	0.68	0.019	89.15	2.49	0.008	5.46	1.22E+02	2.53E+01	79.54	20.42	0.298
	high 1	0.61	0.019	99.61	3.09	0.006	4.09	6.54E+01	2.10E+01	75.37	24.50	0.319
	high 2	0.68	0.021	89.56	2.87	0.008	5.56	1.42E+02	2.93E+01	76.20	23.75	0.289
9	low	0.55	0.044	110.86	9.29	0.014	3.06	3.02E+02	1.17E+02	72.85	27.10	0.144
	high 1	0.50	0.043	121.65	10.54	0.014	2.95	3.36E+02	1.20E+02	70.89	29.09	0.151
	high 2	0.51	0.046	118.69	11.76	0.015	2.38	2.95E+02	1.25E+02	69.67	30.29	0.147
10	low	0.80	0.033	75.56	3.19	0.013	6.96	4.19E+02	6.09E+01	86.56	13.42	0.202
	high 1	0.76	0.024	79.37	2.57	0.009	7.17	2.02E+02	2.85E+01	87.21	12.78	0.215
	high 2	0.78	0.026	76.59	2.51	0.011	8.69	3.01E+02	3.43E+01	89.56	10.44	0.198
average	low	**0.73**	**0.034**	**83.59**	**4.13**	**0.013**	**7.05**	**4.77E+02**	**7.49E+01**	**80.81**	**19.16**	**0.201**
average	high 1	**0.70**	**0.029**	**87.97**	**4.00**	**0.011**	**8.83**	**3.30E+02**	**5.80E+01**	**81.96**	**18.01**	**0.209**
average	high 2	**0.74**	**0.034**	**83.61**	**4.39**	**0.014**	**7.25**	**4.45E+02**	**7.46E+01**	**82.03**	**17.94**	**0.197**

Table #3, Summary of ECG measurements

The mean and standard deviation of the inter beat interval RR show consistent decrease in condition high-1 compared to condition low. In condition high-2 the value is mostly comparable with condition low. The average heart frequency shows consistent increase in condition high-1, while the increase is inconsistent in high-2.

We also observe a good correlation between decrease in heart rate variability and increase in workload. The standard deviations of the inter beat interval, heart rate frequency and the RMSSD in general decreased in condition high-1 as may be expected. Changes in RMSSD are small and the measure seems not te be a clear indicator.

The expected correlation between LF and HF power distribution and high/low workload were not found. These measures seem less suitable to use in a reliable workload indicator. Since the workload may not be uniformly distributed in the defined workload periods, it may be that average values cancel out a lot of the available information.

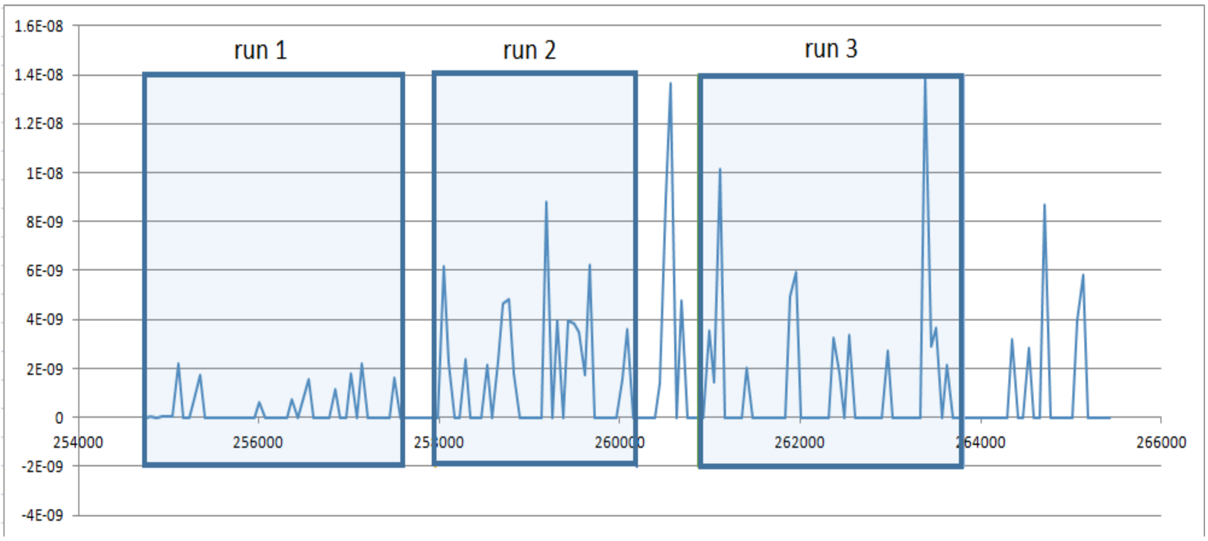
SKIN CONDUCTIVITY ANALYSIS AND RESULTS

The sum of amplitudes of skin conductance responses per minute is plotted against time. This value gives a good indication of the arousal of the candidate. Of the 10 candidates only five were responders. Meaning, that 50% of the measurements gave results that could be analysed. One of them was the participant that was already excluded from the measurements due to sickness, so only results of four candidates remain. The average value of the sum of amplitudes, measured at the dominant hand, of the four remaining candidates is given in table #4. The averages are given for the bow-to-bow pulling and the N-back task scenario. The high workload conditions result in a higher skin conductivity. This is more significant in the N-back task scenario than in the bow-to-bow pulling scenario.

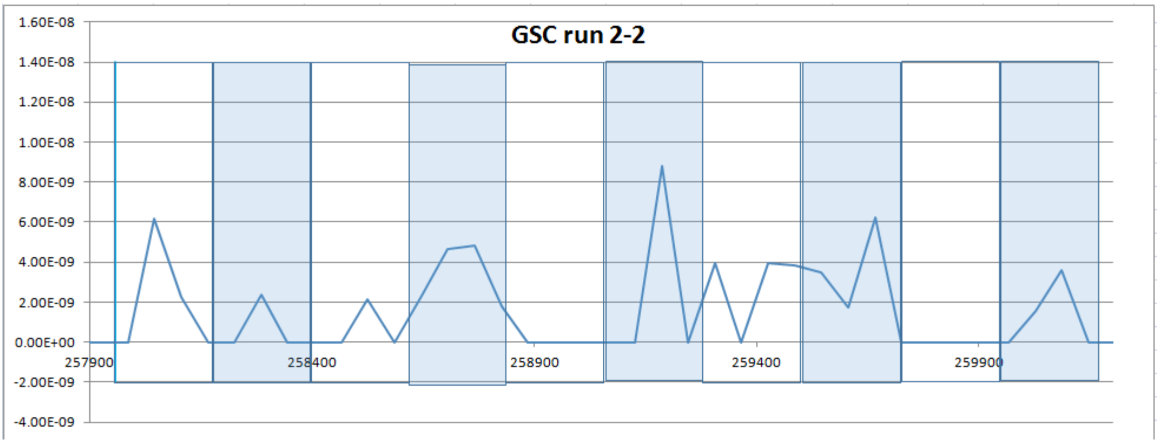
Condition	Bow-bow	N-back
Low	2.012	2.912
High-1	2.227	4.739
High-2	1.710	

Table #4, average values of sum of amplitudes of skin conductance in microSiemens of 4 candidates

An example of the results for one candidate is given in the figures below. The first figure demonstrates the development of the sum of amplitudes of skin conductance responses over time. The blue windows mark the three runs (scenarios) executed in the simulator. The second figure shows a more detailed example of the N-back task for the same candidate. The high workload periods are blue coloured. In most cases, the values are higher during the N-back scenario's. The results for this single candidate suggest a good correlation between the high workload events and the increase in conductivity.



Picture #11, Development of Skin conductivity over time



Picture #12, Detailed plot of Skin conductivity during N-back tasks

COMBINED RESULTS

When results of different physiological measurements and simulator signals are plotted over the same timeline, it is possible to interpret the physiological measures within context of the simulator run. Below an example is given for a candidate executing the bow-to-bow scenario. The low conditions are marked with light blue lines, high-1 and high-2 conditions respectively with green and red lines.

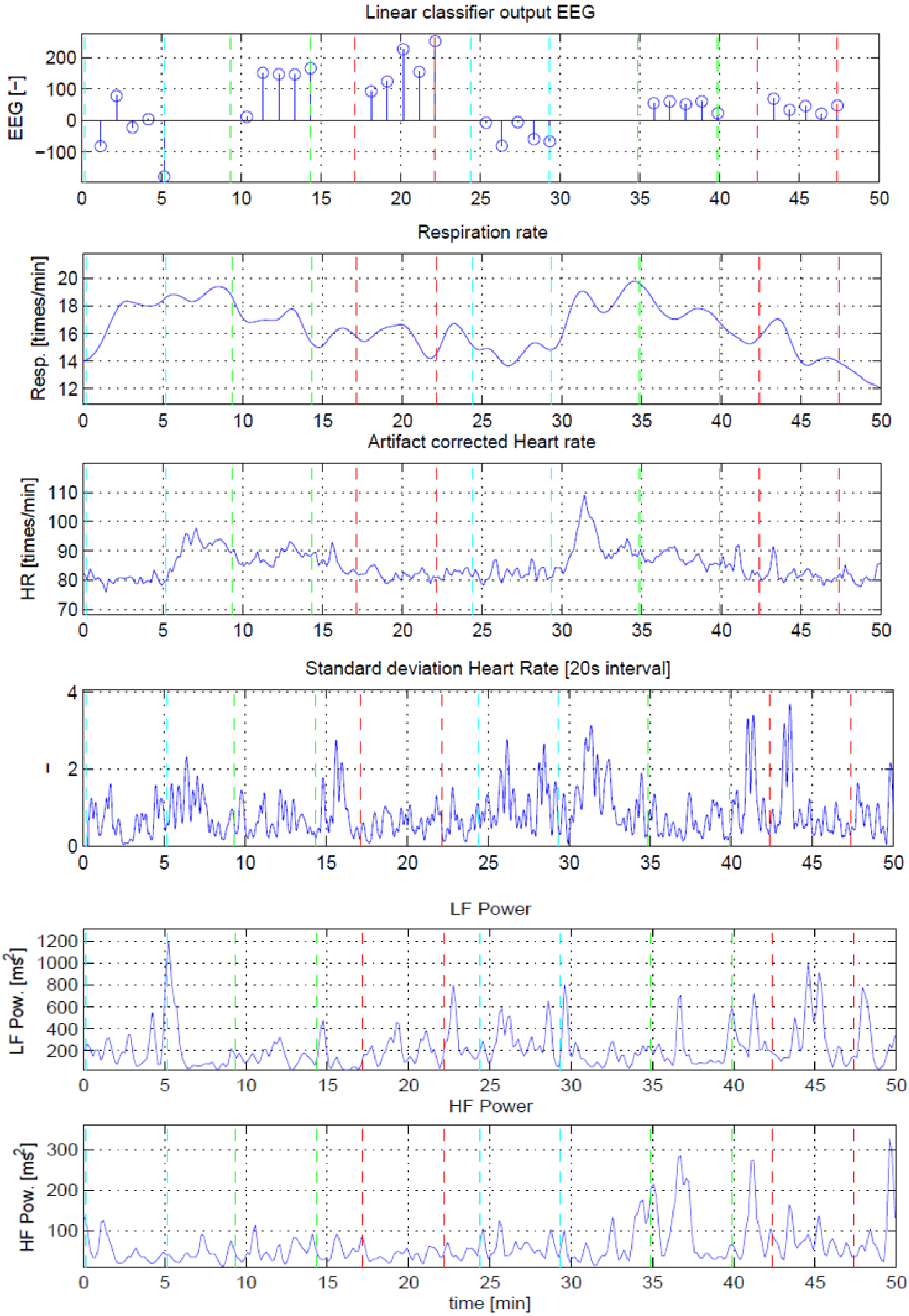
This example demonstrates the common findings that both EEG classifiers and heart frequency are good indicators for mental workload. The decrease in variability of the heart rate frequency was not always as expected. Absolute LF and HF power time traces do not correlate very well with workload.

The indicators give the same signal: keeping station in front of the bow is strenuous.

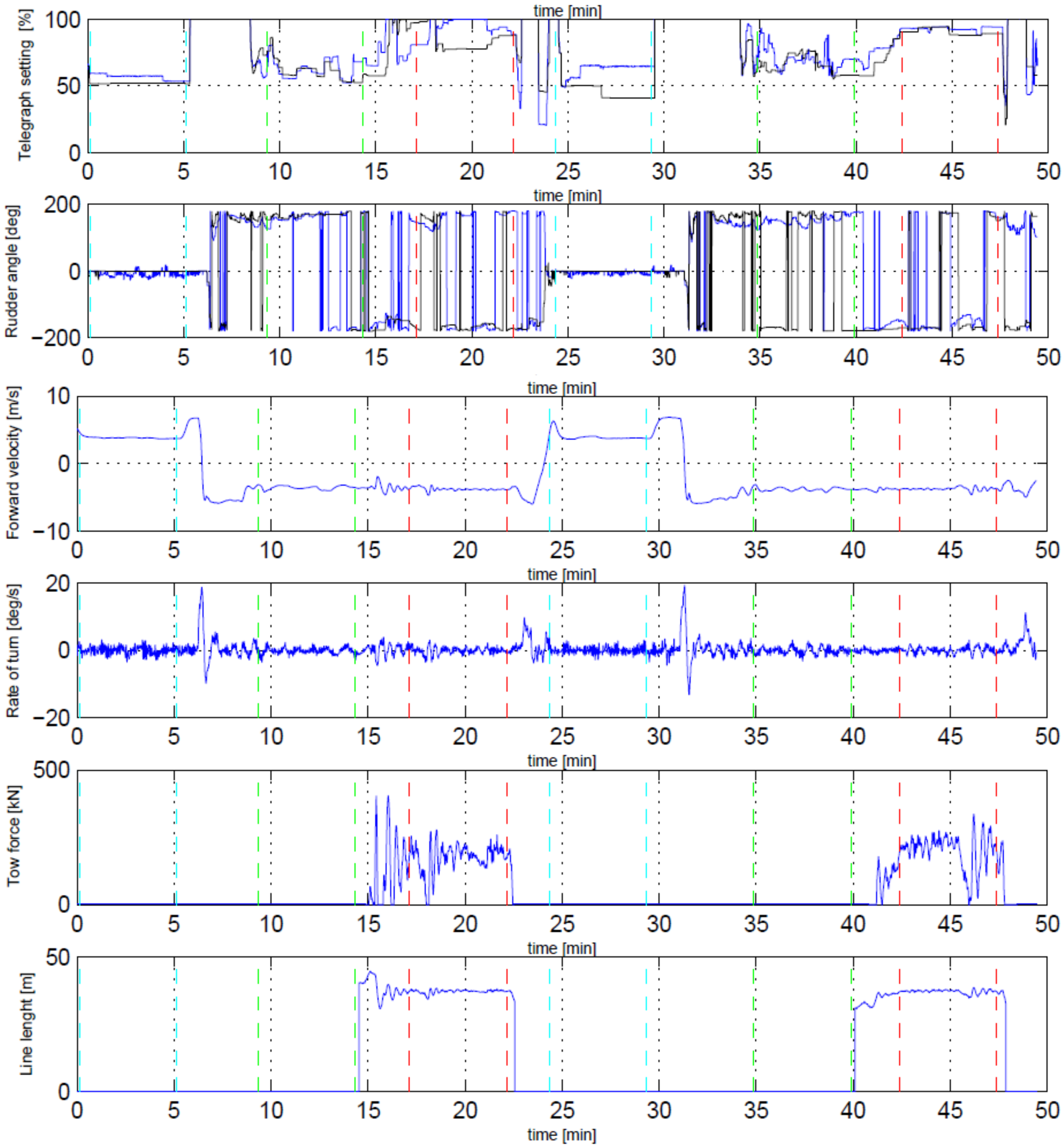
Once we know that we are able to measure workload using physiological measures, we can apply this on the total time trace and study the moments in between the selected 5 minute periods. The additional information about steering and propulsion actions as well as the line length and force help us to complete the total picture of causes and effects on workload.

The highest physiological reactions are seen when the tug is swung around and the tug master performs his approach towards the bow (in between low and high-1 condition). Also for the moment of line connection (between high-1 and high-2) physiological measures indicate a high workload. Once the line is connected the situation becomes easier to handle. On hindsight this can be explained by the fact that the line makes the stationing tasks less difficult.

In periods of increased physiological reactions we also see increased steering and telegraph actions. The changes in rudder- and telegraph settings, and the number of changes within a certain time period, seem to be linked to the workload as well. They are probably valid input for a workload indicator, but this has to be further investigated.



Linear classifier output EEG
EEG [-]
Respiration rate
Resp. [times/min]
Artifact corrected Heart rate
HR [times/min]
Standard deviation Heart Rate [20s interval]
LF Power
LF Pow. [ms²]
HF Power
HF Pow. [ms²]
time [min]



Picture #12, Physiological measurements and simulator signals in bow-to-bow scenario (period 'low' is marked with blue dashed lines, period 'high-1' is marked with green dashed lines and period 'high-2' is marked with red dashed lines.

DISCUSSION OF THE RESULTS

This study shows that it is possible to determine variations in workload measuring different physiological features and that it looks very encouraging that ship related parameters can be used to enhance this process. The outcome can be a relative workload indicator, with which it is possible to determine changes in workload level. Further analyses will be executed by MARIN and TU Berlin to develop the workload indicator.

The experiment also showed that EEG signals are a great contribution, but that it is also very troublesome to carry out experiments using (wet) EEG caps. Not only does it take about 45 minutes to prepare the cap, each participant will also have to go through a calibration session to determine the individual classifiers. However, it is a great tool to calibrate the other measurement techniques and to achieve greater insights in what cognitive processes are involved in manoeuvring a ship.

Heart rate, conductivity and ship signals showed very promising first results, to an extent that the workload indicator may be achieved without the necessity of having to use an EEG system. It should be studied if wire-less dry-cap systems can provide sufficient accuracy in the future, or that other tools can replace the EEG in total.

Although the aim is to develop a generic workload indicator, it is expected that a person specific workload indicator will give better results due to the fact that physiological reactions may differ a lot from person to person. A short calibration session to find the tell tale individual classifiers, may be necessary.

A big by-catch of this study is the amount of appreciation we received from the tug masters who participated in this study. They have to perform a difficult job in sometimes harsh conditions and the work pressure they are under is something that should not be underestimated. They feel that this is not always receiving the attention it deserves.

In this experiment, the captains were sitting in a chair for 3 to 4 hours. Therefore, changes in heart rate due to physical exercise can be excluded. When physical activity is part of the task execution, heart rate itself will be less reliable as a mental workload indicator. Application of the outcome of this study on for example navigators standing on the bridge is not directly possible. It could be necessary to look for other workload measures in other application fields.

ROAD AHEAD

In 2016 MARIN will use machine learning techniques to determine the workload indicator with the signals gathered in this study. This follows the same protocol as used in EEG classification, but now the input signals are not only brain waves, but heart, conductivity, and simulator signals will be included as well. The idea is that the different signals will be weighted and summarized into one signal and cross checked on the simulator data. This signal will be the so-called workload indicator.

We also foresee to review the results of the analysis with the tug masters. The video recordings of the simulations, together with the newly developed workload indicator, will be shown to the tug masters in order to receive their feed-back on it.

It is believed that a reliable workload indicator can be constructed based on the tools that we have studied. However, more tools are available to measure workload. Therefore, we will seek cooperation with other scientific and maritime stake holders to improve the workload indicator. Improvements can be achieved by inclusion of other physiological parameters, like eye-blink rate, pupil diameter etcetera. Another whish is to be able to present the workload indicator real time in combination with video recordings, like the example in the picture below. Further research is necessary to translate the results of our study to other types of ships, operations and sailing tasks. This means more and different equipment and experiments with more candidates. To be able to test the suitability on board, less intrusive tools are necessary.

Picture #13, *Synchronised video recording presented with real time workload indicator.*

ETHICAL ASPECTS:

The conversation always started in the same way, when our test persons got prepared with “hat and wires”: “Now you can look into my brain, can you guess my thoughts?” These questions may seem to be jokes but always come with a glimpse of uncertainty.

It is a big step – adding a clinical or medical touch to the industrial maritime world. Of course we can’t read thoughts, but EEG Data, heart rate, and skin conductivity - they can tell an individual story about a very distinct, highly developed system, namely human biology.

If you are a ship’s engineer, you should learn to deal with systems, cooling waters circuits, exhaust gas systems, fuel etc. In our case, we have to deal with an even more sophisticated and complex individual system. Heartbeat, sweat or brain activity gives us valuable information as to what affects an individual, including job and private life related factors. In this study, the goal was to collect information to evaluate how demanding a certain job task is. This means:

- Dealing with individual personal data,
- Bringing the test person into a stressful situation to observe how and when he reacts under pressure.

It should be regarded as a highly esteemed gesture if someone lets us participate in a situation, where he/she might fail, reaches his/her personal performance limits, e.g. starts sweating, starts showing anger. This is definitely a undesirable situation for the test person, especially since it takes place under observation by a stranger. It needs quite a bit of self-confidence to step voluntarily into such a situation. But fortunately, there are persons, who are open to supporting us, allowing the scientists at the TU Berlin and in Wageningen at MARIN to do their research. They participate because they are convinced of the “Why?” that stands behind this all.

This deserves a huge amount of appreciation and of explanations about what is going to happen with their data. The right of transparency of the “how?”, “why?” and “what for?” belongs to them. An ethical frame and guidelines should be followed.

In our study, the data are anonymized. It is, so to say, a sheltered environment, because there is no connection to any work-relationship, no data flow to any boss or company. We only tested adults. And the amount of discomfort is related to what is expected in reality. The results are used to find an objective scale of high and low workload, independent from individual influences: a workload index for the maritime branch. A tool, which should help to increase safety on board of ships.

What if such a tool is used in the future for assessment purposes or when it is applied in an employee-boss team setting? This would bring us to another discussion. A combination of personal rights, labour and employment rights, as well as philosophy and business ethics and – identity will need to be regulated.

ACKNOWLEDGEMENTS

We wish to thank the tug masters who voluntarily underwent these strenuous tests on the bridge simulator. They contributed tremendously with their patience and cooperative attitude.

REFERENCES

[1] De Waard, D. (1996). The measurement of drivers' mental workload. Groningen, The Netherlands: Traffic Research Centre, University of Groningen.

[2] Rouse, W.B., Edwards, SL & Hammer, J.M. (1993). Modelling the dynamics of mental workload and human performance in complex systems. IEEE transactions on systems, man and cybernetics, 23, 1662-1671.

[3] Wilson, G.F., & Fisher, F. (1991). The use of cardiac and eye blink measures to determine flight segment in F4 crews, Aviation, Space and Environmental Medicine, 62, 959–961.

[4] Revati Shriram, Dr. M. Sundhararajan, Nivedita Daimiwal (1). EEG Based Cognitive Workload Assessment for Maximum Efficiency, IOSR Journal of Electronics and Communication Engineering

[5] Alan Gevins, Michael E. Smith, Harrison Leong, Linda McEvoy, Susan Whit_eld, Robert Du, and Georgia Rush. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. Hum Factors, 40(1), 79-91.

[6] A. Holm, K. Lukander, J. Korpela, M. Sallinen, and K. M. Muller. (2009). Estimating brain load from the EEG. ScientificWorldJournal, 9, 639-651.

[7] M. E. Smith, A. Gevins, H. Brown, A. Karnik, R. Du, and A. S. Gevins. (2001). Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction. Hum Factors, 43, 366-380.

[8] Alan Gevins and Michael E. Smith. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. Theoretical Issues in Ergonomics Science, 4(1-2), 113-131.

[9] David Galin, Jack Johnstone, and Jeannine Herron. (1978). Effects of task difficulty on eeg measures of cerebral engagement. Neuropsychologia, 16(4), 461-472.

[10] Jens Kohlmorgen, Guido Dornhege, Mikio Braun, Benjamin Blankertz, Klaus-Robert Muller, Gabriel Curio, Konrad Hagemann, Andreas Bruns, Michael Schrauf, and Wilhelm Kincses. (2007). Improving human performance in a real operating environment through real-time mental workload detection. In Guido Dornhege, Jose del R. Millan, Thilo Hinterberger, Dennis McFarland, and Klaus-Robert Muller, editors, Toward Brain-Computer Interfacing, pages 409-422. MIT press, Cambridge, MA.

[11] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Muller. (2008). Optimizing spatial filters for robust EEG single-trial analysis. IEEE Signal Process Mag, 25(1), 41-56.

[12] Omkar N Markand. (1990). Alpha rhythms. J Clin Neurophysiol, 7(2), 163-190.

[13] Wilson, G.F. (1992). Applied use of cardiac and respiration measures: practical considerations and precautions. Biological Psychology, 34, 163-178.

[14] Porges, S.W. & Byrne, E.A. (1992). Reasearch methods for measurement of heart rate and respiration. Biological Psychology, 34, 93-130

[15] Backs, R.W. & Seljos, K.A. (1994). Metabolic and cardiorespiratory measures of mental effort: the effects of level of difficulty in a working memory task. International Journal of Psychophysiology 16, 57-68.

[16] Christian Kothe, Lab Streaming Layer, https://github.com/sccn/labstreaminglayer"

[17] Irene Winkler, Stefan Haufe, and Michael Tangermann. (2011). Automatic classification of artifactual ICA-components for artifact removal in EEG signals. Behav Brain Funct, 7(1),30.