# ADVANTAGES AND OBSTACLES OF APPLYING PHYSIOLOGICAL COMPUTING IN REAL WORLD: LESSONS LEARNED FROM SIMULATOR BASED MARITIME TRAINING

(DOI No: 10.3940/rina.ijme.2017.a2.404)

**Y Wu**, **T Miwa**, and **M Uchida**, Graduate School of Maritime Sciences, Kobe University, Japan

**SUMMARY**

While simulator based maritime training is widely implemented under international maritime organization (IMO) convention and model courses, troublesome issues such as objective evaluation of training effectiveness remain unsolved. Physiological computing system (PhyCS) refers to an innovative bidirectional human computer interaction which is achieved by monitoring, analysing, and responding to operators' psychophysiological activities in real-time. With the development of wearable devices, it becomes promising to apply PhyCS, which was considered as a laboratory technology, in real-world scenarios. In our experience utilizing view tracker, portable heart beat sensor, electroencephalogram device, and web-cameras in simulator based maritime training, PhyCS shows potential for advanced applications in operator performance assessment, usability tests, and adaptive training. However, ambulatory working environment, body movement artefact, and model verification are intricate obstacles that constrain its applications in the real world. By examining the advantages and obstacles, this paper aims to develop guidelines to apply PhyCS in the real-world.

## 1. INTRODUCTION

Human error has been identified as the most crucial contributor to accidents in many domains (Baker & McCafferty, 2005; Griffith & Mahadevan, 2011). Providing high quality training is a promising way to improve operator's skills and to reduce human error (Wu *et al.* 2015). In professional training, including training of pilots, advanced seafarers, and surgery operators, high fidelity simulators are widely used to help trainees practice routine and emergency operation procedures with lower cost and shorter time. There are also cases when simulators can provide learning experience beyond that which can be learned in actual systems, for example, in teaching students systematic trouble shooting skills, a simulator that can malfunction any single component in a system may provide more effective training than a real system. In addition, training procedures and instructions can be carried out at a pace that ensure trainees can maintain engagement with training tasks, and keep their mental workload (MWL) at the optimal point, where the learner is neither overloaded nor under loaded. To improve the training regime, institutions need to adjust their pedagogics based on students' mental capacity and engagement throughout the learning process.

Along with the rapid development of virtual reality technologies, many innovative marine simulators have appeared. In the 12$^{th}$ (2015) international conference on engine room simulators, primary manufacturers exhibited prototypes of ship simulators, for instance, head mounted virtual reality devices, computer based 3-dimensional virtual roaming in engine rooms, and simulators with functions of energy saving operation practice. The biggest obstacle for training institutions to adopt these prototypes is proof as to the degree training effectiveness can be improved compared to that of traditional simulators. In other words, the usability of these newly developed products must be carefully studied.

Physiological signals, such as neural brain activities extracted from electrical voltage fluctuations of scalp, eye fixation, pupil diameter, and blink frequency measured by view trackers, and heart rate, provide an unobtrusive and objective method to infer the operator's mental (affective and cognitive) state as well as physical state, although these relationships are mostly implicit. Physiological computing systems (PhyCS) correlate explicit physiological information with implicit operator functional states (OFS). PhyCS is still in its infancy, but it has enormous potential to innovate human computer interaction by extending the communication bandwidth to enable the development of 'smart' technology (Fairclough, 2009). There are numerous types of physiological indices that can be measured in real-time by wearable devices or web cameras. These evaluation indices are often used for MWL (or cognitive workload) measurement (see Table 1) since MWL is considered as one of the core elements of human factor constructs. Exorbitant MWL results in stress, whereas an accumulation of lower workload contributes to task disengagement, boredom, and drowsiness. In a study of BMW group research and technology, Hajek *et al.* (2013) developed a new generation driver assistance system. The system parameters (e.g. cruise speed) adapted to drivers' MWL measured by heart rate, galvanic skin response and respiration, and the results estimated from this physiological data revealed an advantage of workload adaptive cruise control over traditional cruise control.

As shown in Figure 1, the realization of PhyCS generally contains the following steps: a) collect physiological (bioelectrical, biomechanical or biochemical) data through wearable sensors, biomedical devices or image analysis techniques; b) conduct pre-processing and extract valuable features from physiological signals; c) correlate the features with operator's affective/cognitive state based on regression models or machine learning algorithms. The appearance of low cost physiological computing devices, including view tracker, heart beat sensor, and electroencephalogram (EEG) mapping devices are extending the methods of simulator based education and training. For instance, during training to operate an engine control room console, view trackers are used to record the operator's gaze fixation, and from this the instructor can monitor which part of the console the operator pays most attention to. However, actual maritime environment has specific characteristics that we have to consider if we are to apply PhyCS in seafarer training. We have already introduced view tracker, portable heart beat sensors, EEG devices, and web-cameras in the marine engine plant simulator (MEPS) at Kobe University for training and research. This paper aims to report the lessons we have learned from applying PhyCS in simulator based maritime training and to develop guidelines for PhyCS applications in real world scenarios.

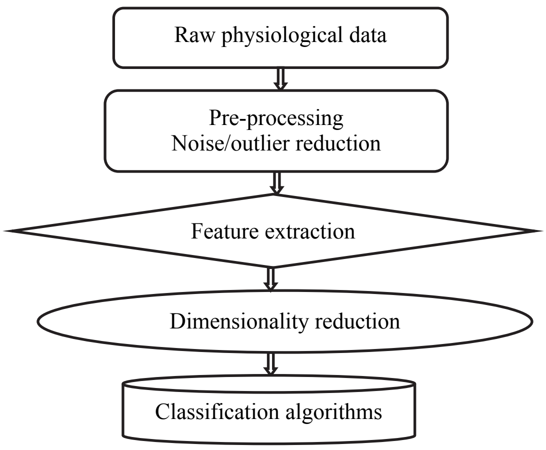


Figure 1: Flow chart of physiological computing

## 2. PHYSIOLOGICAL INDICES

Lean and Shan (2012) briefly reviewed the physiological and biochemical evaluations of human cognitive states and categorized physiological metrics into three main categories based on neurophysiological taxonomy. However, they did not report on the varied validity of these metrics when applied in real or quasi-real environments. In this section, in order to consider practical application, three types of physiological signals and their main features are reported according to the sensors used.

### 2.1 HEART RATE RELATED INDICES

Heart rate and heart rate variability (HRV) analyses are used for evaluating autonomic nervous system activities and are defined as peripheral physiological indices. Heart beat sensors are generally low-cost, simple and user friendly, and inobtrusive. Applications of heart rate related indices are becoming more and more a part of standard physiological monitoring. In addition to absolute heart rate, time domain, frequency domain, and nonlinear indices are also used as physiological computing inputs (Tarvainen *et al.* 2014). Typical heart rate related features include average heart beat interval, standard deviation of heart beat interval, LF/HF ration based on Discrete Fast Fourier Transformation (DFFT), where low frequency (LF) is defined as 0.04-0.15Hz and high frequency (HF) is defined as 0.15-0.4Hz.

### 2.2 BRAIN WAVE INDICES

Functional brain imaging methods including EEG, functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and positron emission tomography (PET), enable the study of cognitive and sensorimotor functions of the human brain across a wide range of behaviours (Kerick *et al.* 2009). EEG is used to map brain electrical activity. By attaching a set of electrodes to specific areas of the scalp, EEG measures voltage fluctuations resulting from ionic current within the neurons of the respective brain area. EEG signal features fall into two categories: short term time domain and power spectrum. Event related potentials (ERP) measures brain response fluctuations that are related to a specific sensory, cognitive or motor event after a particular time delay. Prinzel *et al.* (2003) used the P300 component of ERP to assess participants' task engagement and performance in an adaptive automation situation. EEG waveforms are usually estimated by wavelet transformation or DFFT. Power spectral of EEG are divided into several bands: delta (1-4 Hz), theta (4-7 Hz), alpha (8-15 Hz), beta (16-31 Hz), and gamma (32+), which sometimes can be slightly disparate. The power of these bands or their relative percentage of total power bands are often used as physiological indices (Borghini *et al.* 2014).

### 2.3 EYE TRACKING TECHNOLOGY

View trackers are used to record operators' pupil diameter, blink interval, and gaze fixation. For a typical eye tracking device, a high resolution infrared camera is set to record a video of an eye or face. After image binarization processing, threshold values are chosen to recognize the centre of the pupil based on pixel differences. Thus, both pupil diameter variation and eyeball movement can be continuously recorded. Furthermore, by calibrating the subject's eye fixation point before normal recording, his or her view path can also be recorded. One popular pupil metric that relates to cognition function is saccade, which is the fast unconscious movement of the pupil. Siegenthaler *et al.* (2014) found that task difficulty in mental arithmetic affects micro saccadic rates and magnitudes. Wanyan *et al.* (2014) found that pupil diameter and blink interval are effective to infer human MWL.

 

Table 1: Studies that use physiological indices to infer cognitive states in transportation domain

| Study | Subjective measure | Tasks | Difficulty setting | Samples Analysis Method | Statistical results |
|---|---|---|---|---|---|
| Durantin *et al*. (2014) | NASA TLX | Computer simulated air piloting; n-back sub task and auditory alarm task | With/without Cross wind during piloting; 1/2-back spatial task; (2*2) 4 sessions | 12; ANOVAs; regression | fNIRs<br>Piloting difficulty to Oxygenation $F(1,11)=5.82$<br>Interaction to Oxygenation $F(1,11)=5.11$<br>Oxygenation to Performance $R^2=0.52$<br>Inverted u-shape of oxygenation |
| Stuiver *et al*. (2014) | RSME | 6*1.5h Driving simulator | Fog/no fog Low/high traffic density | 15; Repeated MANOVA 15*6=90 | Interaction to High Frequency spectral HRV $F(1,89)=4.98$<br>Interaction to Blood pressure $F(1,89)=13.42$<br>Ceiling effects of extremely high workload |
| Ayaz *et al*. (2012) | NASA-TLX | n-back (28 sessions) Air traffic control by simulator (6 sessions) | n-back voice/data; number of aircrafts(6,12,18) | 24; ANOVA | fNIRs<br>“n-back” to Oxygenation $F(3,69)=4.37$, $p<0.05$;<br>Vehicle number to Oxygenation $F(2,42)=4.52$ |
| Wanyan et al. (2014) | NASA-TLX | Three phases of flight simulation, monitoring flight indicator | Number of indicators monitored. Four levels | 12; One-way repeated ANOVA | Main effect<br>HR to MW, $p=0.252$. RRCV to MW, $p=0.019$<br>Main effect<br>Pupil diameter $p=0.076$<br>ERP P3 peak amplitude $p=0.049$ |
| Lee and Liu (2003) | NASA-TLX | Four stages of flight simulator. Difficulty by TLX | Take-off, climb and cruise, descent and approach, Landing | 10; One-way ANOVA | $RMS_{R\_R}$<br>$p<0.0001$ |

We reviewed recent MWL evaluation literature, to find effective physiological indices for the transportation domain (Table 1). However, the results are ambiguous because of the disparate experiment setting. Flight simulator and driving simulator are frequently used to simulate the operating environment in research but different physiological metrics are reported as valid and sensitive to infer operators' MWL. The sensitivity of these metrics are tested either by the significance of Analysis of Variance (ANOVA) or regression models in distinctive levels of task difficulty. HRV features and cerebral cortex activity measured by fNIRs or EEG are widely used because the equipment is relatively economical and requires little medical expertise. Compared to that of civil aviation, MWL evaluation research in merchant shipping seems quite inactive (Young *et al*. 2015 Table 2). One of the reasons is that there are considerable obstacles in applying PhyCS in a ship environment, and in Section 4 we will provide recommendations to deal with some of them.

## 3. ADVANTAGES AND APPLICATIONS

### 3.1 PERFORMANCE ASSESSMENT

There are two general categories of operator performance assessment: subjective rating scales and objective measures. Subjective rating scales can be collected based on either operator's autognosis or evaluators' judgement. Their low cost and ease of administration, as well as adaptability, make subjective measures widely used in a variety of safety-critical domains. However, the weakness of a subjective rating scale as a performance assessment method, is its dependence on the operator and their time and ability to record their feelings. Shortcomings such as individual bias, serious intrusiveness and discrete sampling, limit its applicability. Therefore, some other methods have been developed to evaluate operator performance. For instance, in order to avoid drowsiness of officers on watch, the international convention of Safety of Life at Sea (SOLAS) currently requires all ships above 150 tonnes to install a Bridge Navigation Watch Alarm System (BNWAS), to which the officer on watch has to respond by either directly pressing specific buttons or having their movement be detected within a pre-set time interval, say, 12 minutes. However, the high intrusiveness caused by BNWAS and the usability of BNWAS are problematic from the view of ergonomics.

Contrarily, operator's physiological signals can provide an objective, continuous and unobtrusive method to conduct performance assessment of key operators' non-technical skills that are crucial for system safety, such as leadership and communication skills. Different types of physiological information have been used to infer human affective states as well as cognitive states (Mehta & Parasuraman, 2013; Touryan *et al.* 2016; Wei *et al.* 2016). In the transportation domain, researchers successfully used neurophysiological signals to evaluate aircraft pilots' and car drivers' mental workload, fatigue and drowsiness (Borghini *et al.* 2014). In a marine engine room, engineers' cognitive states are crucial because they have to monitor a large number of system parameters and cope with abnormal situations. PhyCS provides insight into operator's cognitive states, and offers a method to objectively evaluate their performance and skills in dynamic human interactive computer system.

### 3.2 USABILITY TEST

With the emphasis on human centred design in many industries, human-oriented systems have also emerged in modern ship design. For a ship, favourable usability means that operators can accomplish required tasks with efficiency, effectiveness, and self-satisfaction by using the limited on-board resources. In developing innovative maritime systems such as the modern ship bridge shown in Figure 2, designers and shipbuilders need to avoid creating distributed interfaces that become technology "barriers" (Lützhöft 2004). Here there are 18 interfaces in total that enable operators to track the status of variables' online through indicators located at different positions on the ship's bridge. However, these screens and indicators may be inadequately placed on the bridge considering the necessity of operator movement and timely information acquisition. One of the most important goals of a usability test is to discover major problems in the user interface that could result in human error, terminate the interaction, and/or lead to user frustration (Papachristos *et al.* 2012). Papachristos *et al.* (2012) also argued for the necessity of a mixed approach, combining questionnaires, gaze tracking and speech recording for a usability evaluation of the ship's bridge. Eye tracking technology has also been used to study pilot scanning across a high fidelity automated 747 cockpit, and a model was developed to predict the distribution of attention so that the design of alerts is noticed (Wickens *et al.* 2009).

In the maritime domain, Gould *et al.* (2009) used HRV and skin conductance as MWL measurements and examined the effects of two different navigation methods, the conventional system using paper charts and an electronic chart display and information system (ECDIS), on workload and performance in simulated high-speed ship navigation. Their results indicated higher workload in conventional navigation, although the difference between the groups was not significant. For bridge and engine room simulators, as higher fidelity does not necessarily mean higher efficiency in training, designers should pay attention to maintain an optimal point of fidelity considering the balance of cost and training improvement. For instance, in whether to choose a simulator with or without vibration and sound feedback, usability should be tested by using PhyCS or some other method so that the conclusion is persuasive.

Figure 2: Modern Ship Bridge with 18 interaction screens (Pan et al. 2015)

### 3.3 ADAPTIVE TRAINING

While simulator based training has been widely used in maritime education institutions, we hope to further improve the effectiveness of training by providing individualized instruction. Wiltshire & Fiore (2014) argued for the advantages of training where the trainee's social and affective cognition state can be handled timely. As learning is most efficient when working memory resources are managed effectively, efficient instruction presents training materials at a pace and in a format that keeps a trainee engaged and motivated without overloading his or her limited working memory resources (Baldwin & Penaranda, 2012). A trainee's functional state can be continuously assessed based on physiological signals. Thus, instructors are able to implement instructions at an adaptive pace according to the trainee's functional state (e.g. mental arousal, boredom) throughout the training process. In a simulated environment, one of the most significant challenges imposed by implementing PhyCS based adaptive training, is that all the data processing and computing must be accomplished in real time, which relies on efficient integration of software and hardware design.

## 4. OBSTACLES

With the development of more user-friendly wearable devices, PhyCS shows particular advantages in reducing human error and improving simulator based training. However, the characteristics of a maritime operation posed inevitable obstacles to the application of this technology in real-world environment where work is accomplished dynamically. Several of these obstacles and recommendations to avoid them are discussed in this section.

### 4.1 REQUIRING AMBULATORY DEVICES

Compared to the work environment of aviation pilots and vehicle drivers, where operators sit in relatively fixed positions and focus mainly on cognitive tasks, a marine engineer has to move around as part of their routine, either in an engine room or engine control room, and their work includes a number of physical tasks. Figure 3 shows MEPS in Kobe University. It consists of an engine room, a centralized control console, and an instructor's space. In both training courses and research experiments, subjects have to complete required tasks in an ambulatory situation. Therefore, wearable devices that reliably collect and wirelessly transform physiological data are necessary for applications in a ship environment. Kerick *et al.* (2009) argued that some neurophysiological measurement technologies, e.g. fMRI, MEG and PET can be generally ruled out due to their machinery size. Furthermore, while cognitive ergonomics studies can be conducted in immobile participants, research on embodied cognition has shown that cognitive processing when moving and interacting in the physical world may have unique characteristics that can only be captured with mobile physiological sensors (Mehta & Parasuraman, 2013).

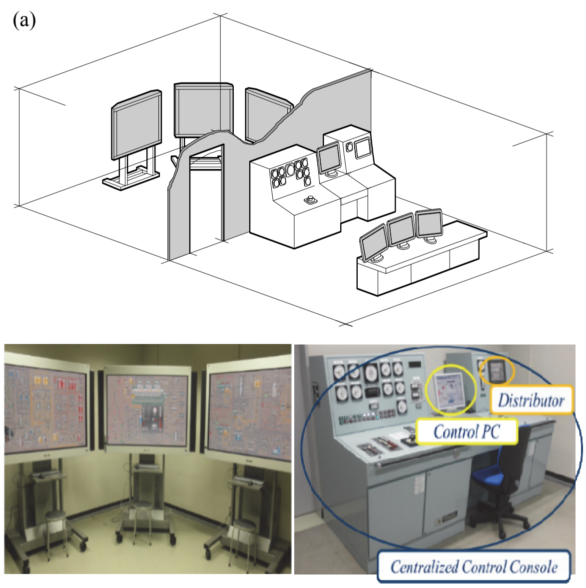


Figure 3: MEPS (a) outline of MEPS; (b) real picture of simulated engine room; (c) control console

To date, various ambulatory sensors, including chest type heart beat sensor (Polar Electro, Finland), glasses like eye trackers (Tobii Technology, Sweden) and integrated cap EEG device (Emotive Systems, Australia), have been manufactured and reported valid in academic publications (Cheng and Vertegaal, 2004; Kingsley *et al.* 2005; Gamelin *et al.* 2006; Ramirez and Vamvakousis, 2012).

### 4.2 BODY MOVEMENT ARTEFACT

Unlike medical electrocardiographic (ECG) devices, portable heart rate monitors such as POLAR RS800 (POLAR, Finland) generally provide extracted heart beat interval (HBI) data rather than raw ECG data. Even a single ectopic beat caused by body movement and/or poor sensor contact can have a serious impact on the interpretation of the results, especially for short-term cognitive state classification. Peltola (2012) argued for the necessity of editing raw HBI data, and appropriate artefact correction methods must be chosen according to different study settings. Although there is still no evidence that one HBI correction method is obviously better than another, three basic principles for HBI editing in PhyCS should be followed. Firstly, automatic algorithms should be performed since manual correction is less accurate and more time-consuming. Secondly, interpolate ectopic beats instead of deletion, since deletion of data points could introduce errors to power spectrum features such as LF/HF. Thirdly, avoid output delays resulting from artefact correction methods that use latter beats to interpolate former beats. Therefore, we recommend using mean value interpolation to correct ectopic beats. The detection and correction of ectopic beats is conducted as follows:

$$if\ x_n > (1+t_1) \times x_{n-1}\ or\ x_n - \bar{x}_n > 3 \cdot SD(x)$$

$$x_n = 1/t_2 \sum_{i=n-t_2}^{n-1} x_i$$

Where $t_1$ and $t_2$ are threshold values that control the intensity of artefact discrimination and correction. The second recognition condition is that the difference to mean is bigger than three times standard deviation of raw data. It is relatively conservative as the raw data also contains artefacts that cause high values of standard deviation. The red line in Figure 4 shows points that are detected as ectopic beats and the green line is the result after artefact correction.

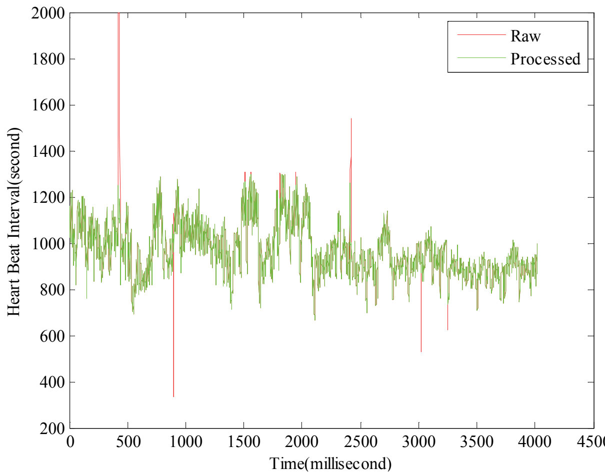


Figure 4: Example of HBI interpolation

An EEG device generally uses silver electrodes to measure the extremely weak signal (µv) of voltage fluctuations along the scalp, and the signal is almost always contaminated by artefacts resulting from different sources. These sources include biological activities (muscle, eyeball, cardiac etc.), baseline artefact, powerline noise, and body movement artefact. In the real working environment, body movement artefact is the sum of electrode-scalp interface impedance fluctuations caused by walking, loud talking, irregular ship motions, and head movement. EEG wave bands of up to 40 Hz have been found useful in evaluating human cognitive states (Borghini *et al.* 2014). It is still difficult to know to what degree and in what power band EEG data is contaminated by body movement artefact. Ferris and colleagues used advanced hardware settings and algorithms (e.g. independent component analysis) to remove gait related movement artefact in experiments of subjects walking and running on a treadmill (Gwin *et al.* 2010; Kline *et al.* 2015). Although their work showed some promising results, regular gait events do not fully represent the complex nature of a working environment on a ship.

We used a portable EEG device (Degital electronic, Japan) with two channels. Channel 1 is for scalp voltage measurement (EEG electrodes), and channel 2 is an accelerometer attached directly to the electrodes to measure electrodes vibration. EEG epochs that are contaminated by movement artefact can be detected based on the power of channel 2. Figure 5 shows the signal measured by EEG electrodes and accelerometer in three body movement conditions: motionless, speaking, and walking around. The respective frequency domain of each condition is estimated from the epochs and marked by the dotted rectangle, of which the standard deviation of time series of channel 2 is relatively large. Compared to the stable signal in motionless condition, the amplitude fluctuations in speaking condition is acute and even more acute when walking around. Accordingly, while EEG wave bands in motionless condition are rarely affected by body movement artefact, speaking may affect EEG wave bands of about <15Hz. Possible harmonic oscillations occur around 5Hz, 9Hz, and 12Hz, which are components of theta wave (4-7Hz) or beta wave (8-15Hz). The influence of body movement artefacts on EEG signal is much more obvious in the walking around situation. EEG signal is almost fully contaminated through all effective wave bands during continuous fast walking, and when turning inside a room. Although it is definitely true that some other EEG devices with more sophisticated circuits might be less vulnerable to body movement artefact, we cannot ignore this problem in research and development (R&D) of applying PhyCS in a practical working environment.

### 4.3 OFS MODELLING AND ITS VERIFICATION

Fairclough (2009) in reviewing the fundamentals of PhyCS explicated the complex relationships between physiological indices and psychological states: one to one, many to one, one to many, and many to many. This requires researchers to focus on functional states that relevant to operator performance and avoid confounding factors affecting the interpretation of physiological signals. In R&D of PhyCS, designing standard tasks and real-world tasks should comply with two basic principles. Firstly, since accuracy of cross-task classification was reported obviously lower than within-task classification (Baldwin & Penaranda, 2012), standard tasks that provide training data should consume cognitive resources of similar quantity and dimension (e.g. visual, auditory) in comparison with those of real-world tasks. Secondly, to obtain ground truth of OFS, task complexity should be manipulated, and verified by other measures such as subjective rating scale and task reaction accuracy. Take the example of MWL assessment, while the complexity of standard tasks can be orderly designed within experimental psychology software tools such as E-prime (Psychology Software Tools, Inc.), task complexity of maritime operations is perhaps the most multifaceted performance shaping factor. There are a wide range of factors, such as a lack of necessary training and experience and poor interaction interface, which can make a task subjectively complex or difficult. Since the research purpose is to elicit different levels of MWL for each subject, we suggest focusing on task analysis when considering complexity. For example, we use the number of operation steps, the number of subsystems involved, and their interactions to quantitatively manipulate the complexity of maritime operation tasks.

Another obstacle to apply PhyCS in maritime ergonomics is the requirement of expertise knowledge in two aspects. Firstly, estimating cognitive states from physiological signals and designing applications based on these models, requires multidiscipline expertise ranging from neurophysiology, statistics, machine learning, experimental psychology, to engineering. It is a significant challenge for researchers to have expertise in all these areas and they are faced with problems designing research protocol, data processing, analysing experiment results, and drawing warranted conclusions. For instance, non-experts, especially those with engineering backgrounds tend to regard neurophysiological data as conveying an objective truth even though other evidence indicates otherwise (Brouwer *et al.* 2015). Secondly, a large number of marine engineers or cadets are essential as subjects for behavioural experiments, however numbers are rarely available even for simulator environments.

 

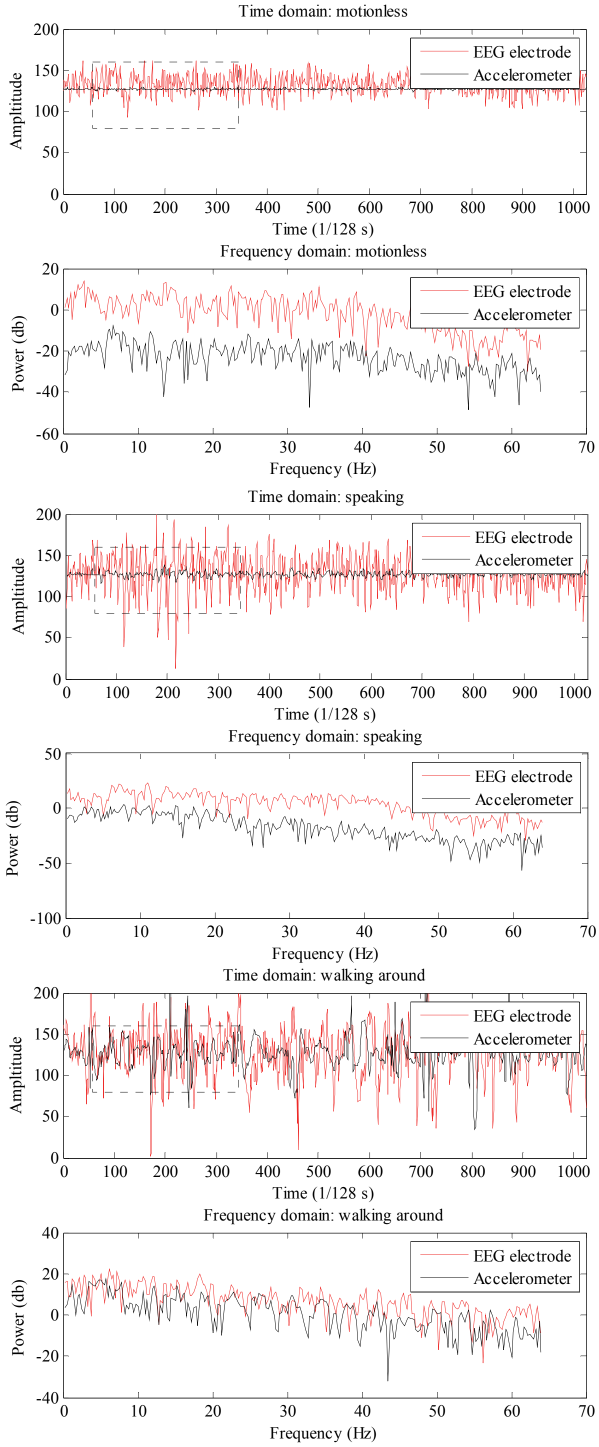


Figure 5: Time domain and frequency domain analysis of EEG in three situations (Top to bottom): a. motionless; b. speaking; c. walking around. Frequency domain is estimated from the epochs marked by dotted rectangle. Red line is EEG electrodes signal, black line is accelerometer signal.

## 5. CONCLUSIONS

With the rapid development of wearable devices and online computing capacity, physiological monitoring is becoming cheaper, more user-friendly, and more reliable. Meanwhile, although human factors have long been found to be the primary cause of maritime accidents, there are few practical proposals to effectively control human error and to improve maritime training. Studying and improving the key operator's functional state is an encouraging way to increase the reliability of complex human computer systems. For instance, some airline companies, seeing the advantages of PhyCS, are now monitoring pilot's physiological signals continuously, but only focus on simple features such as heart rate. One constraint is that there are inevitable obstacles in applying PhyCS in real-world environments compared to neuroscience or psychophysiology that study human cognitive states in strict controlled laboratory settings. Furthermore, the working environment of a ship's crew is even more complicated, leading to more obstacles. To introduce this promising new technology to the maritime domain, this paper proposed advantages of applying PhyCS in three aspects: increase the objectivity of human performance evaluation, improve usability tests, and develop adaptive training systems. Furthermore, according to our research experience and lessons learned from utilizing PhyCS in an engine room simulator environment, several obstacles were discussed: the requirement for ambulatory and wireless devices, the requirement of robust body movement artefact reduction, the modelling and verification, and the requirement of expertise knowledge. Additionally, suggestions to deal with these obstacles were given. For example, interpolation method is suggested to be superior to deletion in handling contaminated time series of heart beat intervals. We believe that this paper offers valuable information for researchers which will help them avoid common pitfalls in this domain and physiological computing systems will ultimately advance toward more sophisticated and nuanced techniques that will be even more effective in solving human factor issues.

## 6. ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI. Grant Number: 25420864

## 7. REFERENCES

1. AYAZ, H., SHEWOKIS, P. A., BUNCE, S., IZZETOGLU, K., WILLEMS, B., and ONARAL, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *Neuroimage*, 59(1), pp.36-47.
2. BAKER, C.C. and MCCAFFERTY, D.B. (2005). Accident database review of human element concerns: What do the results mean for classification? *In Proc. Int Conf. 'Human Factors in Ship Design and Operation*, RINA.
3. BALDWIN, C. L. and PENARANDA, B. N. (2012). Adaptive training using an artificial neural network and EEG metrics for within and

cross-task workload classification. *NeuroImage*, 59(1), pp.48-56.

4. BORGHINI, G., ASTOLFI, L., VECCHIATO, G., MATTIA, D., and BABILONI, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews*, 44, pp.58-75.
5. BROUWER, A. M., ZANDER, T. O., VAN ERP, J. B., KORTELING, J. E., and BRONKHORST, A. W. (2015). Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls. *Frontiers in neuroscience*, 9,p.136
6. CHENG, D. and VERTEGAAL, R. (2004). An eye for an eye: a performance evaluation comparison of the lc technologies and tobii eye trackers. *In Eye Tracking Research & Application: Proceedings of the 2004 symposium on Eye tracking research & applications*, 22, 24, pp.61.
7. DURANTIN, G., GAGNON, J. F., TREMBLAY, S., and DEHAIS, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural brain research*, 259, pp. 6-23.
8. FAIRCLOUGH, S.H. (2009). Fundamentals of physiological computing. *Interacting with computers*, 21(1), pp.133-145.
9. GAMELIN, F.X., BERTHOIN, S.E.R.G. and BOSQUET, L. (2006). Validity of the polar S810 heart rate monitor to measure RR intervals at rest. *Medicine and science in sports and exercise*, 38(5), p.887.
10. GOULD, K. S., RØED, B. K., SAUS, E. R., KOEFOED, V. F., BRIDGER, R. S., and MOEN, B. E. (2009). Effects of navigation method on workload and performance in simulated high-speed ship navigation. *Applied ergonomics*, 40(1), pp.103-114.
11. GRIFFITH, C.D. and MAHADEVAN, S. (2011). Inclusion of fatigue effects in human reliability analysis. *Reliability Engineering & System Safety*, 96(11), pp.1437-1447.
12. GWIN, J.T., GRAMANN, K., MAKEIG, S. and FERRIS, D.P. (2010). Removal of movement artifact from high-density EEG recorded during walking and running. *Journal of neurophysiology*, 103(6), pp.3526-3534.
13. HAJEK, W., GAPONOVA, I., FLEISCHER, K.H. and KREMS, J. (2013). Workload-adaptive cruise control–A new generation of advanced driver assistance systems. *Transportation research part F: traffic psychology and behaviour*, 20, pp.108-120.
14. KERICK, S.E., OIE, K.S. and MCDOWELL, K. (2009). Assessment of EEG signal quality in motion environments (No. ARL-TR-4866). *Army Research Lab, Aberdeen Proving Ground Md, Human Research and Engineering Directorate.*
15. KINGSLEY, M., LEWIS, M. J., and MARSON, R. E. (2005). Comparison of polar 810 s and an ambulatory ECG system for RR interval measurement during progressive exercise. *International journal of sports medicine*, 26(01), pp.39-44.
16. KLINE, J.E., HUANG, H.J., SNYDER, K.L. and FERRIS, D.P. (2015). Isolating gait-related movement artifacts in electroencephalography during human walking. *Journal of neural engineering*, 12(4), p.046022.
17. LEAN, Y. and SHAN, F. (2012). Brief review on physiological and biochemical evaluations of human mental workload. *Human Factors and Ergonomics in Manufacturing & Service Industries*, 22(3), pp.177-187.
18. LEE, Y. H. and LIU, B. S. (2003). Inflight workload assessment: Comparison of subjective and physiological measurements. *Aviation, space, and environmental medicine*, 74(10), pp. 1078-1084.
19. LÜTZHÖFT, M. (2004). "The technology is great when it works": Maritime Technology and Human Integration on the Ship's Bridge. (Ph.D. thesis), Linköping University, Linköping.
20. MEHTA, R.K. and PARASURAMAN, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Frontiers in human neuroscience*, 7, pp.889-889.
21. PAN, Y., KOMANDUR, S., and FINKEN, S. (2015). Complex Systems, Cooperative Work, and Usability. *Journal of Usability Studies*, 10(3), pp. 100-112.
22. PAPACHRISTOS, D., KOUTSABASIS, P. and NIKITAKOS, N. (2012). Usability evaluation at the ship's bridge: A multi-method approach. *In 4th International Symposium on Ship Operations, Management and Economics,* pp. 8-9.
23. PELTOLA, M. (2012). Role of editing of RR intervals in the analysis of heart rate variability. *Frontiers in Physiology*, 3, p.148.
24. PRINZEL, L.J., FREEMAN, F.G., SCERBO, M.W., MIKULKA, P.J. and POPE, A.T. (2003). Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 45(4), pp.601-614.
25. RAMIREZ, R. and VAMVAKOUSIS, Z. (2012). Detecting emotion from EEG signals using the emotive epoc device. *In Brain Informatics* (pp. 175-184). Springer Berlin Heidelberg.

 

26. SIEGENTHALER, E., COSTELA, F. M., MCCAMY, M. B., DI STASI, L. L., OTERO‐MILLAN, J., SONDEREGGER, A., and MARTINEZ, S. (2014). Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *European Journal of Neuroscience*, 39(2), pp. 287-294.
27. STUIVER, A., BROOKHUIS, K. A., DE WAARD, D., and MULDER, B. (2014). Short-term cardiovascular measures for driver support: increasing sensitivity for detecting changes in mental workload. *International Journal of Psychophysiology*, 92(1), pp.35-41.
28. TARVAINEN, M.P., NISKANEN, J.P., LIPPONEN, J.A., RANTA-AHO, P.O. and KARJALAINEN, P.A. (2014). Kubios HRV–heart rate variability analysis software. *Computer methods and programs in biomedicine*, 113(1), pp.210-220.
29. TOURYAN, J., LANCE, B.J., KERICK, S.E., RIES, A.J. and MCDOWELL, K. (2016). Common EEG features for behavioral estimation in disparate, real-world tasks. *Biological psychology*, 114, pp.93-107.
30. WANYAN, X., ZHUANG, D., and ZHANG, H. (2014). Improving pilot mental workload evaluation with combined measures. *Biomedical materials and engineering*, 24(6), pp.2283-2290.
31. WEI, J., CHEN, T., LIU, G. and YANG, J. (2016). Higher-order Multivariable Polynomial Regression to Estimate Human Affective States. *Scientific reports*, 6.
32. WICKENS, C., MCCARLEY J., and STEELMAN-ALLEN K. (2009). "NT-SEEV: A Model of Attention Capture and Noticing On the Flight Deck". *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 53.12, pp.769-773.
33. WILTSHIRE, T.J. and FIORE, S.M. (2014). Social Cognitive and Affective Neuroscience in Human–Machine Systems: A Roadmap for Improving Training, Human–Robot Interaction, and Team Performance. *Human-Machine Systems, IEEE Transactions on*, 44(6), pp.779-787.
34. WU, Y., MIWA, T., SHIMAMOTO, K. and UCHIDA, M. (2015). Development of quantitative team performance evaluation method for ERM. *WMU Journal of Maritime Affairs*, 14(2), pp.333-347.
35. YOUNG, M.S., BROOKHUIS, K.A., WICKENS, C.D. and HANCOCK, P.A. (2015). State of science: mental workload in ergonomics. *Ergonomics*, 58(1), pp.1-17.