Development of operator functional state evaluation methods in the maritime domain

Wu, Yanbin

(Degree)
博士（工学）
(Date of Degree)
2017-09-25
(Date of Publication)
2018-09-01
(Resource Type)
doctoral thesis
(Report Number)
甲第7015号
(URL)
https://hdl.handle.net/20.500.14094/D1007015

Doctoral Dissertation
博　士　論　文

Development of operator functional state evaluation methods in the maritime domain

（船舶運航管理実務者の機能評価手法の開発）

August 2017

Graduate School of Maritime Sciences
Kobe University

WU YANBIN
（毋　岩斌）

Summary

In human machine systems where the risk of accident is closely related to human component such as in maritime operation, it is crucial to maintain an optimal Operator Functional State (OFS). The complexity of these modern systems, the rapid update of contemporary technologies and reduced manning level all contribute to the high cognitive demands experienced by on-board crewmembers. The purpose of this research is to categorize the human factors that lead to suboptimal operator performance and to develop different methods to recognize those suboptimal factors. The evaluation methods of OFS fall into three general categories: subjective rating scales, physiological signal based objective measures, and performance based measures.

In the work environment of maritime domain, operators have to complete required tasks in an ambulatory situation. The necessity of pre-processing physiological signals is emphasized by the fact that many artefacts would decrease the signal quality, especially when the subject has many body movements. In chapter 2, a combination of recursive percentage filter and median filter is used to detect and replace outliers of RR interval series. In an attempt to eliminate artefact of single channel EEG, which is always contaminated across all relevant power bands, an accelerometer was directly attached to EEG electrodes to measure electrodes vibration. A linear model that based on calculating covariance and maximizing independence has been proved effective in reducing artefact of small amplitude across wide range of power bands (1-40 Hz).

Chapter 3 is an experiment study of using physiological features to evaluate operator's mental workload in conducting two kinds of tasks: standard reference task and engine-room simulator task. The difficulty of MEPS task is manipulated by varying the number of operation procedures and the type of pipeline involving in the operation. Six physiological features (alpha wave rate,

beta wave rate, theta wave rate, MHBI, sdHBI, LFHF), a subjective rating scale (NASA-TLX), and performance measures are used to correlate with task demand. The results shows that for n-back task and MEPS task, different features are sensitive to the task difficulty. A ceiling effect of using alpha wave rate to infer operator mental workload is found through the experiment study.

Human fatigue caused by either physical exertion or mental strain is one of the most significant factors that constrain operator's functional capability to fulfil specific tasks. To ensure working performance and to improve occupational health, Chapter 4 aims to develop a quantitative method to evaluate operator fatigue during conducting pipeline works. A Japanese version of RPE scale and heart inter-beat intervals are measured in an experiment study. Hurst exponent (HE) is extracted from detrended fluctuation analysis to define the fractal structures of RR interval time series. The average RR interval highly correlates with RPE scale. Results show that HE during working condition is significantly higher than during rest condition. In addition, a weak positive correlation between Hurst exponent and work performance, which is represented by the torque variance, is found in 5 among the 10 subjects.

In Chapter 5, a real world experiment that studies the mental workload of the first engineer of training ship Fukaemaru is described. One information flow model that divides operator's mental capacity into four channels: visual, auditory, cognitive, and psychomotor (VACP) is used to analyse subject's behavioural information. The weight of each channel is assigned with an orderly scale according to the mental workload exposed to the subject. Individualized combination of physiological features are decided based on the clustering quality in n-back task, which can be quantitatively evaluated from an I-index. Mental workload is estimated by Euclidean distance based classification method and Mahalanobis distance, respectively.

Chapter 6 summarizes the findings of this paper and makes prospective research plans.

Table of contents

List of Tables

List of Figures

ACRONYMS

ANOVA	Analysis of Variance
BNWAS	Bridge Navigation Watch Alarm System
BRM	Bridge Resource Management
DFA	Detrended Fluctuation Analysis
DFFT	Discrete Fast Fourier Transformation
ERM	Engine-room Resource Management
EEG	Electroencephalography
FFT	Fast Fourier Transformation
HBI	Heart Beat Interval
HE	Hurst Exponent
HR	Heart Rate
HRV	Heart Rate Variability
IMO	International Maritime Organization
K-NN	K Nearest Neighbour
LF/HF	Low Frequency / High Frequency ratio
MHBI	mean of Heart Beat Interval
MEPS	Marine Engine Plant Simulator
MWL	Mental Workload
NASA-TLX	National Aeronautics and Space Administration Task Load Index
OFS	Operator Functional State
RPE	Borg’s Rating of Perceived Exertion
RRI	R-R wave (inter-beat) Interval
SD	Standard Deviation
sdHBI	Standard Deviation of Heart Beat Interval
SOLAS	Convention on the Safety of Life at Sea
STCW	International Convention on Standards of Training, Certification and Watch-keeping for Seafarers
VACP	Visual Auditory Cognitive Psychomotor Model

Chapter 1 Research background

1.1 Background

With current sophisticated shipping management and technology development, maritime industry has become an industry with a relatively good safety record compared to that of the last century. Catastrophic disasters such as Titanic (1912, UK) and Toya Maru (1954, Japan) are now part of history. Despite the advances, maritime accidents that lead to injuries or death, environmental pollution and financial losses continue to occur. The American Bureau of Shipping (ABS) analysed the maritime accident reports of the US, UK, Canada and Australia, and concluded that approximately 80 to 85% of all reported accidents involved human error. Of those, about 50% were initially caused by human error. Another 30% of accidents were partially associated with human error [1]. Nevertheless, maritime industry has characteristics that make human factor issues more complicated. Firstly, the safe transportation of cargo and people overseas require the fulfilment of various on-board tasks ranging from navigation, propulsion, cargo handling, to platform maintenance. Secondly, the work environment is characterized by lack of contact with family and friends, by a mix of cultures living and working together, and for the most part by a high degree of monotony [2]. Thirdly, human-computer interaction issues arise continuously, yet long remain unsolved. For instance, in order to avoid the drowsiness of officers on watch, the international convention of Safety of Life at Sea (SOLAS) currently requires all ships above 150 tons 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 letting their movements be detected within a pre-set time interval. However, the intrusiveness and usability of BNWAS are problematic from the view of ergonomics.

In safety-critical systems such as piloting ships in busy ports, we expect that the operator can persistently monitor the system parameters (e.g. ship and wind speed) and adapt to unforeseen changes (e.g. fishing vessels) and avoid any incidents (e.g. collision), in other words, to perform their job perfectly. Hardware reliability has improved to such an extent that abnormal psychophysiological states of the operator are more likely to lead to system performance degradation [3]. It is essential to address ways to maintain optimal operator functional state (OFS) in the context of human-machine systems, where the risk of accident is negatively related to the human component [4].

To prevent marine accidents and to minimize their consequences, IMO came up with a series of conventions that are widely adopted around the world, including MARPOL, SOLAS, and STCW and so on. The IMO's International Convention on Standards of Training, Certification and Watch keeping for Seafarers (STCW), 1978 was the first internationally agreed Convention to address the issue of minimum standards of competence for seafarers. In 1995, the STCW Convention was completely revised and updated to clarify the standards of competence required and provide effective mechanisms for enforcement of its provisions. After that, a comprehensive review of the STCW Convention and the STCW Code commenced in January 2006, and culminated in a Conference of Parties to the STCW Convention that was held in Manila, Philippines from 21 to 25 June 2010, that adopted a significant number of amendments to the STCW Convention. One of the most important changes is that the requirements concerning Bridge Resource Management and Engine-room Resource Management (BRM/ERM) have been introduced into the competence tables as mandatory requirements for navigation and engineer officers. ERM is a widely used approach to achieve ship's safe navigation by effectively managing the resources of personnel, equipment and information in the machinery space. Engine-room resources are personnel resources, equipment resources, consumables, information resources and environmental resources [5].

The core element of ERM and BRM is to appropriately allocate resources, among which personnel resource is the most important, considering the priority of tasks. While the condition of machinery and information resources are relatively straightforward, the condition of every crewmember is implicitly difficult to quantify. From a long-term prospective, the condition of crewmembers is related to non-technical skills, such as communication and leadership skills. The condition of crewmembers in a short-term period, especially when carrying out key operations, is crucial for system safety. If we are aware of a crewmember's physiological and psychological conditions, the allocation of personnel resources can be conducted more reasonably and effectively.

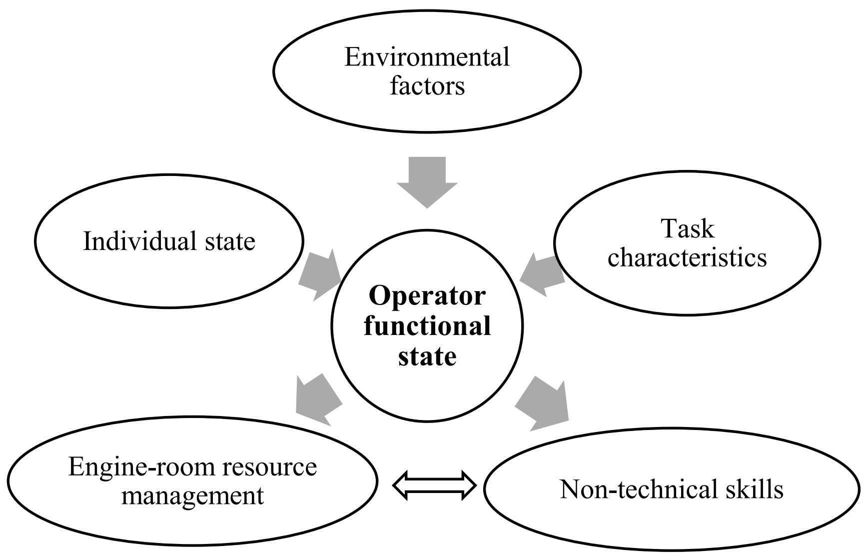


Figure 1-1 Risk factors that affect operator functional state

This paper is structured as follows. Chapter 1 introduces the research background, the risk factors and evaluation methods of OFS, and literature review is conducted. Section 1.4 proposed the potential applications of OFS online measuring technology in the maritime domain. To eliminate automatically the effect of artefact on signal qualities, Chapter 2 studies the characteristics of physiological signal artefacts and developed pre-processing algorithms. Chapter 3 is an experiment study of using physiological features to evaluate operator's mental workload in conducting two kinds of tasks: standard reference task and engine-room simulator task. By using detrended fluctuation analysis, Chapter 4 experimentally studies the heart rate based evaluation of operator's fatigue. In Chapter 5, a real world experiment that studies the mental workload of the first engineer of training ship Fukaemaru is described. One information flow model that divides operator's mental capacity into four channels: visual, auditory, cognitive, and psychomotor (VACP) is used as reference information. Mental workload is estimated by Euclidean distance based method and Mahalanobis distance, respectively. Chapter 6 summarizes this paper and makes prospective research plans.

1.2 Operator functional states

Operator functional states (OFS) refer to the physical, mental, and psychophysiological conditions that may mediate an operator's performance in fulfilling specific tasks. OFS should be regarded as the result of many physiological and psychological processes that regulate brain and body in an attempt to maintain an individual in an optimal condition to meet the demands of the work environment [6].

Different operator's capacity to fulfil some tasks are distinct and can be affected by many risk factors. In 2004, Research and Technology Organization of North Atlantic Treaty Organization (RTO-NATO) conducted a systematic investigation on the multidimensional OFS that can cause performance degradation of military personnel and the methods to detect these factors [7]. Both risk factors and evaluation methods were roughly divided into three categories. In consideration of the working environment of on-board crews, motion sickness is added to risk factors under the category 'individual state' and the time of continuous on-board service is added under the ‘environmental’ risk factors. These risk factors and evaluation methods are summarized in Table 1-1.

Table 1-1 Risk factors and evaluation methods of OFS

Risk factors	
Environmental	Hyperbaric environments, Noise and Vibration, Pharmacological Mediators (Drugs and Medicines), Sustained Acceleration, Thermal Stress, continuous onboard service time.
Individual state	Circadian rhythms, hydration, illness, mental fatigue, sleep loss, motion sickness
Task characteristics	Cognitive load, physical load, situation awareness
Evaluation methods	
Physiological Measures	Actigraphy, cardiorespiratory measures, core temperature, electroencephalography(EEG), electrodermal activity, electromyography(EMG), Eye activity, functional magnetic resonance imaging (fMRIs), near-infrared spectroscopy (fNIRs), oximetry, stress hormones.
Performance tests	Response time, accuracy in display monitoring, memory recall, multi-tasking to evaluate attention and time-sharing resources
Subjective measures	NASA task load index (NASA-TLX), Brooks-Samn Perelli’s Fatigue scale, Profile of mood states, Sleepiness scale, sleep diaries.

In [7], RTO-NATO defines three concepts of OFS to evaluate the OFS variations: background state, baseline state, and operational state. The background state represents the averaged, unloaded (resting) state of the operator, without any responsibilities and goals. An operator’s value of evaluation features of this background state should be advanced measured in order to reflect meaningful changes in the other two states. Although it is expected that some aspects of the personality profile may exhibit small changes from day-to-day, in general the background state would be expected to be stable. Christensen et al. [8] used EEG data obtained from asymptotically trained subjects in performing a complex multitask across five days in one month to classify OFS and their results demonstrate that with proper methods, pattern classification is stable enough across days and weeks to be a valid, useful approach. The operational baseline state is defined as the local non-stressed state of the operator prior to being actively engaged in a task. While clearly related to the background state, baselines may be above or below background levels (background is, theoretically, the average of all baseline states) and are naturally influenced by prior work, temporary individual state factors, and ambient environments [7]. However, if the purpose is to classify OFS into high, normal, low level, this paper suggests the use of calibration tasks that elicit corresponding OFS and use the features as the operational baseline state. The operational state represents the functional state of the operator that to be evaluated while engaged in a task under specific operational conditions and the operational state is critical to the successful implement of operation duties.

1.3 Evaluation methods

There are mainly three types of OFS measures based on the following techniques, subjective self-report, primary/secondary task performance measures, and physiological metrics. The performance measurements are sometimes implemented together with task analysis. In maritime operation, the operators usually devote most of their attention to monitoring automation systems and information processing rather than on making instantaneous behavioral responses. There are almost no performance indices, such as accuracy or reaction time, for complex tasks of maritime operations. Performance indices are therefore unavailable. The largest obstacle in subjective self-reporting as an OFS measure is its dependence on the operator and their time and ability to record their feelings. On the contrary, the importance of continuous online monitoring of OFS as highlighted by the practical applications, physiological metrics offer higher feasibility for measurement of OFS.

Lean and Shan (2012) [9] 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.

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 unobtrusive. 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 [10]. 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.

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 [11]. 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) [12] 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 [13].

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) [14] found that task difficulty in mental arithmetic affects micro saccadic rates and magnitudes. Wanyan et al. (2014) [15] found that pupil diameter and blink interval are effective to infer human MWL.

In literature, flight simulator and driving simulator are frequently used to simulate relevant operation environment in these research and different physiological metrics are reported valid and sensitive to infer operators' MWL. The sensitivity of those metrics are tested either by the significant result of Analysis of Variance (ANOVA) or regression models in distinctive level of task difficulty. HRV features and cerebral cortex activity measured by fNIRs or EEG are widely used because the equipment is relatively low-cost and requires no medical expertise.

Compare to that of civil aviation, MWL evaluation research in merchant shipping seems much inactive (Young et al. 2014 Table 2) [16]. One of the reasons is that there are considerable obstacles for applying MWL research in shipping. The obstacles include: a) more complex working environment in ship engine room; b) require ambulatory physiological sensors and noise reduction algorithms; c) complexity of maritime operation tasks. For instance, the operators generally have to move around engine room and engine control room in engine system operations, and these body movements will cause physiological data contamination.

Table 1-2 Summary of literature

Study	**Subjective measure**	**Tasks**	**Difficulty manipulation**	**Effective subjects**	**Statistical significance (Cardiac)**	**Statistical significance (Ocualr)**	**Statistical significance (Brain wave)**
G. Durantin (2014)[17]	NASA TLX	Computer simulated air piloting; N- back sub task and auditory alarm task	With/without Crosswind during piloting; 1/2-back spatial task; (2*2) 4 sessions	12; ANOVAs		Inverted u-shape of oxygenation	fNIR Piloting difficulty/ oxygenation(6) F(1,11)=5.82 Interaction/Oxygenation(3) F(1,11)=5.11 Oxygenation(3)/Performance R^2=0.52
Arjan Stuiver (2014)[18]	RSME (Zijlstra)	6*1.5h Driving simulator	Fog/no fog Low/high traffic density	15; Repeated MANOVA 15*6=90	Interaction/high Frequency spectral HRV F(1,89)=4.98	Ceiling effects of high load	interaction/blood pressure F(1,89)=13.42
Yufeng Ke (2014)* [19]	NASA-TLX	9(4+4+1)* 3.5 min Verbal, spatial n-back tasks and MATB	N-back Multi attribute task battery	17; Pair t test; Within or cross tasks			EEG Within task p<0.01; Cross-tasks

Table 1-2 Continued in next Page

Table 1-2 continued

Hasan Ayaz (2012)[20]	NASA-TLX	N-back(28 sessions) Air traffic control by simulator (6 sessions)	n-back voice/data based comm number of aircrafts(6, 12,18)	24 ANOVA Subjects as factor			fNIR “n”/oxygenation(2) F(3,69)=4.37, p<0.05; Vehicle number/oxygenation(8) F(2,42)=4.52
K. Ryu (2005)[21]	NASA-TLX	Tracking and mental arithmetic dual tasks	Target speed (3 level) Number digit(2/3, 2 level) 3*2 level	10; Combined measures. subjects as factor.	HRV, /target speed, F(2,18)=5.02, p<0.05; /number digit, null; interaction null.	Blink interval /target speed, F(2,18)=7.64, p<0.01; /number digit, null; interaction, null	EEG, alpha suppression/target speed, null; /number digit, F(1,9)=5.87,p<0.05; Interaction, null
Wanyan, X. et al (2014)[15]	NASA-TLX	Three phases of flight simulation, monitor flight indicator	(9,6,3,0) indicators to be monitored. Four difficulties	12; One-way repeated ANOVA	Main effect HR/MW, p=0.252. RRCV/MW, p=0.019	Main effect Pupil diameter p=0.076	ERP P3 peak amplitude p=0.049
Yung-hui LEE (2003)[22]	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} p<0.0001		

1.4 Potential applications

1.4.1 OFS adaptive automation

The automation methods currently used on board still mainly deal with elementary stages of human cognition, such as information generating and acquisition. Further, there are well-documented incidences of operator in effectiveness, often attributed to “clumsy automation,” because humans have been left with only tasks, which are too difficult or too expensive to automate [23]. It may be possible to involve computers with higher levels of automation such as system decision-making to reduce those problems. Adaptive automation or adaptive systems refer to the idea of an automated system that can adapt to a changing environment [24]. Dorneich et al. [25] (2012) define adaptive systems as those ”allowing the system to invoke varying levels of automation support in real time during task execution, often on the basis of its assessment of the current context…only as needed”. OFS-adaptive automation (OFS-AA) is the adaptive automation system based on the ‘current context’ of operator functional state and some analogous proposals have appeared in the automobile industry, such as BMW’s new generation of driver assistance systems: workload-adaptive cruise control [26]. The biocybernetic loop proposed by Pope et al. (1995) [27] is considered to be seminal research in OFS-AA. The biocybernetic loop [27] is formed by adjusting the mode of operation of a task set (e.g., level of automation) based on the brain activity criterion (EEG-based index of engagement) reflecting an operator’s engagement in the task set.

Fig. 1-2 graphically depicts the conceptual model of OFS-AA, composed of two main parts, physiological computing and system adaption. Physiological computing is the determining

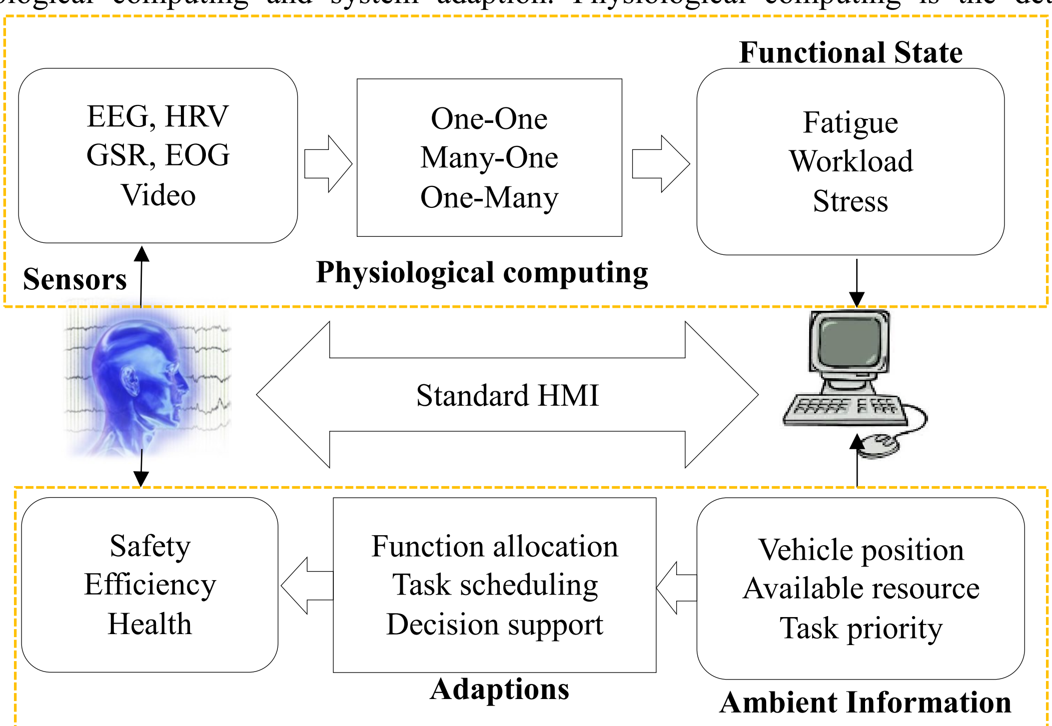


Figure 1-2 Model of operator functional state based adaptive automation system

foundation of adaption strategies and performance-oriented adaption strategies provide feedback to improve the robustness of the former. Physiological computing correlates an operator's physiological changes and their functional states based on simple selected features or complex algorithms such as support vector machine [4] or fuzzy modeling [23]. The physiological metrics mainly include heart rate variability (HRV) recorded from electrocardiogram (ECG), blood pressure, respiration, eye blinks and pupil diameter, skin potential, hemodynamic indices and cerebral cortex indices such as electroencephalogram (EEG) and event-related potentials (ERPs) [9]. The mapping models of these physiological indices and OFS fall into four possible categories: one-to-one (i.e. a physiological variable has a unique isomorphic relation with a psychological state), many-to-one, one-to-many and many-to-many [28].

According to specific OFS and for certain applications, the computer can allocate one or several appropriate adaption strategies, including function allocation, task rescheduling and decision support, to maintain high level performance of the human-computer system. Feigh et al. (2012) [29] defined four main mechanisms of adaption and explained the trigger criterion and possible applications of those strategies. Nevertheless, adaption that may be invoked to adjust users' OFS can be either explicit or implicit. Fairclough (2009) [30] reported the strengths and weaknesses of both approaches, to which the design of explicit and implicit system adaptions for physiological computing system must cater. The conspicuity of explicit adaption at the interface is possibly the more potent technique to directly influence the psychological state of the user. However, it also increases the possibility of 'false alarm'. Contrarily, implicit adaptions represent subtle changes at the interface that may be used frequently without creating the potential for false alarms [30]. Therefore, to ensure the efficacy, as well as the users' trust of system adaptions through a reduction of false alarm, the designers should consider the invoking of explicit adaption, and only opt for them if the implicit ones fail to impact on users' OFS.

Among the cognitive human factor constructs that represent OFS, MWL is foremost. First, Bindewald et al. (2014) [24] proposed that if function allocation between human and machine within an OFS-AA was considered a multi-objective optimization, designers would optimize a combination of performance, safety, and robustness as a function of the task allocated to each component. The limitations of the system and human capability shape this optimization, where a significant component of human capability is quantified in terms of human workload. This claim is similar to the workload restriction of multiple resource theory proposed by Wickens (2008) [31]. Further, despite the apparent effect that exorbitant transitional MWL will result in stress, an accumulation of lower workload or task disengagement will contribute to drowsiness. Therefore, MWL is a causal factor of several different OFS. Hence, development of sensible and diagnostic measurement of MWL are crucial for the realization of OFS-AA.

1.4.2 Usability testing of human-cantered design

As design companies see the benefits of adopting human-oriented design methods instead of technology-oriented ones, they pay more attention on testing the usability of their products. A common goal, for experts in HCI who conduct research on designing and evaluating user interfaces is to design an interface transparently [32], allowing the user to focus their MWL on the underlying task rather than on understanding and interpreting the user interface [33]. Correspondingly, human-oriented systems have also emerged in modern ship design. Figure1-3 shows multiple, distributed interaction interface with several interacting screens on a ship bridge. For a ship, favourable usability means that operators can accomplish required tasks using the limited on-board resources with efficiency, effectiveness and self-satisfaction. In the development of new innovative maritime systems, usability experts need to avoid creating distributed interfaces that can become technology "barriers". The object is to support the operator and make technology as user friendly as possible.

Figure 1-3 Modern ship's bridge, 18 screens in an operation room

(adapted from Pan et al. 2015 [34])

Usability testing of maritime products should be undertaken not only by designers but also by the operators (ship crew) who would use it. This is especially true for scenario-based products for situations of high uncertainty, such as collision avoidance systems. Robert et al. (2003) [35] using a ship control simulator found that higher levels of collision threat and impaired performance on the secondary oil pressure monitoring were associated with markedly increased ratings of MWL. Their work has relevance for the design of collision avoidance systems [35]. Gould et al. (2009) [36] 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. They reported that ECDIS navigation significantly improved course-keeping performance. They also used heart rate variability (HRV) and skin conductance as MWL measurements and the results indicated higher workload in conventional navigation, although the difference between the groups was not significant. In the engine room of a ship, the on watch engineer is required to simultaneously monitor and control a few complex sub-systems within one or several interaction interfaces, i.e. central cooling system and lubrication oil system. With the increase of human-cantered control systems, it should be a necessary requirement to conduct usability testing concerning MWL before adoption.

1.4.3 Maritime training

In professional education such as in the training of pilots, advanced seafarers and surgery operators, high fidelity simulators are widely used to help the 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 real 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. Therefore, although many companies focus on developing the fidelity of simulators, high fidelity does not necessarily bring high effectiveness of training. Training procedures and instructions should be carried out to pace and train the trainees to maintain engagement with the operational process, and keep their MWL at the optimal point, where the learner is neither overloaded nor under loaded. Wiltshire and Fiore (2014) [37] argued for the advantages of training where the trainee's social and affective cognition state can be handled timely. To improve the training regime, training institutions need to adjust their syllabus and pedagogics based on individual mental capacity and students' engagement throughout the learning process. A record of psychophysiological data such as brain activity and pupil movement can be used for a mental workload model, which can provide more reliable and objective data in addition to the usual subjective or experts' ratings. It may also improve the following training related issues:

- evaluate the effectiveness of training regimes
- compare differences and similarities of simulator operation and real operation
- elaborate on responses of experienced operators and novices
- choose optimized pedagogics for individualized teaching

References

[1] Baker CC, Seah AK. 2004. Maritime accidents and human performance: the statistical trail. In MarTech Conference, Singapore.

[2] Grech M, Horberry T, Koester T. 2008. Human factors in the maritime domain. CRC Press.

[3] Byrne, Evan A., and Raja Parasuraman. "Psychophysiology and adaptive automation." Biological psychology 42.3 (1996): 249-268.

[4] Yin, Zhong, and Jianhua Zhang. "Identification of temporal variations in mental workload using locally-linear-embedding-based EEG feature reduction and support-vector-machine-based clustering and classification techniques." Computer methods and programs in biomedicine 115.3 (2014): 119-134.

[5] Wu, Yanbin, Takashi Miwa, Kaoru Shimamoto, and Makoto Uchida. "Development of quantitative team performance evaluation method for ERM." WMU Journal of Maritime Affairs 14, no. 2 (2015): 333-347.

[6] Gaillard, A.W.K., & Kramer, A.F. (2000). Theoretical and Methodological Issues in Psychophysiological Research. In R.W. Backs & W. Boucsein (Eds.), Engineering Psychophysiology: Issues and Applications (pp. 31-58). Mahwah, New Jersey: Lawrence Erlbaum Associates.

[7] Operator Functional State Assessment (L'evaluation de l'aptitude operationnelle de l'operateur humain), NATO RESEARCH AND TECHNOLOGY ORGANIZATION NEUILLY-SUR-SEINE (FRANCE), ADA422195 , Feb. 2004

[8] Christensen, J. C., Estepp, J. R., Wilson, G. F., & Russell, C. A. (2012). The effects of day-to-day variability of physiological data on operator functional state classification. NeuroImage, 59(1), 57-63.

[9] 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.

[10] Tarvainen, Mika P., Juha-Pekka Niskanen, Jukka A. Lipponen, Perttu O. Ranta-Aho, and Pasi A. Karjalainen. "Kubios HRV–heart rate variability analysis software." Computer methods and programs in biomedicine 113, no. 1 (2014): 210-220.

[11] Kerick SE, Oie KS, McDowell K. 2009. Assessment of EEG signal quality in motion environments. ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH AND ENGINEERING DIRECTORATE. (No. ARL-TR-4866).

[12] Prinzel III, L. J., Freeman, F. G., Scerbo, M. W., Mikulka, P. J., & Pope, A. T. (2003). Effects of a psychophysiological system for adaptive automation on performance, workload, and the event-related potential P300 component. Human factors, 45(4), 601-614.

[13] 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.

[14] Siegenthaler, E., Costela, F. M., McCamy, M. B., Di Stasi, L. L., Otero‐Millan, J., Sonderegger, A., ... & Martinez‐Conde, S. (2014). Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. European Journal of Neuroscience, 39(2), 287-294.

[15] Wanyan, X., Zhuang, D., & Zhang, H. (2014). Improving pilot mental workload evaluation with combined measures. Bio-medical materials and engineering, 24(6), 2283-2290.

[16] Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. Ergonomics, 58(1), 1-17.

[17] Durantin, G., Gagnon, J. F., Tremblay, S., & Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. Behavioural brain research, 259, 16-23.

[18] Stuiver, A., Brookhuis, K. A., de Waard, D., & Mulder, B. (2014). Short-term cardiovascular measures for driver support: increasing sensitivity for detecting changes in mental workload. International Journal of Psychophysiology, 92(1), 35-41.

[19] Ke, Y., Qi, H., He, F., Liu, S., Zhao, X., Zhou, P., ... & Ming, D. (2014). An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task. Frontiers in human neuroscience, 8.

[20] Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. Neuroimage, 59(1), 36-47.

[21] Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. International Journal of Industrial Ergonomics, 35(11), 991-1009.

[22] Lee, Y. H., & Liu, B. S. (2003). Inflight workload assessment: Comparison of subjective and physiological measurements. Aviation, space, and environmental medicine, 74(10), 1078-1084.

[23] Ting, C. H., Mahfouf, M., Nassef, A., Linkens, D. A., Panoutsos, G., Nickel, P., ... & Hockey, G. R. J. (2010). Real-time adaptive automation system based on identification of operator functional state in simulated process control operations. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 40(2), 251-262.

[24] Bindewald, J. M., Miller, M. E., & Peterson, G. L. (2014). A function-to-task process model for adaptive automation system design. International Journal of Human-Computer Studies, 72(12), 822-834.

[25] Dorneich, M. C., Ververs, P. M., Mathan, S., Whitlow, S., & Hayes, C. C. (2012). Considering etiquette in the design of an adaptive system. Journal of Cognitive Engineering and Decision Making, 6(2), 243-265.

[26] Hajek, W., Gaponova, I., Fleischer, K. H., & Krems, J. (2013). Workload-adaptive cruise control–A new generation of advanced driver assistance systems. Transportation research part F: traffic psychology and behaviour, 20, 108-120.

[27] Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. Biological psychology, 40(1), 187-195.

[28] Cacioppo, J. T., Tassinary, L. G., & Berntson, G. (Eds.). (2007). Handbook of psychophysiology. Cambridge University Press.

[29] Feigh, K. M., Dorneich, M. C., & Hayes, C. C. (2012). Toward a characterization of adaptive systems: A framework for researchers and system designers. Human Factors, 54(6), 1008-1024.

[30] Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. International Journal of Psychophysiology, 56(2), 171-184.

[31] Wickens, C. D. (2008). Multiple resources and mental workload. Human Factors: The Journal of the Human Factors and Ergonomics Society, 50(3), 449-455.

[32] Shneiderman, B. (2010). Designing the user interface: strategies for effective human-computer interaction. Pearson Education India.

[33] Hirshfield, L. M., Solovey, E. T., Girouard, A., Kebinger, J., Jacob, R. J., Sassaroli, A., & Fantini, S. (2009, April). Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near infrared spectroscopy. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2185-2194). ACM.

[34] Pan, Y., Komandur, S., & Finken, S. (2015). Complex systems, cooperative work, and usability. Journal of Usability Studies, 10(3), 100-112.

[35] Robert, G., Hockey, J., Healey, A., Crawshaw, M., Wastell, D. G., & Sauer, J. (2003). Cognitive demands of collision avoidance in simulated ship control. Human Factors: The Journal of the Human Factors and Ergonomics Society, 45(2), 252-265.

[36] Gould, K. S., Røed, B. K., Saus, E. R., Koefoed, V. F., Bridger, R. S., & Moen, B. E. (2009). Effects of navigation method on workload and performance in simulated high-speed ship navigation. Applied ergonomics, 40(1), 103-114.

[37] 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.

Chapter 2 Physiological features and the Pre-processing of physiological signals

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 in engine control room, and their work includes a number of physical tasks. Therefore, wearable devices that reliably collect and wirelessly transform physiological data are necessary for applications in a ship environment. Kerick et al. (2009) [1] argued that some neurophysiological measurement technologies, e.g. fMRI, MEG and PET could be generally ruled out due to their machinery size. 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) [2] argued for the necessity of editing raw HBI data, and appropriate artefact correction methods must be chosen according to different study settings.

A chest strap heart rate sensor RS800CX (Polar Electronics) was used to continuously measure the heart rate and a portable EEG device was used to measure brain electric voltage fluctuations caused by neural activities. EEG was measured by an ambulatory device (Digital Medic Inc.) consisted of a headset (17*16*1.5 cm 80g) and main part (5.8*12*2.4 cm 95g) with a sampling frequency of 128Hz. The headset is comprised of two electrodes and an accelerometer attached to the electrodes. The main part wirelessly connects to an online monitoring computer by Direct Sequence Spread Spectrum (DSSS). Effective wireless transmission distance were 10 meters. The detected brain area was posterior parietal cortex (P3-O1, P4-O2 of standard EEG channel nomenclature). Digital Medic Inc. (2013) [3] compared the measurement result in 128Hz between this EEG device and a medical one, the average correlation coefficient was 0.94. In the experiment, the subject can wear these two sensors at the same time as shown in Fig. 2-1.

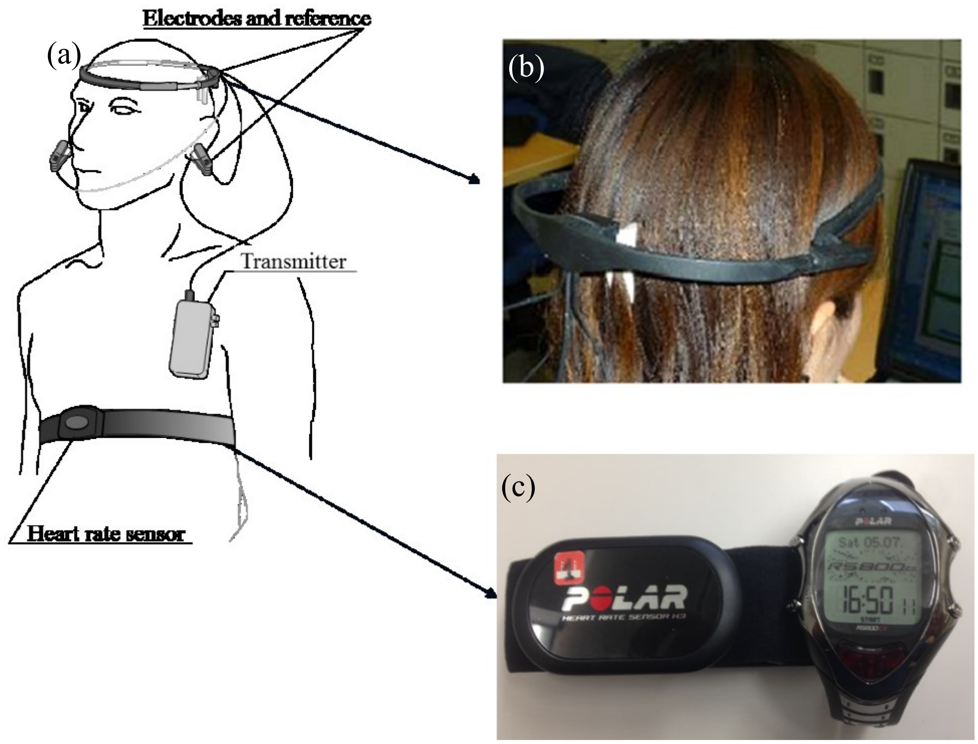


Figure 2-1 (a) sketch of subject wearing EEG device and hear rate sensor; (b) Muse EEG device; (c) Polar chest strap heart rate monitor (RS800CX)

2.1 Heart Rate Monitoring and Signal Pre-processing

RR interval series in units of milliseconds were extracted and analysed post-experiment. In a practical environment, the RR interval series is often contaminated by artefacts caused by poor skin-electrode contact, body motions, and/or sweating. The necessity of pre-processing of the RR interval series has been widely agreed upon since the loss of even single ectopic RR interval data point can have a serious impact on the interpretation of the results, especially for short-term analysis [2]. In addition, the deletion of ectopic beats can cause information loss and error of spectrum features. Therefore, automatic recognition and replacement of outliers must be conducted before extracting features from RR interval data. In this study, we utilized a pre-processing method that combined recursive percentage filter and median filter, which was similar to that proposed by Mishra and Swati [4]. As shown in equation (2-1), consider an RR interval series (n) , the first step is to recognize any data point that is more than t1 larger or smaller than the last sample, then replace it using a moving average window. $abs\{\cdot\}$ is absolute value operator and w_m is length of the moving window.

$$if\ \frac{abs\{x(n)-x(n-1)\}}{x(n-1)} > t_1\ ,\ \hat{x}(n) = mean\left\{x(n+m): abs\{m\} \leq \frac{w_m-1}{2}\right\}. \qquad (2\text{-}1)$$

The second step is to segment the original data into 5-minute samples and a median based pulse rejection filter is applied [5]. The recognition of outlier is based on equation (2-2), where med $[\cdot]$ is the median operator and x_m is the median value of the segmented signal (n) . The recognized outlier is then replaced by the median value of the moving window as shown in equation (2-3) and (2-4).

$$D(n) = \frac{abs\{x(n) - x_m\}}{1.483med[abs\{x(n) - x_m\}]} \qquad (2\text{-}2)$$

$$\hat{x}(n) = \begin{cases} x(n) & if\ D(n) \leq t_2 \\ x_i(n) & if D(n) > t_2 \end{cases} \qquad (2\text{-}3)$$

$$x_i(n) = med\left\{x(n+m): abs\{m\} \leq \frac{w_m - 1}{2}\right\} \qquad (2\text{-}4)$$

The threshold value t_1 was set as 30%, t_2 was set as four and window length w_m was five. According to the above method, one typical example of RR interval series preprocessing is shown in Figure 2-2. The total length of measured signal is 3445, and 9 points are edited by percentage filter, after 23 points are edited by median filter.

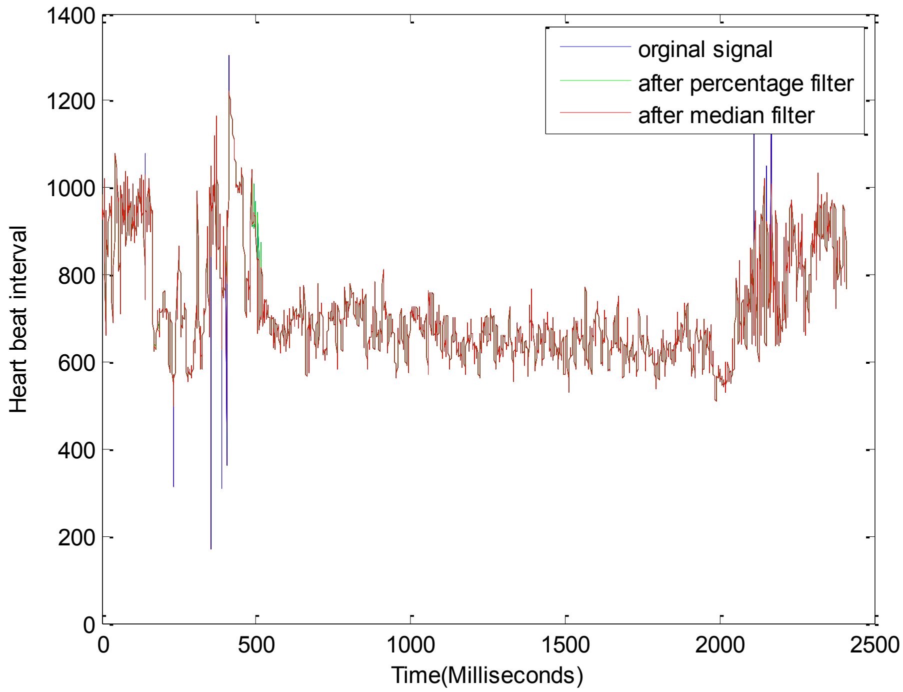


Figure 2-2 one example of RR interval time series pre-processing

2.2 EEG artefact reduction

An EEG device generally uses silver electrodes to measure the extremely weak signal (μv) of voltage fluctuations along the scalp, and the signal is usually 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 [6]. 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 [7,8]. 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 (Digital 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 2-3 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. Since different body movement can cause distinct EEG signal contamination in a wide range of power bands, it is not possible to use specific filters to effectively reduce artefact and remain the EEG signal unaffected. Figure 2-3 shows 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.

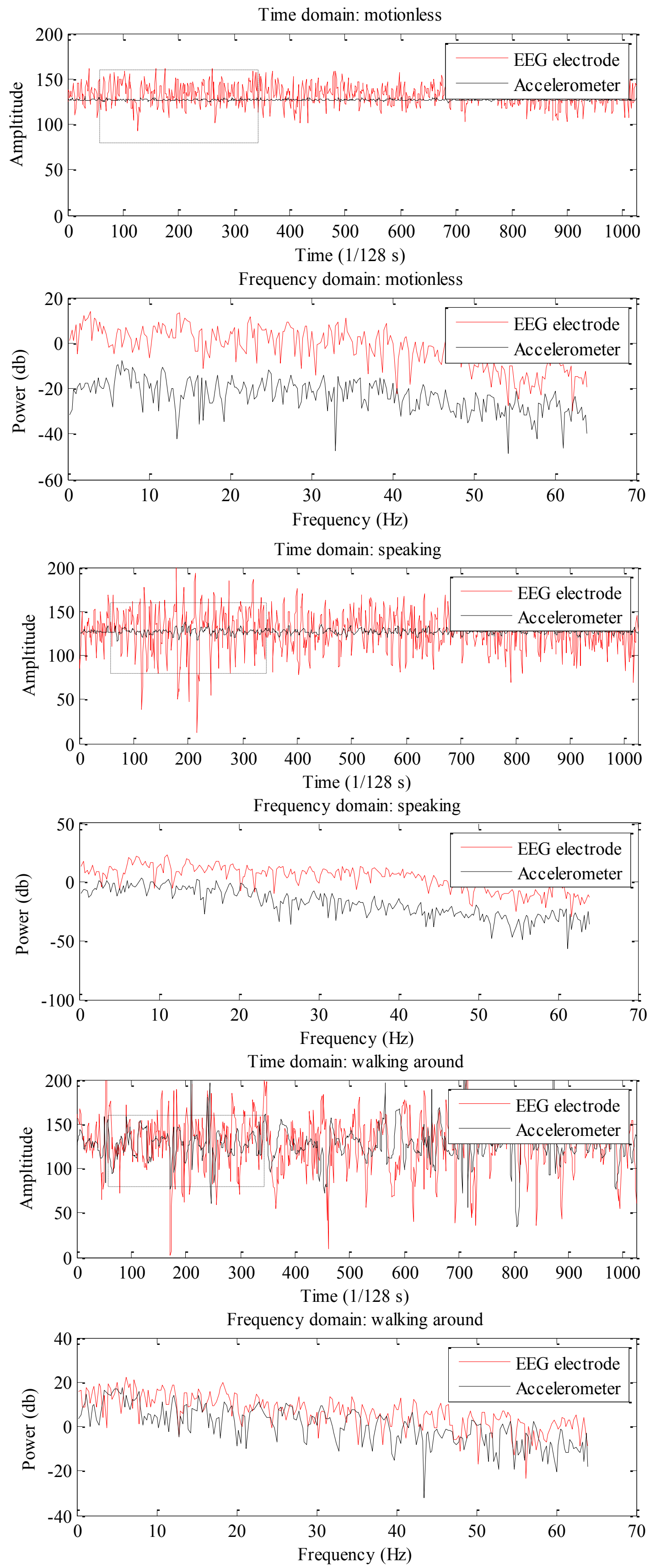


Figure 2-3 Time domain and frequency domain analysis of EEG in three situations

To eliminate the effect of body movement artefact on EEG signal, the following linear method is proposed. If it is hypothesized that the signal measured by EEG electrode *E* (channel 1) is the linear sum of clean EEG signal *S* and artefact caused by electrode movement measured accelerometer *V* (channel 2), we can have the following linear model with two unknown constants:

$$E = S + k \cdot V + b \tag{2-5}$$

Calculate the covariance between *E* and *V,*

$$cov(E,V) = cov(S,V) + k \cdot cov(\mathrm{V},\mathrm{V}) + cov(b,V) \tag{2-6}$$

As S and V are from different resource and if we maximize the independence between S and V, we can have

$$k = cov\left(E,V\right) / D\left(V\right) \tag{2-7}$$

$$S + b = E - k \cdot V \tag{2-8}$$

In actual situation, both head movement and verbal communication happen occasionally, resulting intermittent contaminated signal. Therefore, the solution of k and b can be different when signals are contaminated by different source. Epochs are recognized as contaminated by body movement artefact when the power of channel 2 exceeds a pre-set threshold value. The standard deviation of channel 2 signal is calculated with a step of 0.25 seconds to detect contaminated epochs and the corresponding epochs of channel 1 is processed by the linear model.

In an experiment, a subject was required to sit still and read a same page of a book in two conditions: read silently and read loudly. In read loudly situation, the subject occasionally shake his head. Figure 2-4 shows the power spectrum of a 2-second segment of reading silently condition. The Red line is the power spectrum of original signal measured by channel 1 and black line is measured result of channel 2. Channel 1 falls between 0-15 dB in the power bands of 0-40Hz, which is found related to different physiological or cognitive activities, and becomes smaller than 0 dB in >40Hz band. Channel 2 measured by accelerometer is smaller than -10 dB and rarely overlaps with that of channel 1. In this condition, we consider the signal measured by EEG electrode are clean EEG signal and it is not contaminated by vibration artefact measured by accelerometer. Besides, the peak of power in 60Hz is artefact caused by the alternating electricity line, the frequency of which in Kansei area Japan is 60Hz. Since we only extract EEG physiological features from power bands of <40Hz, this artefact is not processed.

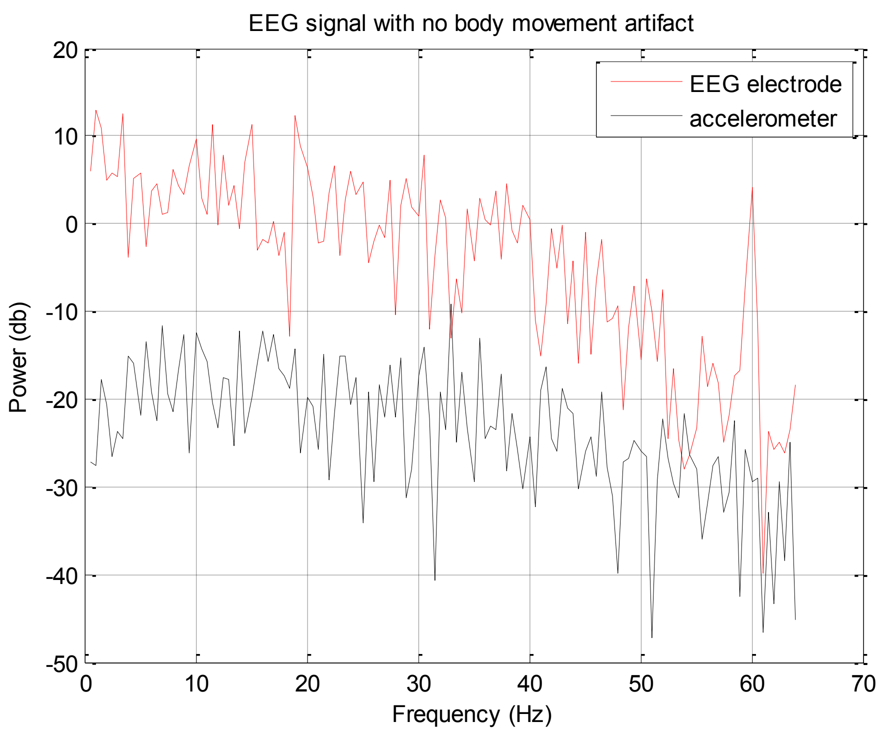


Figure 2-4 Power spectrum of a 2-second segment: subject kept still and read a book silently

Figure 2-5 and Figure 2-6 shows the two cases of reading loudly condition. In Figure 2-5 subject kept still and read a book loudly, and in Figure 2-6 head movement happened once. Using the above linear model, Figure 2-5 also shows the pre-processing result of two epochs that are detected as contaminated by body movement artefact. In Figure 2-5, black line (vibration artefact) is large enough to influence the red line (EEG electrode) at around 18 Hz, and channel 1 remain unaffected for other power bands. In addition, the difference between red line (original EEG) and green line (pre-processed signal) is bigger at corresponding power band compared to other power bands. In Figure 2-6, the subject shake his head for one time and the power of original EEG is abnormally high (up to 28 dB) in 0-5Hz band. The large difference between green line and red line indicates that the regression effect is obvious. The processed signal (green line) is more close to the nature of clean EEG power spectrum as shown in Figure 2-4.

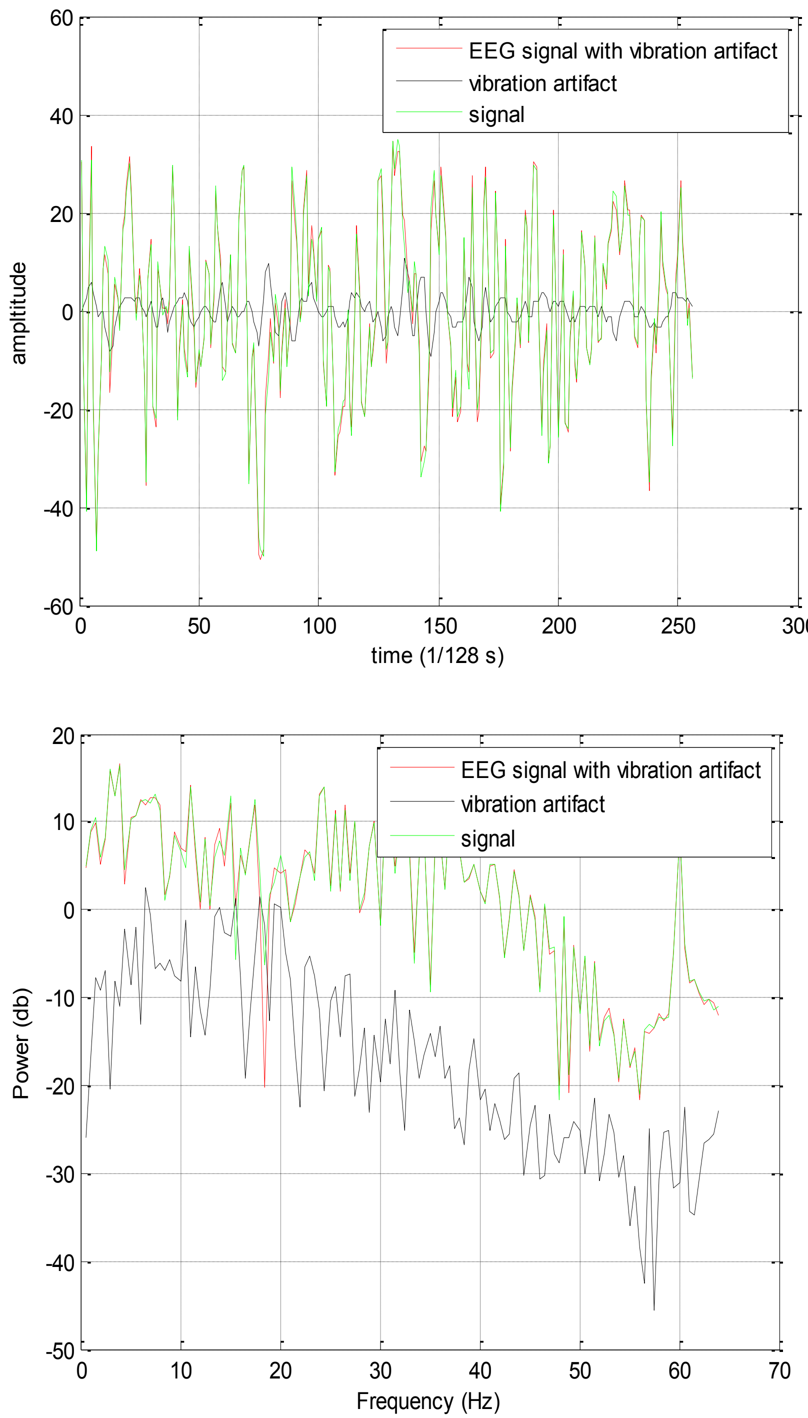
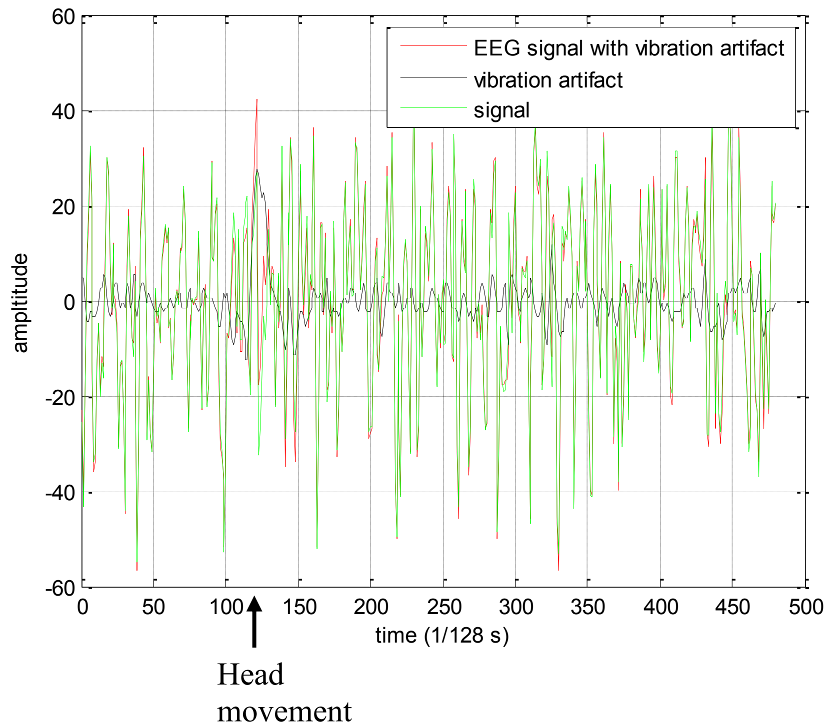


Figure 2-5 Processing of contaminated EEG signals. Subject kept still and read a book loudly k=0.486, b=68.7. Top: time domain; bottom: frequency domain



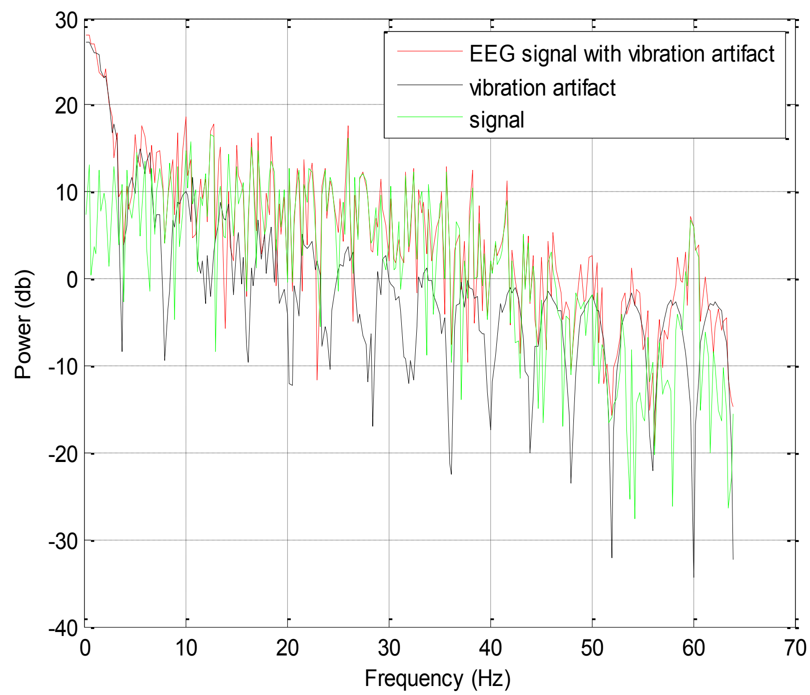


Figure 2-6 Processing of contaminated EEG signals: head movement happened once in reading a book loudly k=0.552, b=59.5. Top: time domain; bottom: frequency domain

Reference

[1] Kerick SE, Oie KS, McDowell K. 2009. Assessment of EEG signal quality in motion environments. ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH AND ENGINEERING DIRECTORATE. (No. ARL-TR-4866).

[2] Peltola, M. A. (2012). Role of editing of R–R intervals in the analysis of heart rate variability. Frontiers in physiology, 3.

[3] Digital Medic Inc [Internet]. Comparison with medical electroencephalogram [cited 2016 Mar 23], available from: http://www.digital-medic.com/products/comparison.

[4] Mishra, A., & Swati, D. (2015). The recursive combination filter approach of pre-processing for the estimation of standard deviation of RR series. Australasian physical & engineering sciences in medicine, 38(3), 413-423.

[5] Mcnames, J., Thong, T., & Aboy, M. (2004, September). Impulse rejection filter for artifact removal in spectral analysis of biomedical signals. In Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE (Vol. 1, pp. 145-148). IEEE.

[6] 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.

[7] 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.

[8] 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.

Chapter 3 Evaluation of mental workload of standard task and simulator task

3.1 Introduction

Generally, mental workload (MWL) or cognitive workload is described as the degree of mismatch between operator's available cognitive resources and those that are demanded for specific tasks. From an operational point of view, workload refers to the individual effort and subjective experience of a particular person performing tasks under a certain working environment and in time during which those duties must be accomplished [1]. What we should keep in mind, is that whilst one side of the 'mismatch' of MWL theory refers to the difficulty, amount, duration and types of specific tasks, usually named as objective task load, the other side refers to the individual cognition capacities. In other words, the same tasks may arouse different levels of individual MWL because operators may have unlikeness in cognitive capabilities, efforts and skills, or they may be in different personal states such as fatigue, excitement and stress. A more general definition is provided by Young and Stanton (2005) [2], who suggested that MWL reflects "the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support, and past experience".

Considering the operation characteristics of shipping, both under load and overload of key operators are potential precursors to human error, as the former leads to reduced vigilance and engagement to task, and the latter leads to inadequate cognitive resources to cope with an emergency. However, a continuous and exact relationship between MWL and performance degradation is rarely easy to build. Young et al. (2014) [3] reviewed related research and proposed a model of performance, task demand and arousal level (see Figure 3-1), which is originated from De Waard (1996) [4] that there is an optimum range of MWL associated with good performance.

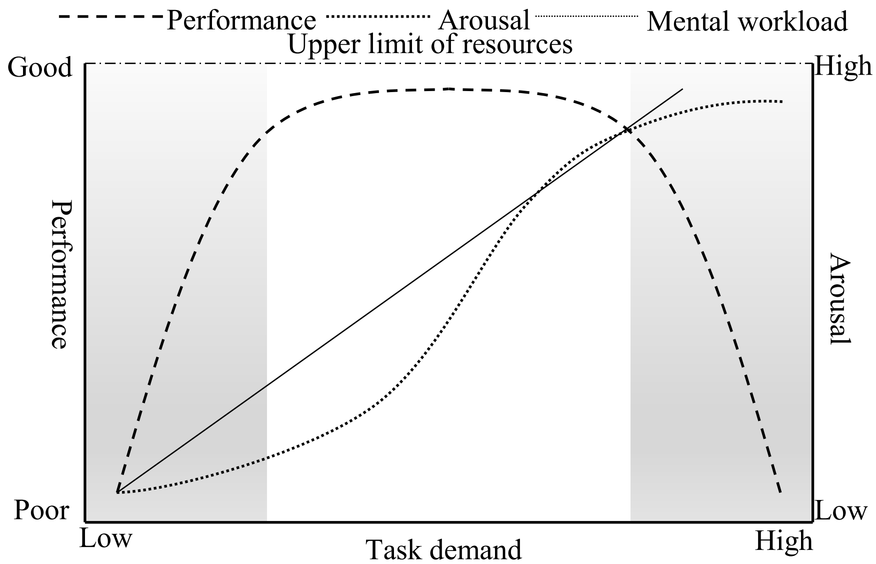


Figure 3-1 Model of task demands, arousal level and performance

(adapted from Young et al. 2014 [3])

An appropriate human information-processing model is crucial to understand the mechanism of operators limited cognitive resources because the channels or dimensions of human information processing determine that of cognitive resources. Mesulam (1998) [5] reviewed research articles in neuropsychology and reported that the human brain contains at least five anatomically distinct networks. Each of the five networks is related to a distinct brain area and is responsible for a different cognition function. For instance, the network for spatial awareness is based in trans-modal epicentres in the posterior parietal cortex, and in the frontal eye fields. The working memory-executive function network is in epicentres in the lateral prefrontal cortex and perhaps the posterior parietal cortex [5]. To some extent, these anatomical findings support the development of multiple resource models proposed by Wickens (1980) [6], first a 3-dimensional model which evolved to a 4-dimensional model by adding visual channels (focal and ambient vision) [7]. In general, the 4-dimensional multiple resource model differentiates pools of mental resources based on four dichotomous dimensions:

- Stage of processing: perception, cognition, responding
- Codes of information processing: spatial and verbal
- Perception modalities dimension: visual and auditory
- Visual Channels: focal and ambient versions

Multiple resource theory makes an important contribution to MWL by predicting the types of performance degradation once the multi-tasking overload has been reached. Many maritime operations need operators' multi-tasking during stressed situations. For example, when leaving or entering a port, navigation officer may simultaneously need to monitor the radar and communicate with engine room regarding changes of ship speed. Additionally, the dimensions of the multiple resource model coincide with relatively straightforward decisions that designers could make when configuring tasks or work spaces to support multi-tasking activities, such as to design the communication methods (auditory or visual) in emergency operation procedures [7].

However, compare to that of civil aviation and road transportation, MWL evaluation research in maritime domain seems much inactive [3] (Table 2). The obstacles of measuring MWL in shipping include: (1) working environment is complex (2) ambulatory physiological sensors are required (3) complexity setting of maritime tasks is challenging. To address these issues, this Chapter studies the conceptual models of MWL and develops MWL measures in standard psychological task and engine-room simulator task. In section 3.2, details of experiment setting are introduced. Section 3.3 describes the performance measures, subjective measure, and physiological measures developed to evaluate MWL in the two types of task. Experiment results are summarized in section 3.4 and section 3.5 is the discussion of these results.

3.2 Experiment

Ten males and one female voluntarily participated in this experiment (age 23.2±2.4). Seven participants were senior students in the bachelor's program at Marine engineering and four graduate participants from a similar program. They all had three to five months on-board cadet experience and grasped the knowledge and technical skills of ship engine system operation. One participant's EEG data were contaminated and excluded from the analysis. For the female participant, HRV data were not available. Therefore, there were 10 sets of data available for both EEG and HRV analysis.

All experiments were conducted in a Marine Engine Plant Simulator (MEPS) room with constant illumination and temperature (Figure 3-2). The experiment tasks were two types: standard visual n-back task and MEPS task each with four difficulty levels. Each experiment for a subject lasted approximately 90 minutes.

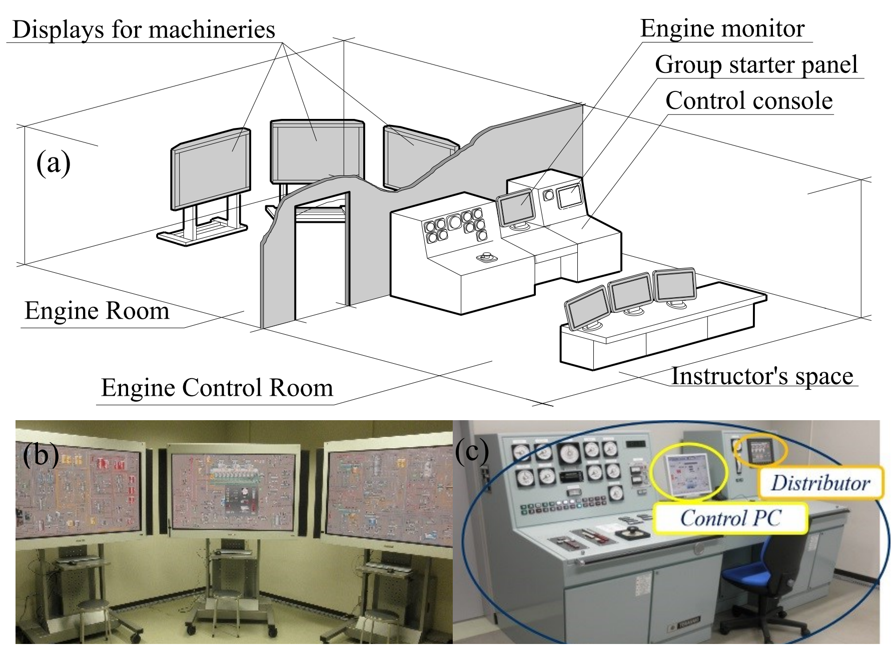


Figure 3-2 (a) outline of MEPS; (b) photograph of simulated engine room; (c) control console

The experiment purpose outlines and their rights were firstly explained to participants. After they agreed to sign the consent form, experiment instructor introduced them n-back task and the participants learned how to fill NASA-TLX. It took about 120 seconds for the participants to wear physiological sensors. The EEG electrodes impedance was then adjusted to acceptable value. Between n-back task and MEPS task, the participants took a rest for about 5 minutes.

N-back task appears to originate as a paradigmatic case for quantifying working memory through brain activity monitoring research [8]. In the most typical variant of n-back task, participants are required to monitor and remember a series of stimuli and to respond whenever a stimulus is presented that is the same as the one presented n trials previously, where n is a specified integer, usually 0,1,2 and 3. In this research n-back task was designed and presented by E-prime 2.0 (Psychology Software Tools), in which capital letters were presented one-by-one for 500ms on a computer screen and blank time between stimulus was 2000ms. Participants were asked to monitor the presented letter and to press predefined target-key when the letter was a target and to press non-target-key when it was not. The task difficulty was incrementally controlled by varying the number of ‘n’ from 0 to 3. In the easiest 0-back task, participants were required to remember a specific letter (X) as the target and they had to press target-key when the letter was presented. In the1-back task, 2-back and 3-back tasks, the target was defined as the letter presented 1, 2 and 3 times before, respectively. Before the beginning of formal experiment, participants were free to practice until they wanted to begin formal experiment. The formal experiment of each difficulty level lasted 75-78 seconds.

Figure 3-3 Subject conducting MEPS tasks

MEPS was a full scale simulator of the engine system of a real ocean-going container ship. MEPS tasks were designed based on the combination of engine system components. Four routine tasks of engine department with different level of difficulty were designed: (1) Transfer diesel oil from settling tank to service tank; (2) Prepare and start the central cooling system; (3) Start diesel engine of NO.2 generator; (4) Start lubrication oil purifier. These tasks were denoted as MEPS-1, MEPS-2, MEPS-3 and MEPS-4 respectively. MEPS-1 required 5 procedures within fuel oil pipes, MEPS-2 required 14 procedures within sea water pipes and fresh water pipes, MEPS-3 required 14 procedures within lubrication oil pipes, fuel oil pipes, compressed air pipes and cooling water pipes, and MEPS-4 required 12 procedures within fresh water pipes, sludge pipes, steam pipes and lubrication oil pipes. Despite the inaccuracy to hastily conclude that one task was absolutely more demanding than another, we generally expected that task with less operation procedures and involving pipes was less demanding than that with more of those. As all participants had marine operation skills, practice session was not arranged except an instruction of task goal.

Table 3-1 Operation procedures of MEPS1

	[MEPS1]
	Diesel Oil (D.O.) TRANS LINE scenario:INDOCK2 (Power supply is required)
□	(D03) OUTLET VALVE on D.O TANK Open
□	(D04) SUC. VALVE for DO TRANS. PUMP Open
□	(D05) DISCH. VALVE for DO TRANS. PUMP Open
□	(D06) INLET VALVE for DO SETT.TANK Open
□	D.O. TRANS. PUMP Start

Table 3-2 Operation procedures of MEPS2

	[MEPS2]
	CCFW & SW LINE scenario:INDOCK2 (Power supply is required)
	・CENTRAL CSW LINE (S/B PUMP is not need)
□	(C01) HIGH SEA CHEST OUTLET VALVE remote Opened
■	*(C02) HIGH SEA CHEST OUTLET VALVE Opend @ECC
□	(C09) No.1 CENTRAL FW CLR SW INLET VALVE Open
□	(C10) No.1 CENTRAL FW CLR SW OUTLET VALVE Open
□	(C11) No.2 CENTRAL FW CLR SW INLET VALVE Open
□	(C12) No.2 CENTRAL FW CLR SW OUTLET VALVE Open
□	(C13) MAIN SW CIRC. LINE SW over board valve Open
□	(C05) No.1 CENTRAL CSW PUMP INLET VALVE Open
□	No.1 CENTRAL CSW PUMP Start(low)
□	(C06) No.1 CENTRAL CSW PUMP OUTLET VALVE Open
■	*(C07) No.2 CENTRAL CSW PUMP INLET VALVE Open
■	*(C08) No.2 CENTRAL CSW PUMP OUTLET VALVE Open
■	*No.2 CENTRAL CSW PUMP STANDBY @GSP
■	*(No.1 No.2 CENTRAL CSW PUMP AUTO)
	・M/E LO COOLER CFW LINE (S/B PUMP is not need)
□	(C20) No.1 CENTRAL FW CLR INLET VALVE Open
□	(C21) No.1 CENTRAL FW CLR OUTLET VALVE Open
□	(C22) No.2 CENTRAL FW CLR INLET VALVE Open
□	(C23) No.2 CENTRAL FW CLR OUTLET VALVE Open
□	(C37) M/E LO CLR INLET VALVE Open
□	(C38) from M/E LO CLR OUTLET VALVE Open
□	(C) No.1 CENTRAL CFW PUMP INLET VALVE Open
□	No.1 CENTRAL CFW PUMP Start (low)
□	(C) No.1 CENTRAL CFW PUMP OUTLET VALVE Open

Table 3-3 Operation procedures of MEPS3

[MEPS3]

START No.2 DIESEL GENERATOR. Scenario: IN PORT

(D/G control position REMOTO → LOCAL)

・No.2 D/G LO LINE

□ (L89) No.2 D/G PRIMING LO PUMP OUTLET VALVE Open

□ (L90) No.2 D/G LO PUMP OUTLET VALVE Open

*No.2 LO PRIM PUMP AUTO on GSP

・No.2 D/G FO LINE

□ (D21) DRIVEN PUMP INLET VALVE (No.2 D/G) Open

□ (D35) DRIVEN PUMP OUTLET VALVE (No.2 D/G) Open

□ (F80) No.2 D/G SUPPLY VALVE Open

□ (D23) No.2 D/G OUTLET VALVE (to DO RETURN RECEIVER) Open

・No.2 D/G AIR LINE

□ (A09) No.2 D/G INLET (double valve) D/G side Open

・No.2 D/G CFW LINE

□ (C55) INLET VALVE for No.2 D/G LO CLR Open

□ (C56) OUTLET VALVE No.2 D/G AIR CLR Open

□ (C59) OUTLET VALVE No.2 D/G Open

□ (C58) OUTLET VALVE No.2 D/G CFW PUMP Open

・TURNING

□ TURNING BAR SET on No.2 D/G OPERATION PANEL

□ No.2 D/G Start

Table 3-4 Operation procedures of MEPS4

	[MEPS4]
	START LO PURIFIER. Scenario:IN PORT
	・OPERATION WATER
□	(W18) FW INLET VALVE for No.1 LO PURIFIER Open
	・SLUDGE LINE
□	(L47) No.1 LO PURIFIER OUTLET VALVE Open
	・ STEAM LINE for LO HEATER
□	(S40) INLET VALVE for No.1 LO PURIFIER HEATER Open
	・NO.1 LO PURIFIER LINE
■	*(L17) OUTLET VALVE on M/E LO SUMP TANK Opened
■	*(L18) SUPPLY VALVE to LO PURIFIER Opened
■	*(L06) INLET VALVE on M/E LO SUMP TANK Opened
□	(L19) INTERMID. VALVE to LO PURIFIER Open
□	(L20) INLET VALVE for No.1 LO PURIFIER Open
□	(L20) INLET VALVE for No.1 LO PURIFIER GEAR PUMP Open
□	(L22) INLET VALVE for No.1 LO PURIFIER Open
□	(L26) RETUERN VALVE from 3WAY VALVE to GEAR PUMP Open
□	(L23) DISCHARGE VALVE from No.1 LO PURIFIER Open
□	(L24) INTERMID. VALVE (non return) from NO.1 LO PURIFIER Open
□	(L25) INTERMID. VALVE to M/E LO SUMP TANK Open
□	No.1 LO PURIFIER Start

3.3 Measures

NASA task load index is one of the most widely used subjective MWL measures and its validity has been recurrently tested in laboratory and field studies since its first proposal, developed by Hart and Staveland [9]. Nevertheless, quite lot variants of original NASA-TLX technique, appear in various industries. Some noticeable developments are the software-based rating techniques and the simplified ones [10]. The NASA-TLX uses six dimensions to assess mental workload: mental demand, physical demand, temporal demand, performance, effort, and frustration. Table 3-2 shows the definitions of NASA-TLX dimensions. Twenty-step bipolar scales are used to obtain ratings for these dimensions. A score from 0 to 100 (assigned to the nearest point 5) is obtained on each scale. A weighting procedure is used to combine the six individual scale ratings into a global score; this procedure requires a paired comparison task to be performed prior to the workload assessments. Paired comparisons require the operator to choose which dimension is more relevant to workload across all pairs of the six dimensions. The number of times a dimension is chosen as more relevant is the weighting of that dimension scale for a given task for that operator. A workload score from 0 to 100 is obtained for each rated task by multiplying the weight by the individual dimension scale score, summing across scales, and dividing by 15 (the total number of paired comparisons).

Table 3-5 Six dimensions of NASA-TLX

Title	Description
Mental Demand	How much mental and perceptual activity was required (e.g. Thinking, deciding, calculating, remembering, looking, searching.)? Was the task easy or demanding, simple or complex, exacting or forgiving
Physical Demand	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals?

Table 3-5 continued in the next page

Table 3-5 continued

Effort	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

A paper-based Japanese version of NASA-TLX (show in Appendix 1) was used in this experiment to collect the participants' feeling immediately after each difficulty level of task. A weighted score of six MWL dimensions was calculated.

EEG was measured by an ambulatory device (Digital Medic Inc.) consisted of a headset (17*16*1.5 cm 80g) and main part (5.8*12*2.4 cm 95g) with a sampling frequency of 128Hz. The headset is comprised of two electrodes and an accelerometer attached to the electrodes. Body movement such as walking and jogging has been found to be one of the most influential origin of EEG signal artefact [11]. When the acceleration of the electrodes exceeds certain value, corresponding epochs of EEG data were removed from further analysis. The main part wirelessly connects to an online monitoring computer by Direct Sequence Spread Spectrum (DSSS). Effective wireless transmission distance were 10 meters. The detected brain area was posterior parietal cortex (P3-O1, P4-O2 of standard EEG channel nomenclature). Digital Medic Inc. [12] compared the measurement result in 128Hz between this EEG device and a medical one, the average correlation coefficient was 0.94. Time domain data for each difficulty level of two types of tasks were discrete fast Fourier transformed (DFFT) to three wave bands in every second. These bands were theta (4-7Hz), alpha (8-13Hz) and beta (20-25Hz) and the power spectral was estimated based on period-gram method. Alpha wave rate was defined as

$$r_\alpha = \frac{P_\alpha}{P_\theta + P_\alpha + P_\beta} \tag{3-1}$$

P_α was the power of alpha band in one second. The average rate of each wave band, $\frac{\sum r_\alpha}{l}$, $\frac{\sum r_\beta}{l}$, $\frac{\sum r_\theta}{l}$ (l was time length of each task) were chosen as EEG indices of each task.

HRV data was measured by a chest belt heart beat sensor RS800CX (Polar Electronics). Raw, continuous HRV data for each difficulty level of two types of tasks were cropped to epochs when participants were performing tasks. Three measures were analysed: (1) mean value of heart beat interval (MHBI) (2) standard deviation of heart beat interval (SDHBI) (3) FFT-based LF/HF ration. Low Frequency (LF) was defined as 0.04-0.15Hz and High Frequency (HF) was defined as 0.15-0.4Hz.

3.4 Results

Less than one percent of EEG signal for n-back task and 5 percent for MEPS task was contaminated by body movement and removed from analysis.

3.4.1 Descriptive statistics

Table 3-6 Descriptive statistics of n-back task and MEPS task each with four difficulty levels

Mean	n-back task				MEPS task			
SD	0-back	1-back	2-back	3-back	1(5+1)	2(14+2)	3(14+4)	4(12+4)
Performance	0.3	0.6	1.6	3.8	72.3	141.6	227.7	220.7
	0.7	0.7	1.3	2.6	61.8	118.2	148.6	174.6
NASA-TLX	4.2	5.9	9.6	12.6	4.4	6.2	7.6	6.9
	2.0	2.5	2.8	2.5	4.9	5.6	5.7	5.6
Alpha wave rate	54.8	52.0	50.2	49.4	43.6	48.0	41.3	44.3
(%)	6.6	5.7	4.0	6.9	4.2	4.2	3.9	2.8
Beta wave rate	18.2	19.1	21.1	20.2	19.1	20.8	15.6	15.6
(%)	7.3	8.5	7.3	8.7	7.1	7.1	5.4	4.7
Theta wave rate	26.9	29.1	29.2	30.5	35.5	31.2	42.9	40.6
(%)	9.1	8.3	8.2	10.7	9.0	8.5	6.8	5.9
MHBI(ms)	760.6	776.1	724.7	762.8	741.1	746.5	729.9	734.6
	163.2	142.7	117.9	104.7	109.7	105.5	97.6	93.5
SDHBI	41.9	48.0	43.2	46.9	43.0	41.2	43.8	42.2
	19.6	34.6	20.5	16.9	19.6	12.8	13.1	12.4
LF/HF	1.8	1.8	2.6	4.2	4.5	5.4	5.0	5.4
	1.5	1.4	2.8	4.2	4.0	4.4	4.7	5.6

The results (mean and standard deviation of 10 participants) of seven measures, NASA-TLX, alpha wave rate, beta wave rate, theta wave rate, MHBI, SDHBI, and LF/HF were summarized in Table 3-6. The objective difficulty of MEPS tasks were manipulated by the sum of operation steps and pipe types, noted in brackets after task serial number, in incremental order of MEPS-1, MEPS-2, MEPS-4, and MEPS-3. The performance of n-back task was evaluated by the number of mistaken responses while the performance of MEPS task was evaluated by total time that operator consumed to achieve task goal. The unit of MEPS task performance measure was second. As shown in Figure

3-4, the average value of NASA-TLX and performance measure corresponded with objective difficulty for both n-back task and MEPS task.

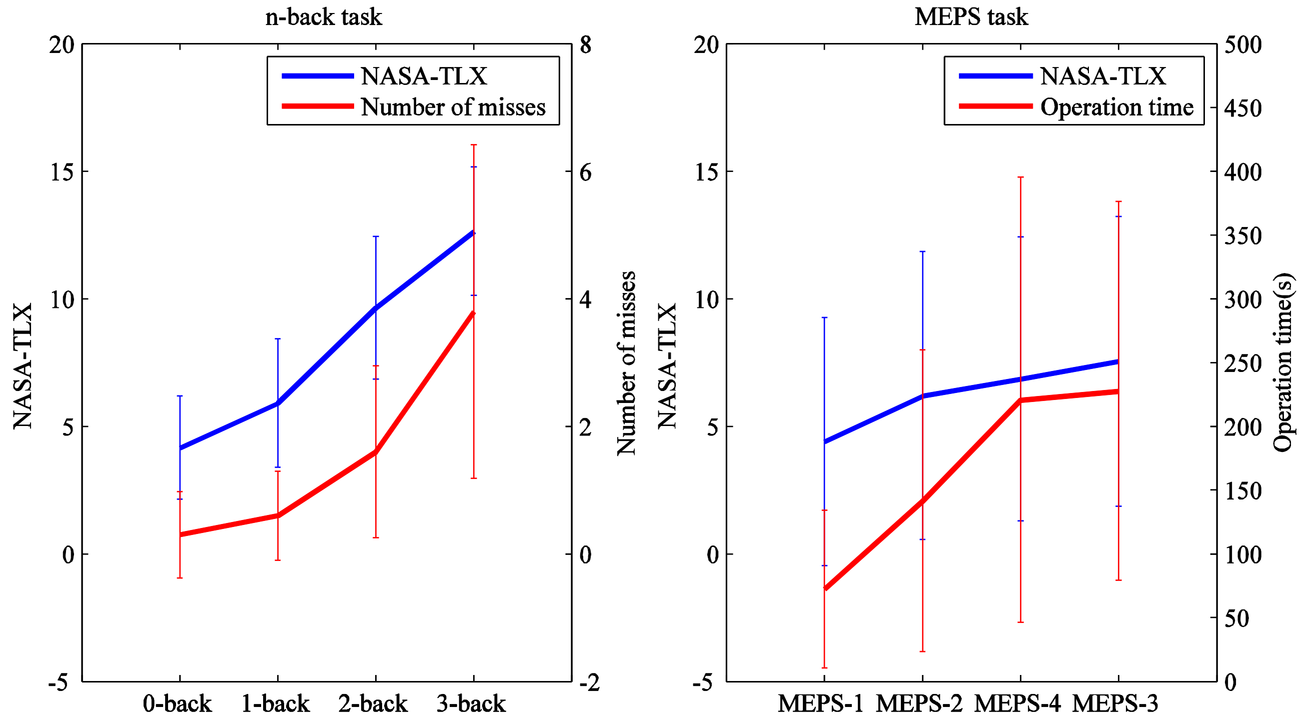


Figure 3-4 Result of NASA-TLX and task performance, error bar is standard deviation

3.4.2 Sensitivity of MWL measures

If MWL measures were sensitive to task demand, the average value under different level of task should be significantly different. The null hypothesis of ANOVA was then "there is no significant difference between MWL measures that under each difficulty level of task". Each subject's physiological data and NASA-TLX score for eight tasks were pre-processed through zero-mean normalization to eliminate the effect of individual difference. Since the measurements of n-back task performance and MEPS task performance were in different orders of magnitude, performance measures were normalized within task type. The zero-mean normalization function was

$x^* = \frac{x-\mu}{\sigma}$, where μ is the mean value and σ is the standard deviation. For n-back tasks and MEPS tasks, the main effect of difficulty level on the seven measures was tested separately using one-way repeated measure ANOVA, significance criterion is set at p=0.05.

Table 3-7 Main effect analysis of variance for n-back task and MEPS task

Measures	n-back task	MEPS task
Performance	$F(3,36)=9.66$, $p<<0.001$	$F(3,36)=33.9$, $p<<0.001$
NASA-TLX	$F(3,36)=21.7$, $p<<0.001$	Non-significant
Alpha wave rate	$F(3,36)=14.2$, $p<0.001$	$F(3,36)=17.4$, $p<<0.001$
Beta wave rate	Non-significant	$F(3,36)=7.03$, $p<0.001$
Theta wave rate	Non-significant	$F(3,36)=18.0$, $p<0.001$
MHBI	Non-significant	Non-significant
SDHBI	Non-significant	Non-significant
LF/HF	Non-significant	Non-significant

Table 3-7 shows the main effect results of ANOVA. For n-back tasks, main effect on performance measures, NASA-TLX, and alpha wave rate were significant. Post-hoc tests for pairwise comparison (6 pairs) showed that beta wave rate difference between 0-2 back (p=0.015) tasks was significant, theta wave rate difference between 0-3 back (p=0.033) tasks was significant. For heart ratio related measures, post-hoc tests showed that MHBI difference for 1-2 back (p=0.036) and 2-3 back (p=0.039) were significant, and LF/HF difference between 0-3 back (p=0.021) was significant.

For MEPS tasks, significant main effect on performance measures and all three EEG related measures were found while main effect on NASA-TLX, NHBI, SDHBI and LF/HF were not significant. Post-hoc tests for pairwise comparison showed that NASA-TLX difference of MEPS1-MEPS3 ($p=0.011$) and MEPS1-MEPS4 ($p=0.030$) were significant. Post-hoc tests found no significant difference for HRV related measures.

3.4.3 Validity of MWL measures

Pearson's correlation coefficient was calculated to examine the linear correlation among performance measures, NASA-TLX, and physiological indices (Table 3-8). Each subject's data was pre-processed through within task type zero-mean normalization. There were four difficulty levels and ten participants, and then the data length was 4*10=40. For n-back task, performance measure correlated strongly with NASA-TLX at a significance level ($r=0.634$, $p<<0.001$), indicating that with the increase of subjective MWL score, the number of mistaken response also increased. Alpha wave rate ($r=-0.653$, $p<<0.001$), beta wave rate ($r=0.404$, $p=0.01$) and LF/HF ($r=0.313$, $p=0.049$) also showed different degree of correlation with NASA-TLX. The strong negative correlation of alpha wave rate indicated that higher task demand induced higher subjective MWL and lower alpha wave rate. This finding of alpha wave suppression corresponded with the findings of Fairclough et al. (2005) [13] and Slobounov et al. (2000) [14], who reported that EEG spectral power in the alpha band decreased during complex and cognitively demanding tasks. For MEPS task, performance measure correlated strongly with NASA-TLX at a significance level ($r=0.717$, $p<<0.001$), indicating that with the increase of subjective MWL score, the complete time of MEPS operation increased. Compare to that of n-back task, the correlation between alpha wave rate and NASA-TLX of MEPS task was not significant ($p>0.5$), and the correlation between beta wave rate and NASA-TLX was negative ($r=-0.565$, $p<0.001$). Theta wave rate seemed to be the universal physiological indices in both n-back task ($r=0.404$, $p=0.01$) and MEPS task ($r=0.531$, $p<0.001$).

Table 3-8 Pearson's correlation coefficients between performance measures, NASA-TLX and Physiological indices

Measures	n-back task		MEPS task	
	Performance	NASA-TLX	Performance	NASA-TLX
NASA-TLX	0.634	1.000	0.717	1.000
Alpha wave rate	-0.393	-0.653	-0.254	-0.063
Beta wave rate	0.241	0.404	-0.525	-0.565
Theta wave rate	0.170	0.237	0.568	0.531
MHBI	0.062	-0.219	-0.263	-0.149
SDHBI	0.060	-0.081	-0.008	0.093
LF/HF	0.275	0.313	0.032	-0.016

3.4.4 Clustering analysis

To further study the physiological response during the process of conducting n-back tasks, physiological features were calculated in a window of 4-second. Then each data point is considered as a sample of a corresponding cluster and each physiological feature is considered as a variable. As the experiment time of n-back tasks were 76-78 seconds, there were 19 data points of each level of task. Cluster centre is calculated as the mean value of each cluster. As shown in Figure 3-5 shows the clustering result of subject 8. Figure 3-5 (a), along with the difficulty increase from 0-

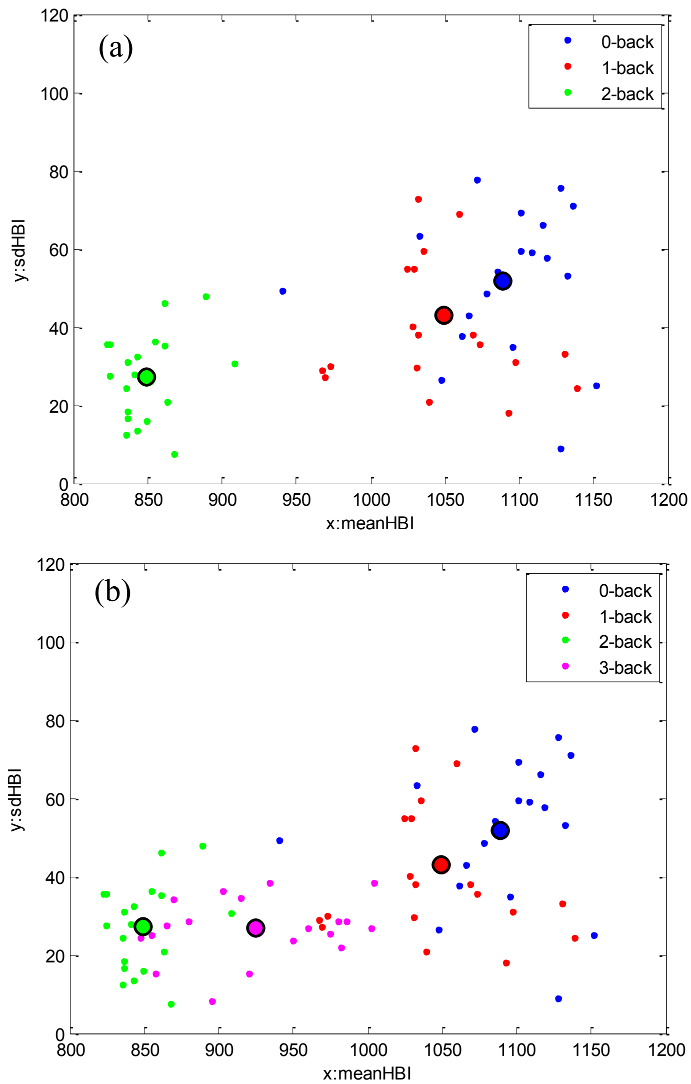


Figure 3-5 Subject 8: ceiling effect. (a) without 3-back cluster; (b) with 3-back cluster

back to 2-back task, both sdHBI and MHBI shows decreasing tendency. However, the clustering of 3-back task falls between 2-back and 1-back task, which is inconsistent with the difficulty setting.

3.5 Discussions

Participants reported higher MWL in n-back task than in MEPS task while they had lower alpha wave suppression in n-back task than in MEPS task. This contradiction between subjective report and physiological indices could be explained by that MEPS task required more than one channel of cognitive information processing that corresponded with multidimensional limited cognitive resource model proposed by Wickens (1984) [6], first a 3-dimensional model which evolved to a 4-dimensional model by adding visual channels [7]. In general, the 4-dimensional limited resource model differentiates pools of cognitive resources based on four dichotomous dimensions: Stage of processing: perception, cognition, responding; Codes of information processing: spatial and verbal; Perception modalities dimension: visual and auditory; Visual Channels: focal and ambient versions. Additionally, there were more significant physiological indices in MEPS task than in n-back task, and the physiological indices correlated to subjective mental workload were different in MEPS and n-back task. Chen and Epps (2014) [15] found that cognitive load measurement was affected by different task types through using pupil diameter and blink measures to infer cognitive load and perceptual load. Similarly, this disparity could be explained by that MEPS task required visual and auditory perception, spatial information processing, and long and short term working memory while n-back task only required visual perception, verbal information processing, and short-term working memory.

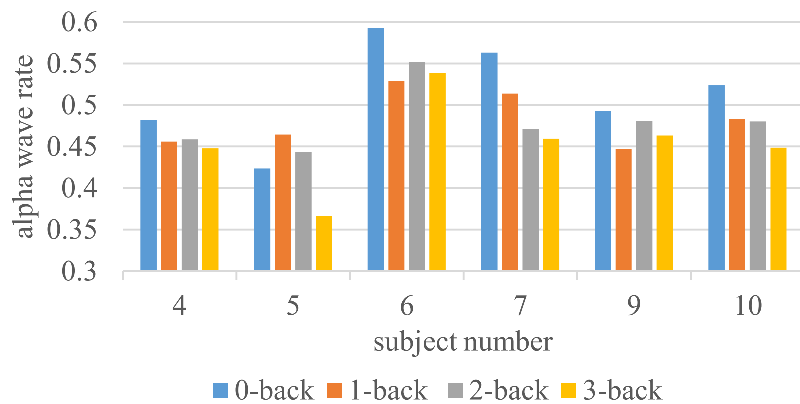


Figure 3-6 Alpha wave suppression with increasing task demand

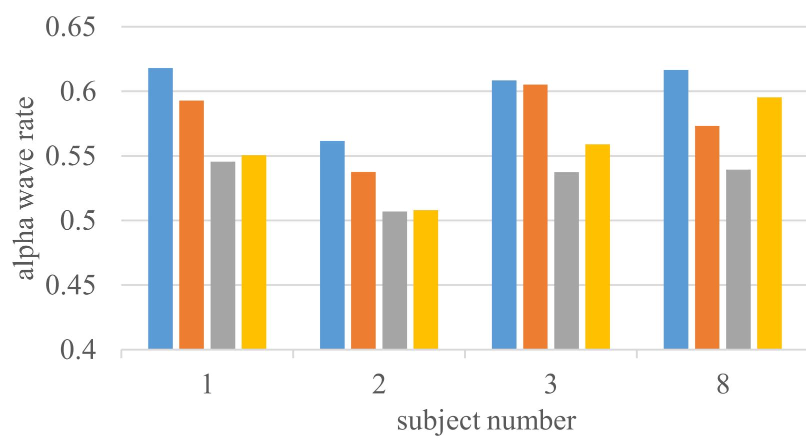


Figure 3-7Alpha wave suppression with increasing task demand,

results of 4 subjects show ceiling effect

As shown in Figure 3-7, four participants had higher alpha wave rate in 3-back task than in 2-back task, and another two participants had higher alpha wave rate in 2-back task than in 1-back task. When task demand exceeded the capacity of mental resource, the operator failed to catch up with the pace of task and showed lower activation level, and this was generally associated with different degree of performance degradation. However, individuals' capacity of mental resources were different. In Figure 3-6, four participants' alpha wave rate (subject 4, 5, 7, and 10) showed decreasing tendency from 0-back to 3-back task, indicating that even 3-back task did not exceed their mental capacity. This effect was reported by Stuiver et al. (2014) [16] as a ceiling effect and by Durantin et al. (2014) as an inverted U-shape curve of using physiological indices to infer human MWL. Ceiling effect was also found in MEPS tasks, two participants had higher alpha wave rate in the task that they subjectively reported as most difficult (highest NASA-TLX).

Regarding the implications of the findings for maritime safety, measuring operators' MWL based on physiological metrics makes OFS-AA systems more applicable to be adopted in maritime operation. In MEPS task, while the difficulty level did not have main effect on subjective ratings, EEG indices showed higher sensitivity. This result evidently supported the opinion of De Waard (2014) [17] that self-report scales, which are used on-board to ensure seafarer's rest time, alone cannot capture MWL. Besides, alpha wave rate was extracted from EEG measured by a portable and low-cost device, the shortcomings of physiological data collection, including vulnerability to artefact and prolonged device wearing time, were alleviated in this research. Shipping companies have long been suffering from a lack of competent seafarers because the training, evaluation and certificating of seafarers were carried out in a relatively low level. With knowing the trainee's real-time MWL, the instructors are able to conduct an instruction in an appropriate pace for seafarer's training in simulator environment. Besides, objective measurement of MWL can also provide a more reliable method to evaluate the trainee's competence of technical and nontechnical skills.

This paper conducted a systematic research on MWL to fill the gap between practical applications and the theoretic human factor models. An experiment study was designed to further extend the available physiological metrics based on wearable and portable devices that are feasible in maritime environment. Task demand of maritime operation was successfully manipulated by involving different number of operation steps and pipeline types. ANOVA was used to test the sensitivity of MWL measures and Pearson's correlation coefficient was calculated between NASA-TLX, performance measures, and physiological indices to check the validity of these measures. The conclusions of this paper include:

(1) Mental workload models have potential applications in shipping industry, but these applications require reliable, sensitive, and online measurement of MWL.

(2) Alpha band wave suppression and subjective self-report MWL are sensitive to n-back task demand while heart ratio related measures are not. Alpha (8-13Hz) band wave suppression, beta wave band (20-25Hz), and LF/HF correlate with subjective MWL in n-back task

(3) Three EEG features are sensitive to MEPS task demand. Beta band (20-25Hz) and theta band (4-7Hz) correlate with subjective MWL in MEPS task.

(4) Ceiling effect of using physiological metrics to infer human MWL are found.

As different physiological indices are found to be sensitive for n-back task and MEPS task, one important implication of this finding is that development of cross-task MWL measures must be implemented based on sufficient analysis of task attributes. Maritime operations require multidimensional mental resource whilst standard tasks, such as n-back task and mental arithmetic task, require one or two dimensions in a relatively constant manner. A combination of reference tasks that can effectively account for complex operation tasks should be utilized to provide training data for MWL prediction.

References

[1] 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.

[2] Mark S. Young and Neville A. Stanton (2004). Mental workload. In: Handbook of human factors and ergonomics methods. CRC press, pp. 39.1-39.9.

[3] Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. Ergonomics, 58(1), 1-17.

[4] De Waard, D., & Lewis-Evans, B. (2014). Self-report scales alone cannot capture mental workload. Cognition, Technology & Work, 16(3), 303-305.

[5] Mesulam, M. M. (1998). From sensation to cognition. Brain, 121(6), 1013-1052.

[6] Wickens, C. D. (1980). The structure of attentional resources. Attention and performance VIII, 8, 239-257.

[7] Wickens, C. D. (2008). Multiple resources and mental workload. Human Factors: The Journal of the Human Factors and Ergonomics Society, 50(3), 449-455.

[8] Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N ‐ back working memory paradigm: A meta ‐ analysis of normative functional neuroimaging studies. Human brain mapping, 25(1), 46-59.

[9] Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology, 52, 139-183.

[10] Cao, Alex, et al. (2009). NASA TLX: Software for assessing subjective mental workload. Behavior research methods 41(1), 113-117.

[11] Kerick SE, Oie KS, McDowell K. 2009. Assessment of EEG signal quality in motion environments. ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD HUMAN RESEARCH AND ENGINEERING DIRECTORATE. (No. ARL-TR-4866).

[12] Digital Medic Inc [Internet]. Comparison with medical electroencephalogram [cited 2016 Mar 23], available from: http://www.digital-medic.com/products/comparison.

[13] Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. International Journal of Psychophysiology, 56(2), 171-184.

[14] Slobounov, S. M., Fukada, K., Simon, R., Rearick, M., & Ray, W. (2000). Neurophysiological and behavioral indices of time pressure effects on visuomotor task performance. Cognitive Brain Research, 9(3), 287-298.

[15] Chen, S., & Epps, J. (2014). Using task-induced pupil diameter and blink rate to infer cognitive load. Human–Computer Interaction, 29(4), 390-413.

[16] Stuiver, A., Brookhuis, K. A., de Waard, D., & Mulder, B. (2014). Short-term cardiovascular measures for driver support: increasing sensitivity for detecting changes in mental workload. International Journal of Psychophysiology, 92(1), 35-41.

[17] De Waard, D., & Lewis-Evans, B. (2014). Self-report scales alone cannot capture mental workload. Cognition, Technology & Work, 16(3), 303-305.

Chapter 4 Evaluation of Operator Fatigue and Performance during Pipeline Work

4.1 Introduction

Fatigue has been identified as one of the most significant factors that leads to accidents in a wide range of industries. Just like many other conceptual constructs of human factor, fatigue is rarely well defined since its complexity in practical situations. Fatigue is caused by many factors and linked to a series of degradation of human capability. It generally falls into two categories: mental and physical according to the definition given by the International Maritime Organization, “A reduction in physical and/or mental capability as the result of physical, mental or emotional exertion which may impair nearly all physical abilities including: strength; speed; reaction time; coordination; decision making; or balance” [1]. Furthermore, operators always have to utilize both physical exertion and mental attention when conducting real-world tasks, such as connecting and tightening pipe flanges, whose perfect sealability is crucial for the safety of the entire power plant. Generally, an operator’s competence to fulfil a specific task degrades along with as physical and mental fatigue accumulates.

In the maintenance of power plants, the operators have to conduct much pipeline work to keep the machinery system run normally, especially for ocean-going ships that are characterized by a closed environment. The engineer crew have to fulfil specific maintenance work with limited personnel and equipment resource. In many circumstances, they continuously work for prolonged time in a high intensity. Operators would seek for a rest to recovery from fatigue when they feel tired, but sometimes their subjective feelings may drift from the objective degree of fatigue and affect their work performance. An objective detection and alarm of a high degree of fatigue can improve this situation in two aspects: the operator can actively adopt a rest and/or the manager can take intervention counter measures.

The Borg's Rating of Perceived Exertion (RPE) is one subjective method to measure an individual's perception of exertion during physical work or exercise [2]. RPE provides a way to measure the intensity of physical exertion and has been found to correlate positively with an increase of heart rate (HR) [3]. However, the accumulation of fatigue and exhaustion also happens when workload decreases and HR keeps constant. HR dynamics is a complex interactive process that reflects instantaneous changes of body position, physical movement, and mental state. It is therefore inaccurate to simply use the value of HR as an objective measure of physical fatigue. In addition to absolute HR value, features extracted from time/frequency/nonlinear based methods of heart rate variability (HRV) have been developed for clinical diagnosis and athletic exercise [4,5]. Based on detrended fluctuation analysis (DFA), Chen et al. [6, 7] studied the fractal properties and developed a cardiac stress index (CSI) to measure a subject's cardiac stress status during a cycling exercise of a relatively short term. Nevertheless, little research attention has been paid to an operator's short-term fatigue in actual or quasi-actual working situations.

This paper aims to study the accumulation of fatigue and its effect on working accuracy during pipeline work. Section 2 elaborates the methods and materials, including the pre-processing of original RR interval series, the methods of DFA, and the experiment setting. Section 3 presents and discusses results of the experimental study. Finally, conclusions are drawn in section 4.

4.2 Methods and Materials

4.2.1 Experiment Settings

Eleven male university students (age 26±2.8) voluntarily participated in the experiment study, which was approved and conducted complying with the Kobe University Guidance of Research on Human Subjects. All experiments were conducted between 1-4 o'clock p.m. Caffeine intake and strenuous exercise were prohibited on the day of the experiment. Figure 4-1 shows the flow of the experiment settings. Before the participants signed the consent form, experiment purpose, its outline and their rights were explained. The experiments were arranged in the following temporal order: instruction and preparation (wear heart rate sensor), practice, two-minute rest, formal experiment, and two-minute recovery baseline. The recovery baseline condition was measured three minutes after the formal experiment. Heart rate monitoring started before the two-minute rest.

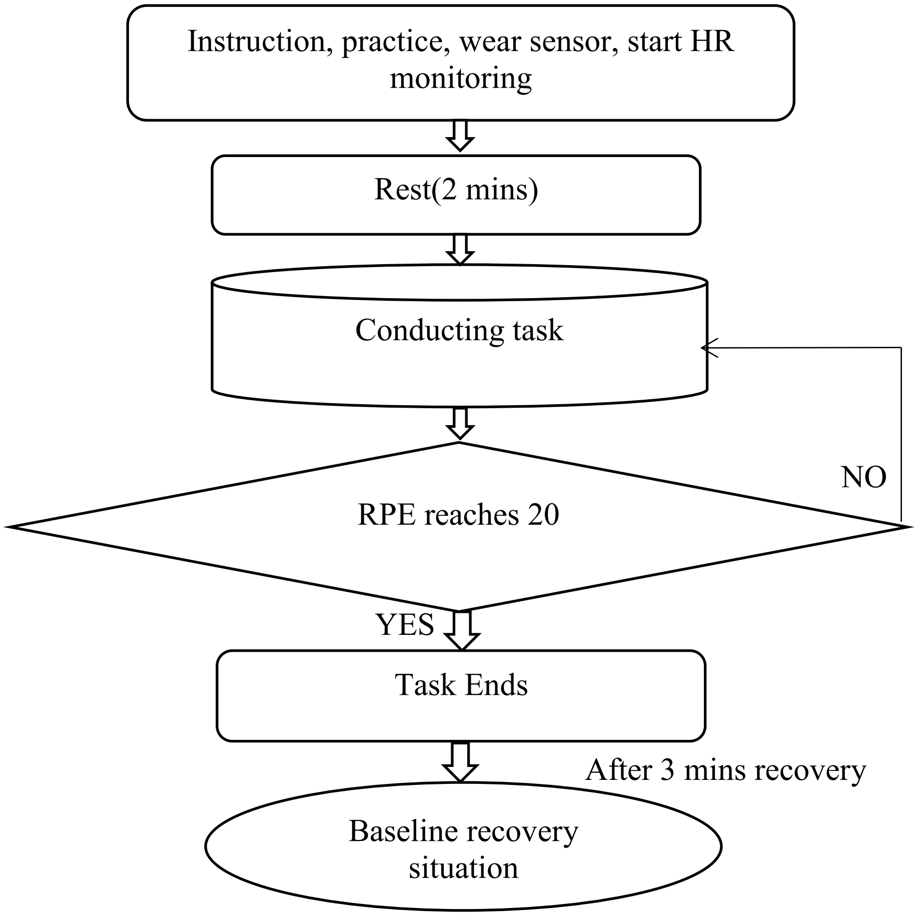


Figure 4-1 Flow chart of experiment settings

Figure 4-2 shows the experiment apparatus, each flange that connected pipes consisted of four bolts. The bolts need to be tightened evenly and moderately to ensure sealability. In the practice session, the participants used a torque-measuring wrench to get used to a torque of 20Nm. During the formal experiment, the participants were instructed to tighten each of the bolts in a diagonally pattern, as evenly as possible using two traditional wrenches. The torque variance of the four bolts on each flange was used as a measure of work performance. Participants reported their subjective degree of fatigue corresponding to a number in Borg's RPE scale every two flanges. The experiment ended if a participant reported an RPE scale of 20 (exhausted) or finished tightening all 26 flanges.

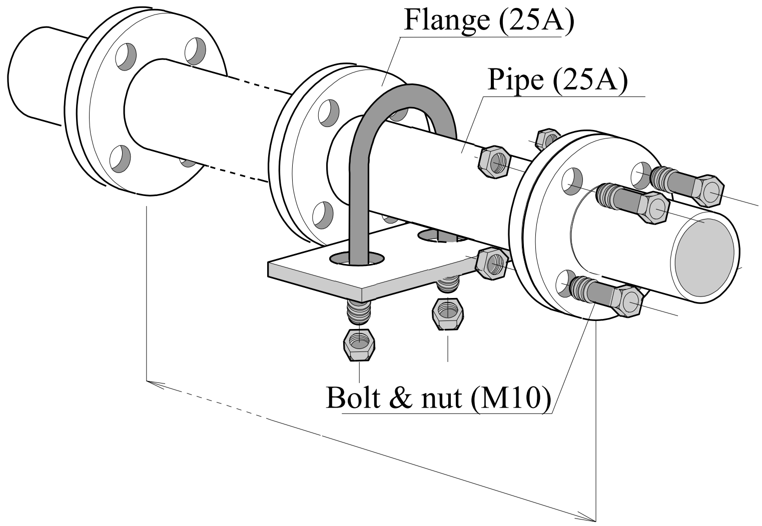


Figure 4-2 Sketch of experiment apparatus, flanges and bolts

4.2.2 Detrended Fluctuation Analysis (DFA)

DFA was firstly proposed by Peng et al. [8] and has been widely used to study the fractal properties and the long-term autocorrelations of a nonstationary time series. Fractal geometry was originally used to depict the roughness of a surface, but is applicable to time series data of the following essence: a process with stronger fractal characteristics does not adhere to equilibrium around any specific scale such as a constant heart rate [9]. According to [8,10], monofractal DFA consists of four steps as shown in equation (4-1) to (4-4). First, calculate the cumulative deviation of signal x, where $\bar{x}$ is the mean of x:

$$Y(i) = \sum_{k=1}^{i}[x_k - \bar{x}] \, , i = 1, \ldots, N \tag{4-1}$$

Then divide $Y(i)$ into $N_s = int(N/s)$ nonoverlapping segments of length s. Calculate the local trend for each N_s segment by a least-square fit of the series and determine the variance for each segment v=1,…,Ns:

$$F^2(v,s) = \frac{1}{s}\sum_{i=1}^{s}\{Y[(v-1)s+i] - y_v(i)\}^2 \tag{4-2}$$

For monofractal analysis, average $F^2(v,s)$ over all segments to obtain the second order fluctuation $F(s)$:

$$F(s) = \sqrt{\frac{1}{N_s}\sum_{v=1}^{N_s} F^2(v,s)} \tag{4-3}$$

Hurst exponent (HE) h that characterizes fractal properties is then extracted from the slope by fitting the log-log linear relationship between $F(s)$ and s. s ranged from 4 to 60 with a step size of 2 in this paper.

$$h = \frac{log_2^{F(s)}}{log_2^{s}} + c \tag{4-4}$$

According to [11], HE of biomedical signals generally ranges from 0.5 to 1.5. HE is 0.5 when the signal is white noise (Gauss distributed), while 1.0 indicates pink noise and 1.5 indicates brown noise.

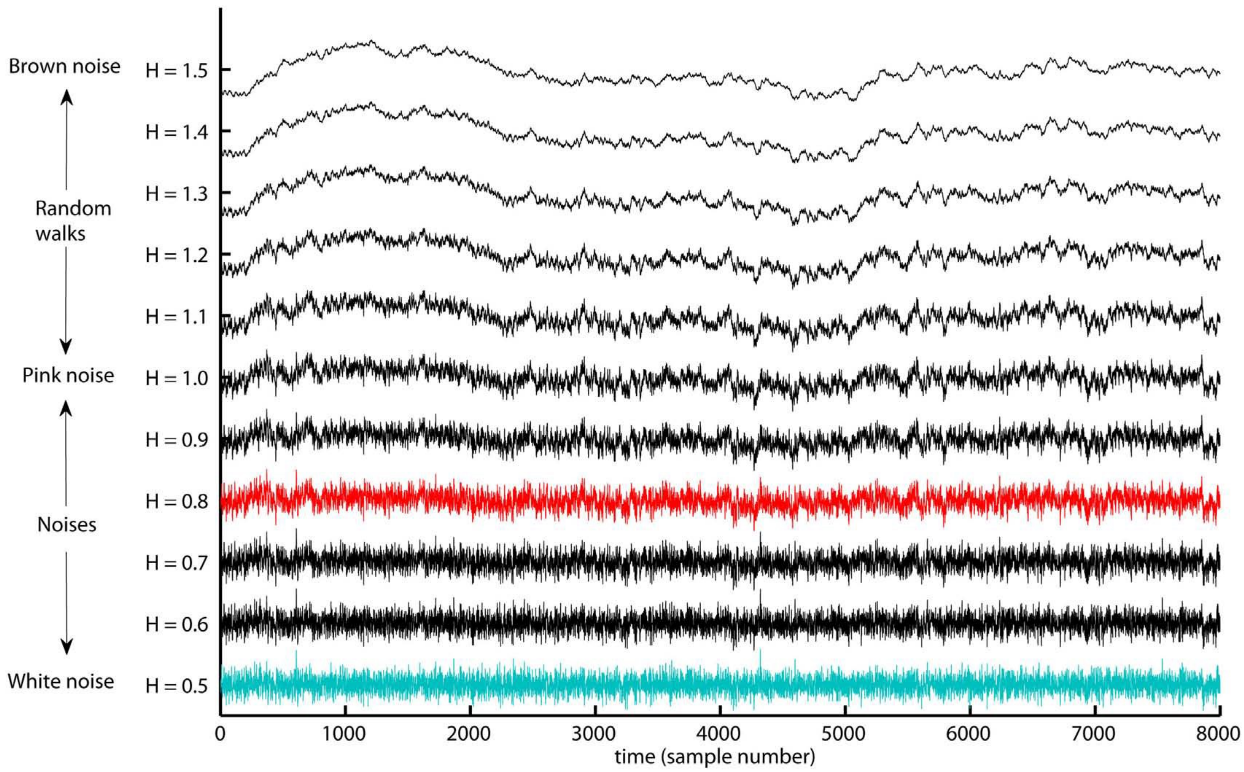


Figure 4-3 The range of Hurst exponents defines a continuum of fractal structures between white noise (H =0.5) and Brown noise (H =1.5). [14]

4.2.3 Borg's rating of perceived exertion

The Borg's RPE is a ratio scaling methods to describe how subjective intensity varies with the physical intensity in exercise, heavy physical work, and diagnostic situations. The scale is a linearly increase from 6 to 20 (total 15 scales) rating of the interaction of physical exertion and muscle fatigue. The semantic definitions and the corresponding scale number is shown in Table 4-1.

Table 4-1 The 15-grade rating and semantic definitions of Borg's RPE scale

Rating	Perceived exertion
6	No exertion at all
7	Extremely light
8	
9	Very light
10	
11	Light
12	
13	Somewhat hard
14	
15	Hard
16	
17	Very hard
18	
19	Extremely hard
20	Maximal exertion

4.3 Results

A Japanese translation of Borg's RPE proposed and tested by [3] (Scheme C) was used as a subjective fatigue measure. Experiment data of one participant was not correctly recorded and was excluded from further analysis. Ten sets of experiment data are available for analysis and all data are presented in the form of means ± standard deviation (SD) over ten participants. RR interval data series were preprocessed using the combined filter (Section 2.2) to eliminate the effect of outliers on signal quality. The threshold value t_1 was set as 30%, t_2 was set as four and window length w_m was five. Paired t test was used to check the statistical significance and p=0.05. The experiment ended after 24.6±2.5 flanges were tightened. Seven among the ten participants finished all 26 flanges without being exhausted.

4.3.1 Borg's RPE and Heart rate

HR of rest, pipeline work, and recovery baseline were 72.5±8.0, 98.0±11, and 83.7±12, respectively. The maximal rate of increase in heart rate during work condition was 62%±15%, indicating that heart rate significantly increased to adapt to workload. After cropping RR interval series into epochs of tightening every two flanges, Borg's RPE highly correlated with the decrease of RR interval. The Pearson's correlation coefficient was -0.81±0.2 and corresponds with the former study [2,3] that shows RPE scale is able to track the changes in HR. However, working performance (variance of torque) correlated neither with RPE scale nor with mean RR interval as shown in Figure 4-4.

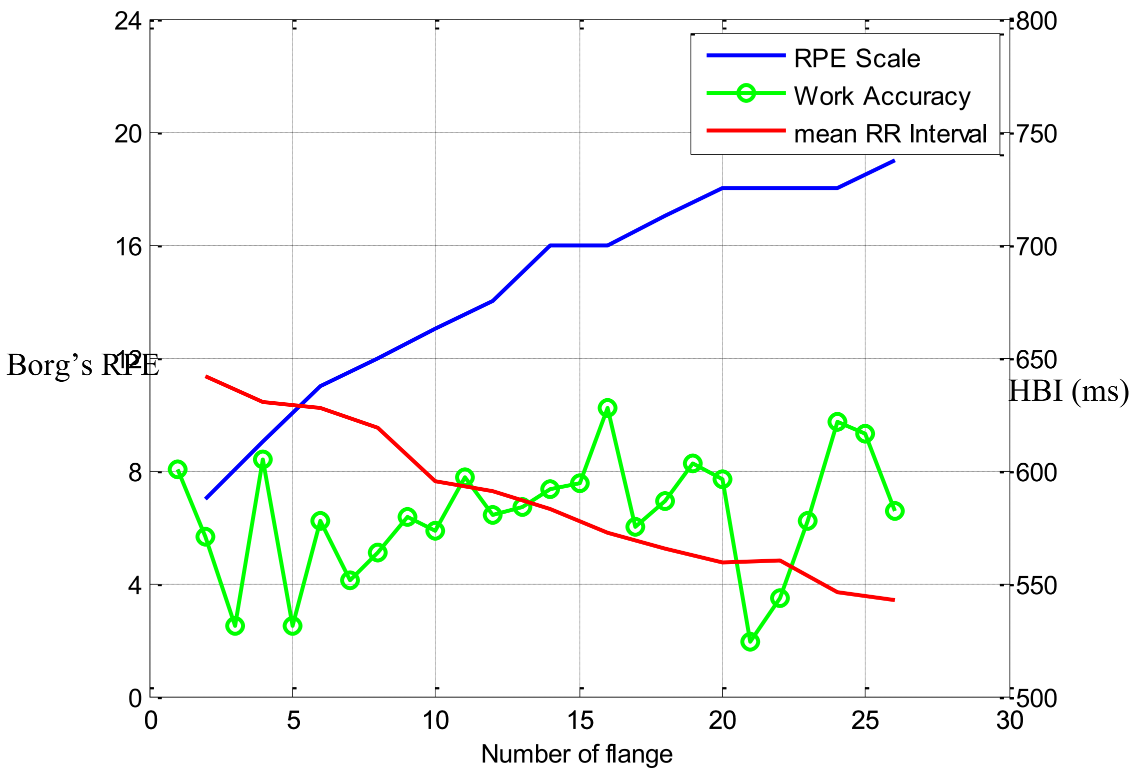


Figure 4-4 Participant 6: Negative correlation between mean RR interval and RPE scale. Green line shows SD of tightened torque of each flange and is considered as a measure of work accuracy

4.3.2 Working Accuracy and Hurst Exponent

SD of tightened torque of each flange was considered as one performance measure since evenly tightened bolts were crucial for the sealability of connecting flanges. The first and the latter half of torque variance were 5.6±2.1 Nm and 6.2±2.0 Nm respectively, indicating that the bolts were more evenly tightened in the first half when participants' degree of fatigue was lower, although the difference is not statistically significant (p=0.16) (Figure 4-5 b).

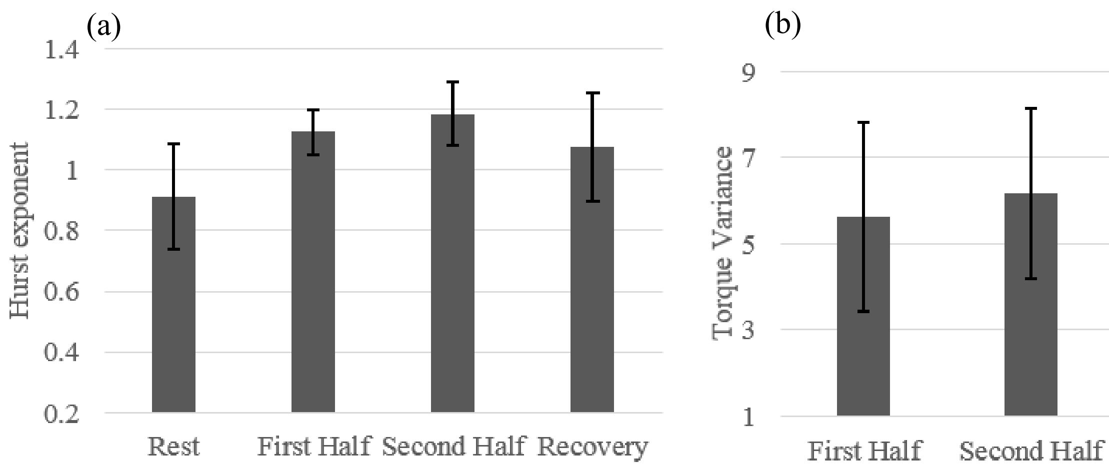


Figure 4-5 (a). HE of rest, first half of pipeline work, second half of pipeline work, and recovery condition. (b). Torque variance during first half and second half of pipeline work. Error bar is SD over all participants

To check the robustness of applying DFA to RR interval series, we randomly shuffled the RR interval series of one experiment 100 times. Randomly shuffled signals almost become white noise as h = 0.52±0.01. HE of rest, first half of pipeline work, second half of pipeline work, and recovery baseline were 0.91±0.17, 1.12±0.07, 1.19±0.10, and 1.08±0.18, respectively. HE of working condition was significantly higher than that of the rest condition (p=0.003). This result corresponds with the findings of [12], in which HE was found higher under the shooting exercise than under the rest condition. In addition, HE during the second half of pipeline work was significantly higher than the first half (p=0.04), which might be explained by the accumulation of fatigue.

Furthermore, RR interval series were cropped into segments that corresponded with the onset of tightening each flange and ended 40 seconds after the accomplishment of the tightening of each flange. HEs were then calculated as shown in Figure 4-6 and Figure 4-7. A weak positive correlation between HE and torque variance was found for five participants. The average Pearson's correlation coefficient was 0.26±0.1. This indicates that the participant's working accuracy was lower when HE was higher, although this relationship of cause-and-effect is unclear.

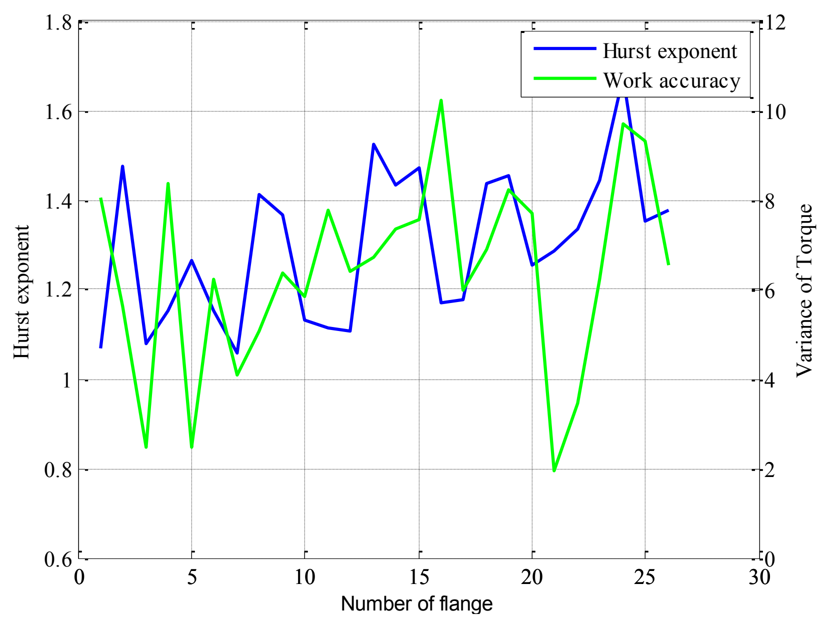


Figure 4-6 Participant 6: Correlation between Hurst exponent and work accuracy

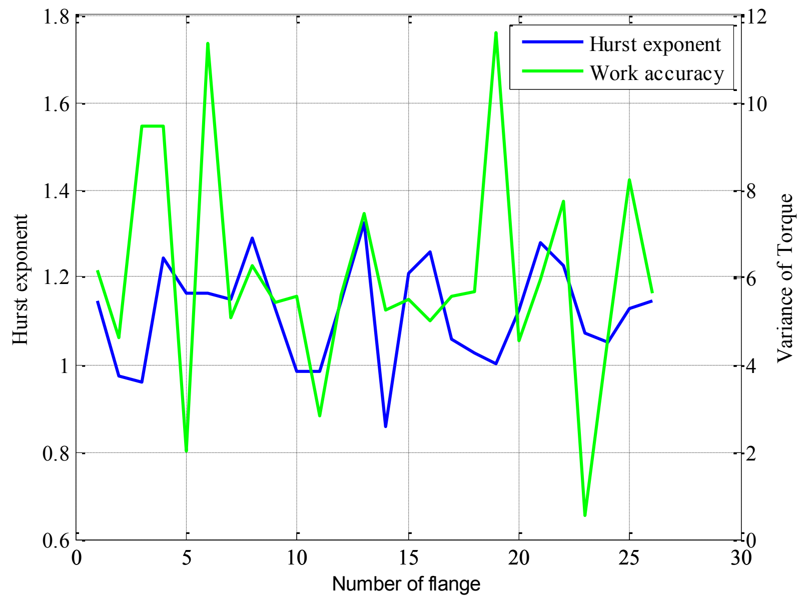


Figure 4-7 Participant 9: Correlation between Hurst exponent and work accuracy

4.3.3 Cardiac Stress Index (CSI)

Chen et al. developed a cardiac stress index to monitor human cardiac stress online during cycling exercise [6] and CSI was further proved in [7]. According to [7], CSI was defined as

$$\mathrm{CSI} = \frac{Number\ of\ events\ with\ h\ lower\ than\ 1}{Total\ number\ of\ events} \tag{4-5}$$

In [7], HE h was calculated in a one-minute sliding window with a step of 20 seconds. In this study, tightening of every flange was considered as an event and the RR interval series were cropped into corresponding segments with a delay of 40 seconds. CSI of each participant was then calculated. However, from the beginning to the end of the pipeline work, CSI showed a decreasing trend rather than an increasing trend although the RPE scale indicated an increase in subjective fatigue. This inconsistency with former study [6,7] may be caused by the following: First, the Hurst exponent extracted from DFA is affected by the setting of scale s in equation (4-2) and (4-3), especially when the signal length is different. Second, in [6,7], the experiment task was a cycling exercise that required relatively monotonous physical exertion of the lower extremities, while this paper studied a real-world task that required both complex physical exertion and mental attention.

4.4 Discussions

Early detection and avoidance of an operator's fatigue and the following performance degradation would be helpful to improve safety as well as operator's comfort. In an effort to develop a quantitative evaluation of the degree of operator fatigue in the conducting of pipeline work, an experiment was designed and conducted in this paper. Borg's RPE scale, performance measure, and RR interval series were measured and analysed. These are the results:

In the continuous pipeline work, RPE scale generally increases while RR interval decreases, and they are highly correlated (Pearson's correlation coefficient $r=-0.81\pm0.2$);

HE of working condition ($h1=1.16\pm0.08$) is significantly (paired T test, $p=0.003$) higher than baseline condition ($h2=0.91\pm0.17$), which indicates that RR interval series show more auto-correlation structures in working condition compared to rest condition;

Working performance in the first half (5.6±2.1 Nm) is better than the latter half (6.2±2.0 Nm), indicating that the bolts are more evenly tightened in the first half when participants' degree of fatigue is lower, but the difference is not statistically significant;

CSI derived from cycling exercise is not applicable to pipeline work in this study.

The main limitation of this on-going study is the small sample size and this paper fails to develop a cross-individual regression model to predict working accuracy. Another limitation is that an error of performance measure existed since the measured torque of the tightened bolts was affected by the order of measuring. We expect to solve these limitations in a future study.

References

[1] IMO.: International Maritime Organization, Guidance on Fatigue Mitigation and Management. 1. MSC/CIRC. 1014. London: IMO (2001)

[2] Borg, G. A.: Psychophysical bases of perceived exertion. Med Sci Sports Exerc, 14(5), 377-381 (1982)

[3] Onodera, K., Miyashita, M. [In Japanese]: A study on Japanese scale for rating of perceived exertion in endurance exercise. Jpn J Physio edu, 21(4), 191-203 (1976)

[4] Captur, G., Karperien, A. L., Hughes, A. D., Francis, D. P., Moon, J. C.: The fractal heart embracing mathematics in the cardiology clinic. Nat Rev Cardiol., 14(1), 56-64 (2017)

[5] Peçanha, T., Bartels, R., Brito, L. C., Paula-Ribeiro, M., Oliveira, R. S., Goldberger, J. J.: Methods of assessment of the post-exercise cardiac autonomic recovery: A methodological review. Int. J. Cardiol., 227, 795-802 (2017)

[6] Chen, S. W., Liaw, J. W., Chang, Y. J., Chuang, L. L., Chien, C. T.: Combined heart rate variability and dynamic measures for quantitatively characterizing the cardiac stress status during cycling exercise. Comput Biol Med, 63, 133-142 (2015)

[7] Chen, S. W., Liaw, J. W., Chang, Y. J., Chan, H. L., Chiu, L. Y.: A Cycling Movement Based System for Real-Time Muscle Fatigue and Cardiac Stress Monitoring and Analysis. PLoS ONE, 10(6) (2015)

[8] Peng, C. K., Havlin, S., Stanley, H. E., Goldberger, A. L.: Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos: An Interdisciplinary J Nonlinear Sci, 5(1), 82-87 (1995)

[9] Wallot, S., Fusaroli, R., Tylén, K., Jegindø, E. M.: Using complexity metrics with RR intervals and BPM heart rate measures. Front. Physiol, 4, 211 (2013)

[10] Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., Stanley, H. E.: Multifractal detrended fluctuation analysis of nonstationary time series. Physica A, 316(1), 87-114 (2002)

[11] Ihlen, E. A.: Introduction to multifractal detrended fluctuation analysis in matlab. Front. Physiol, 3(3), 141 (2012)

[12] Zhuang, J. J., Ning, X. B., He, A. J., Zou, M., Sun, B., Wu, X. H.: Alteration in scaling behavior of short-term heartbeat time series for professional shooting athletes from rest to exercise. Physica A, 387(26), 6553-6557 (2008)

Chapter 5 Estimation of mental workload in actual ship operation

In chapter 3, experiments were conducted in laboratory and simulator environment. Physiological signals were successfully measured under a relative controlled situation. However, real on-board environment is subject to many noise resources such as vibration and ship motions.

For different individual subject, reference tasks whose difficulty is manipulated in a same range and pace, may elicit different arousal level. This is partially reflected by the results in chapter 3 that subjects show different tendency of ceiling effect. In addition, task of different difficulty level can elicit similar arousal level, represented by close values of physiological features.

5.1 Experiment settings

The first engineer of the training ship 'Fukae Maru' participated in the experiment as the subject. The gross tonnage of Fukae maru is 674 ton, with a length of 49.95 meter. The main engine is a 4-stroke diesel engine, and two diesel generators and a shaft generator are installed. Fukae maru has a bow thruster and a stern thruster to increase manoeuvrability. The engine control console is installed at the bridge, but the distributor panel of electrical devices is installed inside the engine room. The subject was male, 60 years old, right handed, and had no disability. Caffeine intake and strenuous exercise were prohibited on the day of the experiment. Before the participants signed the consent form, experiment purpose, its outline and his rights were explained. The voyage was arranged for a training course of large rudder-angle steering, started from around 15:00 pm and ended at 17:00 pm. The subject conducted n-back tasks at around 13:00 pm on the ship.

Figure 5-1 Experiment apparatus, training ship fukae maru

Figure 5-2 Experiment apparatus, engine control console and distributed monitoring screen

During the leaving port and entering port condition, the subject was in charge of operating and monitoring the engine system in the engine control console area. The subject had to verbal communicate with other engineers inside the engine room, the on-duty navigation officers, and the chief engineer. He was free to walk around in the ship bridge, where the engine control console was set. Video recording and measurement of heart rate and EEG simultaneously started at 14:57 and lasted 36 minutes and 13 seconds in leaving port condition. The entering port condition started at 16:25 and lasted 18 minutes and 41 seconds. Major events of the operation are summarized in Table 5-1, where the time is presented in the format of 2-digit minutes and seconds.

Table 5-1 Major events of leaving port and entering port operations

Leaving Port		**Entering Port**	
Time	**Events**	**Time**	**Events**
0000	Start Recording	0:00	Start Recording
0022	Start operation sequence	0037	Start Diesel generator 2(D.G 2)
0109 0114	Clutch disengagement stop**	0155 0220	Dailogue
0115 0152	Test hydraulic system of controllable pitch propeller (CPP)	0206 0210	Anwser call from E.R: confirm start diesel generator
0158	Give start main engine order to E.R	0222	Button operation parallel generators
0202	Watch main engine(M.E) monitor	0253 0257	Stand up and serch for C.E
0216	Start M.E**	0342	Report to C.E: D.G 2 started
0236	Turn prime pump to auto mode	0405 0413	Dialogue with C.E
0550	Receive report: M.E normal	0414	Entering port ordered
0553 0720	Report to captain: M.E normal Walk around bridge	0424 0443	Stand up and dialogue with C.E
1153	Leaving port ordered	0526 0648	Dialogue with C.E
2313	Clutch engagement reported navigation officer (N.O)	0758 0811	Prepare for starting S.G
2324 2335	Prepare for starting shaft generator (S.G)	0812	Button operation: start S.G
2338	Start shaft generator	0843	Report to C.E: S.G normally start
2411	Report to N.O: SG start	0845	Report to N.O: S.G normal
2425	Transfer M.E control authority to Bridge, confirm from N.O	0852	Bow thruster start, reported from N.O
2439	CPP operated, reported from N.O	1420	Anwser call from E.R
2447	Start thruster, reported from N.O	1516 1518	Dialogue with C.E
2635	Leaving Port, from C.E	1531 1535	Dialogue with C.E

Table 5-1 continued in next page

Table 5-1 continued

3524	Call E.R: leaving port finish	1815	Button operation: stop S.G
3710	Bow thruster stop, reported from captain	1820	S.G clutch disengagement**
3717	Button operation: stop S.G	1828	Finish Engine
3728	S.G clutch disengagement**		

** Operated by Chief Engineer (C.E), First engineer (subject) confirm the operation

The RS800CX (Polar Electronic) and the portable EEG device (Digital Medic Inc.) were used to continuously measure the subject's RR interval series and EEG in conducting experiment tasks. Notice that there exist slight delays among the recording of EEG, HR, and video data. The delay is processed to align the calculation of physiological features in the following data-analysis.

5.2 VACP Models

Table 5-2 Checklist of VACP model

INPUT	**VISUAL**			
	No.	Description	Action Code	Weight
	V1	Visual Detection	Gaze <=2s	1
	V2	Visual Discrimination	Gaze >2s, Static, Target-oriented	2
	V3	Visual Tracking	Gaze >2s, Dynamic, Target-oriented. Moving ship	3
	V4	Visual Read, Searching, Orienting	Dynamic, Visually high attentive. Find out target	4
	AUDITORY			
	No.	Description	Action Code	Weight
	A1	Auditory Detection	Digital Signal, Sound. e.g. Telephone ring	1
	A2	Auditory Verification	Auditory Feedback. e.g. confirm order	2
	A3	Auditory Decoding	Speech, Semantic Content	3
	A4	Auditory Interpretation	Sound patterns, Auditory high attentive	4
	Auditory Disturbance		Background noise (non-directive)	each 1
	COGNITIVE			
	No.	Description	Action Code	Weight
	C1	Automatic, Alternative Selection	e.g. start one pump between two	1
	C2	Sign/Signal Recognition	e.g. Alarm information CFW pressure	2
	C3	Evaluation/Judgement (Single Aspect)	e.g. A→B, linear cause and effect relation	3
	C4	Evaluation/Judgement (Several Aspect)	e.g. A→B←C, interactive complex relationship	4
OUTPUT	**PSYCHOMOTOR**			
	No.	Description	Action Code	Weight
	P1	Discrete Actuation, Speech, Walk	e.g. Push Button, Talk, Walk, momonitoring	1
	P2	Continuous Adjusting	Unimanual, e.g. Governer operation	2
	P3	Symbolic Production	e.g. Writing, typing	3
	P4	Convergent Multiple Operations	>=2 Extremities, 1 goal	4
	P5	Divergent Multiple Operations	>=2 Extremities, >=1 goal, multi-tasking	5

Based on the multiple resource model [1], the basic idea of evaluating operator's information processing density is to divide the resource pool into four different channels: Visual, Auditory, Cognitive, Psychomotor (VACP).VACP model originates from the study of workload components in the operation of light weight helicopter [2]. They recognized four workload components: visual, auditory, cognitive, and psychomotor, and the respective overload threshold is defined. The semantic describe of the interval scale was further improved in [3]. There are also variants of VACP model adapted to apply in different context. For instance, Pfeffer et al. (2013) [4] replaced Cognitive component with a Haptic component and applied this VAHP model to the evaluation of workload in Anaesthesiology. Based on their study, a modified VACP model and semantic definitions of the weights of each component were designed as shown in Table 5-2. The weight of each channel is assigned with an orderly number according to the level of mental workload exposed to the subject. The standards of each scale are adjusted according to the characteristics of ship operation. As the engine control console is set on the ship bridge, where the captain communicate with the N.O occasionally, background noises on the bridge that affect the subject's MWL is defined as V1 (Visual channel, weight 1, same in the following). Some operations require utilization of different channels. For example, when the subject verbal confirmed orders from C.E by face-to-face, the auditory weight is A2. However, when the subject confirm information by telephone, VACP score should include a P1 since the operator have to use on hand to hold the telephone. Besides, P1 is added to VACP when the subject moved his view focus from one screen to another screen. Furthermore, notice that V4 is defined as the discrimination of high attentional auditory signal, mainly the pre-set alarm signals, which rarely happen in actual situation. The weight of cognitive item is decided based on the background knowledge. Large weight cognitive activities would happen more frequently in emergency. Since leaving port and entering port are relatively routine operation procedures, the weight of cognitive item is generally small through the experiment.

In leaving port and entering port operation, we continuously recorded the video of the first engineer and the engine control console. The video data was analyzed after experiment and VACP score were summed each for a four-second window.

The total VACP score is calculated by summing up the weight of four channels,

$$VACP = w_V + w_A + w_C + w_P$$

5.3 Evaluation of clustering quality

The results of Chapter 3 shows that different physiological features are sensitive to mental workload for different subjects. In using a combination of physiological features to evaluate MWL, the effectiveness must be quantitatively defined. In cross-task classification, training data are from standard n-back task whose difficulty is orderly manipulated. Data sets with a large number of features and a limited number of observations, usually many features are not useful for producing a desired learning result and the limited observations may lead the learning algorithm to overfit to the noise. Furthermore, a smaller number of features can also reduce computation complexity and increase comprehensibility. Thus, it is necessary to choose a proper combination of features to estimate OFS. If the data points of one difficulty level cluster more compact, then it is more accurate to use these data points as training data of the corresponded classification label. Therefore, this section proposes an I-index to evaluate the quality of clustering in reference tasks.

5.3.1 I-index

The index I is a composition of three factors, namely, 1/K, E1/EK and DK. The first factor will try to reduce index I as K is increased. The second factor consists of the ratio of E1, which is constant for a given data set, and EK, which decreases with increase in K. Hence, because of this term, index I increases as EK decreases. This, in turn, indicates that formation of more numbers of clusters, which are compact in nature, would be encouraged. Finally, the third factor, DK (which measures the maximum separation between two clusters over all possible pairs of clusters), will increase with the value of K. However, note that this value is upper bounded by the maximum separation between two points in the data set. Thus, the three factors are found to compete with and balance each other critically [5]. The power p is used to control the contrast between the different cluster configurations. In this article, we have taken p=2. The i-index is considered

$$I(\mathrm{K}) = \left(\frac{1}{K} \times \frac{E_1}{E_K} \times D_K \right)^p , \tag{5-1}$$

Where E1 is to assume that all data points belong to one single cluster, and calculated as

$$E_1 = \sum_{j=1}^{n} \left\| x_j - z_1 \right\| \tag{5-2}$$

Ek is the sum of the within cluster distance in the kth cluster,

$$E_K = \sum_{k=1}^{K} \sum_{j=1}^{n} \left\| x_j - z_k \right\| \tag{5-3}$$

Dk is the maximum distance between different clusters, calculated as

$$D_K = \max_{i,j=1}^{K} \left\| z_i - z_j \right\| \tag{5-4}$$

K is Number of clusters,

z_j is center of jth cluster,

x_j is the points in jth cluster.

5.3.2 Euclidian distance and Mahalanobis distance

Two methods are used to calculate the distance. Euclidean distance is generally defined as point-to-point distance. Euclidean distance is also known as geometrical distance. Before calculating Euclidean distance, the variables of input features must be normalized to eliminate the effect of different scales and units. In this chapter, for a single feature, the 0-1 normalization is utilized across all data included in analysis. Equation (5-5) shows the definition of 0-1 normalization, where $x_{\max}$, $x_{\min}$ is the maximum and minimum value of a series x_i , and z_i is the corresponded 0-1 normalized data. Equation (5-6) shows the definition of Euclidean distance between two data points x_i and y_i .

$$z_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{5-5}$$

$$d_E(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{5-6}$$

The Mahalanobis distance is a measure of the distance between a point P and a sample G of a distribution, introduced by Indian statistician P. C. Mahalanobis in 1936. It is a multi-dimensional generalization of the idea of measuring how many standard deviations away a point is from the mean of a sample. Equation (5-7) shows the definition of Mahalanobis distance, where μ is the means vector of variables, Σ is the covariance matrix of sample G. In calculating Mahalanobis distance, it is not necessary to normalize data series.

$$d_{M1}(x, \mathrm{G}) = \sqrt{(x-\mu)^T \Sigma^{-1} (x-\mu)} \tag{5-7}$$

Mahalanobis distance between two samples G_1, G_2 from a distribution can also be calculated as shown in (5-8), where the covariance matrix of this distribution can be estimated from the respective samples as shown in Equation (5-9).

$$d_{M2}(G_1, G_2) = \sqrt{(\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2)} \tag{5-8}$$

$$\Sigma = \frac{(n_1 - 1)\Sigma_1 + (n_2 - 1)\Sigma_2}{n_1 + n_2 - 2} \tag{5-9}$$

Where μ_1, μ_2 are the mean vector, Σ_1 and Σ_2 are the covariance matrices of G_1 and G_2. n_1 and n_2 are the respective size of sample G_1 and G_2.

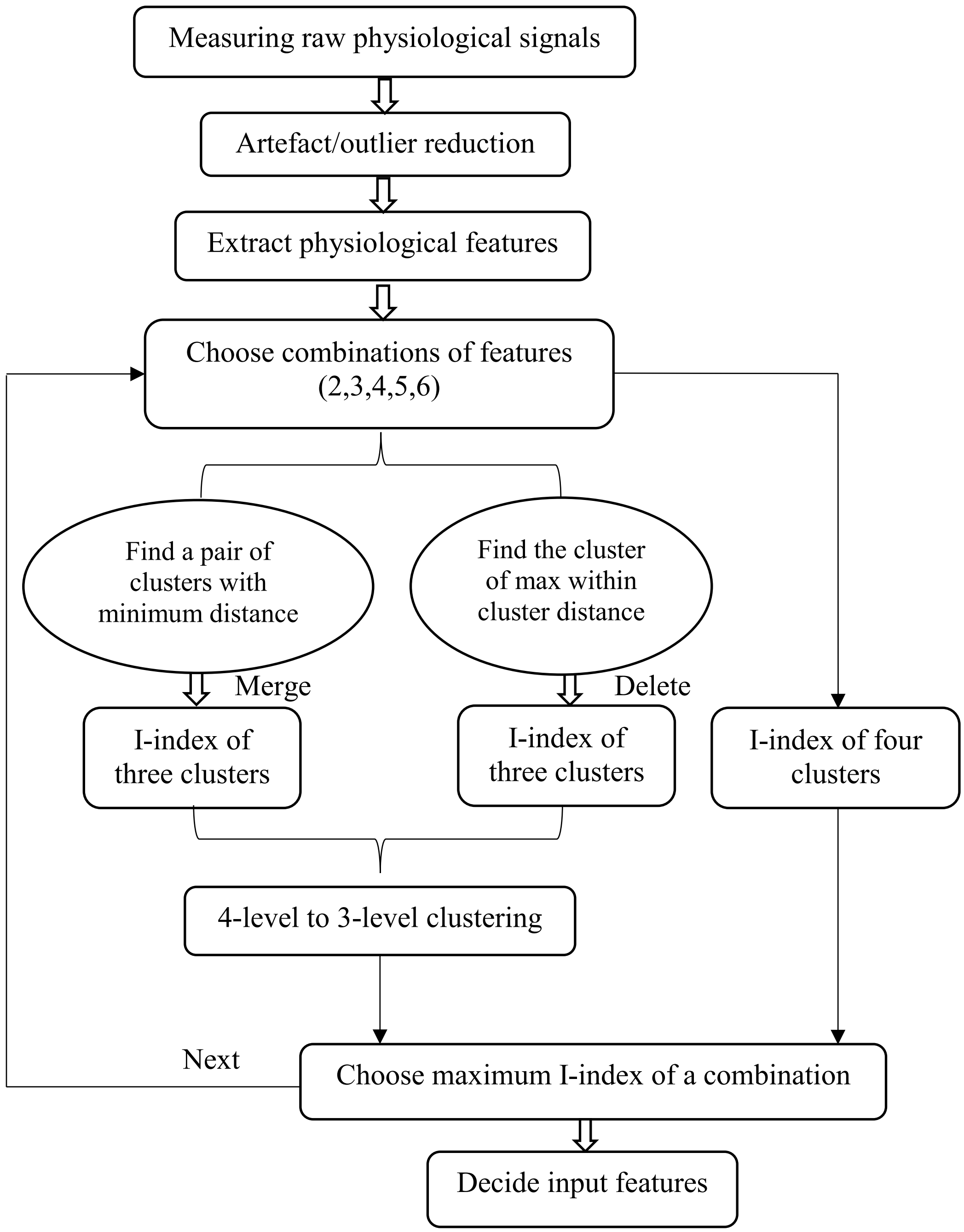


Figure 5-3 Flow chart of clustering quality analysis and feature selection

Figure 5-3 shows the flow chart of the feature selection process. After extracting the six features from the clean HBI and EEG signals, the combination consisted of 2, 3 ,4, 5, and all 6 features are formed successively. The number of combination each with 2, 3, 4, 5, 6 input features are 15, 20, 15, 6, and 1 respectively. For each combination, I-index are calculated in three conditions and the maximum I-index is chosen as the I-index of the combination. First, find out a pair of clusters with minimum distance and merge these two clusters. Second, find out the cluster that have maximum within cluster distance and delete this cluster from the 4-level clustering. Third, keep the 4-level clustering. Last, compare the I-index of each combination and choose the combination and method that make the I-index maximum.

5.3.3 N-back task clustering

Six physiological features were extracted for both n-back task, leaving port, and entering port operation tasks. Three HR based features are same to that in Chapter 4, namely MHBI, sdHBI, LF/HF ratio. MHBI and sdHBI were calculated in a non-overlapped 4-second window. To ensure the minimum length (30 seconds) for calculating frequency domain of RR interval, a moving window with 13-second advance and 13-second delay, centered at the four-second segment, was used to calculate LFHF. For EEG based features, time domain data for each difficulty level of two types of tasks were DFFT to three wave bands in a non-overlapped 4-second window. These bands were theta (4-7Hz), alpha (8-13Hz) and beta (20-25Hz) and the power spectral was estimated based on period-gram method. Alpha wave rate r_α was defined as

$$r_\alpha = \frac{P_\alpha}{P_{sum}}$$

$P_\alpha + P_\beta + P_\theta$ was the power of alpha band in one second and P_{sum} was the sum of power bands from 1Hz to 30Hz. Notice that herein P_{sum} is used as the denominator instead of $P_\alpha + P_\beta + P_\theta$ because if so one EEG feature will be linear represented by the other two EEG features in the use of $P_\alpha + P_\beta + P_\theta$. Under this circumstance, the input matrix will not be full rank, and the Mahalanobis distance will not be available. The length of each level of n-back task is 77-78 seconds, thus there are 19 data points for each level.

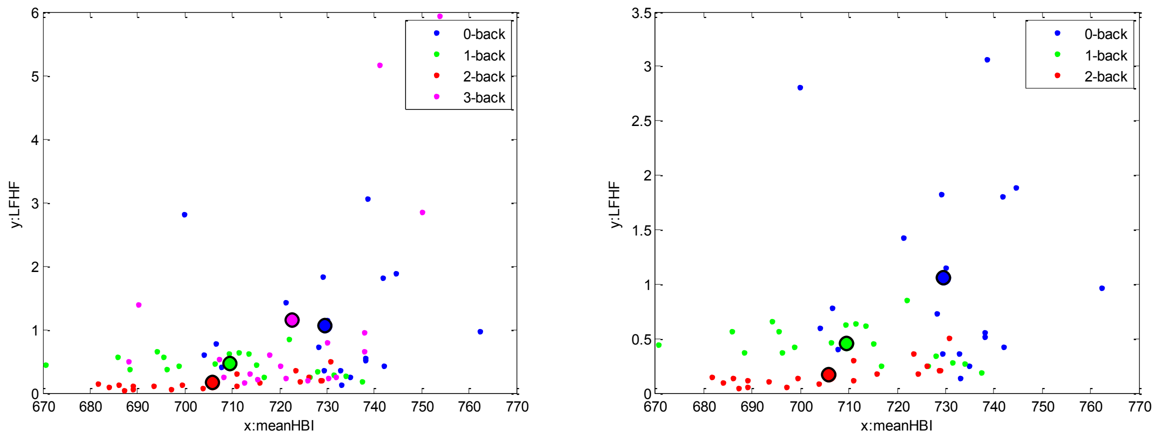


Figure 5-4 Clustering quality, using MHBI and LFHF as features

By calculating Euclidean distance, Figure 5-4 shows the clustering quality in using a combination of two features, MHBI and LFHF. The I-index of four-level clustering is 0.83 and increases to 1.54 if the cluster of 3-back task is delete from analysis. Therefore, if it is limited that only use a combination of two features as input, then the training strategy is to use 0-back, 1-back, and 2-back as low, normal, high mental workload cluster, and the input features are MHBI and LFHF(as shown in Table 5-3). The I-index of four-level clustering I_4 = 1.00. The cluster of 1-back and 2-back task is found with minimum between-cluster distance, therefore 1-back and 2-back are combined as the normal MWL cluster while 0-back is considered as the low condition and 3-back is considered as high MWL condition. The maximum I-index of merge-method is I_m = 1.18 when use the combination of [alpha wave rate, theta wave rate, MHBI, sdHBI, LFHF] as the input features. On the other hand, by delete the cluster of 3-back, the maximum I-index is I_d=1.66 when use the combination of [alpha wave rate, theta wave rate, MHBI, sdHBI, LFHF] as the input features. Since $I_d > I_m > I_4$, delete-method is used to construct the three-level training data. In detail, feature vector of [r_α , r_θ ,MHBI, sdHBI, LFHF] under 0-back, 1-back, and 2-back task are considered as low, normal, and high MWL training data set respectively.

Table 5-3 I-index results of selection process

	Combination	Merge	Delete	Four cluster
Euclidean distance	alpha, theta, MHBI, sdHBI, LFHF	1-back and 2-back 1.18	4-back 1.66	1.00
Mahalanobis distance	alpha, beta, MHBI, sdHBI, LFHF	2-back & 3-back 0.60	4-back 0.80	0.57

5.4 Estimation of mental workload

The estimation of mental workload was calculated by two methods: k-NN algorithms and Euclidean distance, and Mahalanobis distance based estimation

5.4.1 Classification based on Euclidean distance and k-NN algorithms

The k nearest neighbour (k-NN) classifier is one of the most basic pattern recognition methods, first proposed by Cover and Hart in 1967 [6]. The basic principle of k-NN is the intuitive idea that data points belong to a same cluster should be close to each other in the feature space. The input of k-NN classifier is feature vector f , corresponding to a point in the feature space. The output is classification label l , which can be dichotomous or multiple levelled. In this paper, cluster label set c =[low, normal, high]. The training data set is built as shown in Equation (5-10)

$$T = \{(f_1, l_1), (f_2, l_2), ..., (f_N, l_N)\} \tag{5-10}$$

Where N is the size of training data set. If we have a point (f_t, l_t) to be classified, find out k nearest points and denote this neighbour-space as $N_k(f)$. The estimated label is

$$l = \arg\max_{c_j} \sum_{f_j \in N_k(f)} I(l_i = c_j), i = 1,2,...,N; j = 1,2,...K \tag{5-11}$$

Where $I(l_i = c_j)$=1 when $l_i = c_j$ and $I(l_i = c_j)$=0 when $l_i \neq c_j$.

In the choice of parameter k, a smaller k would cause the increase of estimation error. In other words, the predicted label would be very sensitive to the outlier points within the neighbour-space $N_k(f)$. A bigger k would cause the increase of approximation error, namely even point far from test data would also have impact on the predicted result. The classification model would become simpler when parameter k is bigger.

To decide the choice of parameter k, ten-fold cross validation is used to check the within n-back task classification accuracy. Ten-fold cross validation is to randomly and averagely divide the sample into ten sets, nine sets of data as training data and the left one set is used as test data set. This validation process is repeated ten times to ensure all data has been used as test data. The parameter k varies from 1 to 19, and the result shows that the error rate is minimum when k=3.

The input feature vector is [r_α, r_θ,MHBI, sdHBI, LFHF], parameter k = 3, and the *knnclassify* function integrated in matlab is used as calculation algorithms.

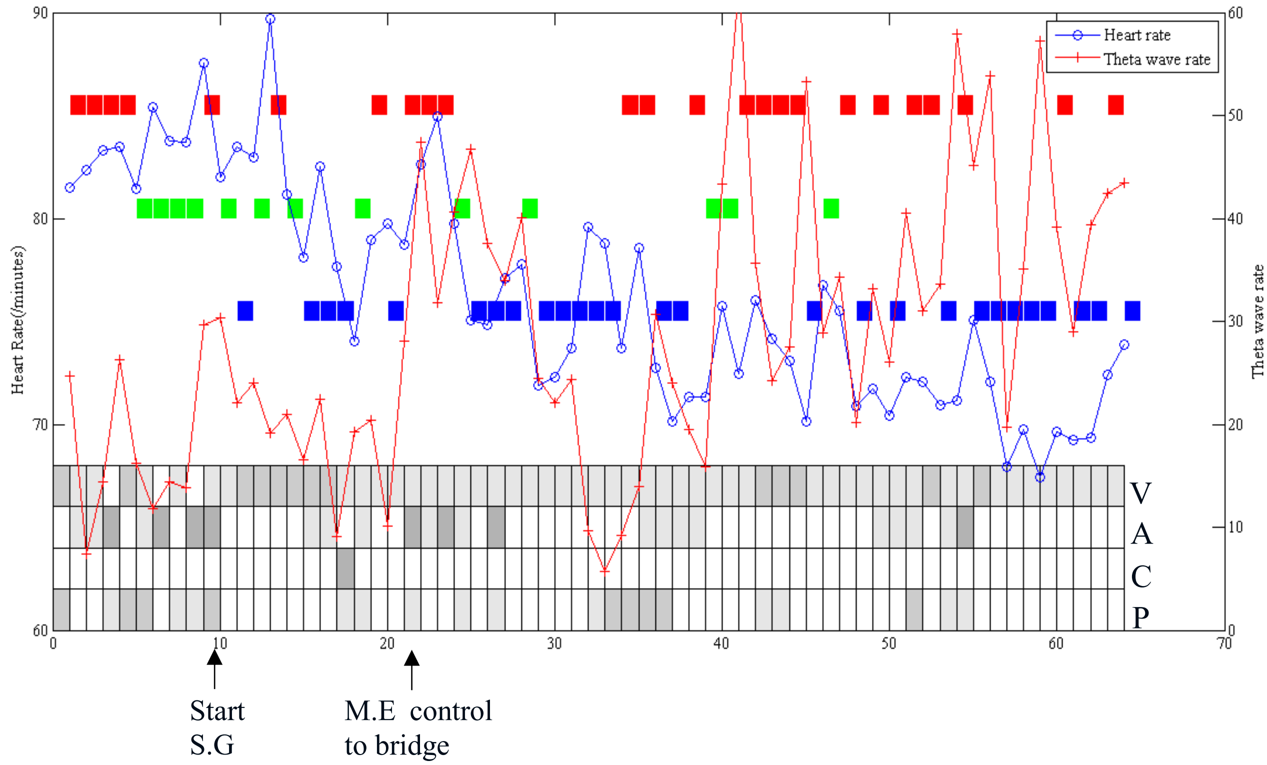


Figure 5-5 Estimated mental workload based on Euclidean distance: Leaving port

Figure 5-5 shows one analysis result of leaving port condition. The classification period is 23:00-27:16, lasted 256 seconds. By using the training data of n-back task, MWL in conducting maritime operation tasks is estimated. The red, green, and blue rectangle represent high, normal, and low MWL. The blue and red line show the instantaneous change of heart rate and theta wave rate. The rectangles with different gray scale represent different weights of VACP score. Higher gray scale represents higher weight of VACP. The red, green, and blue squares in the upper part of Figure 5-5 represents the estimated MWL is high, normal, and low respectively.

The linear correlation between estimated MWL and VACP score is checked by calculating Pearson' correlation coefficient r. A low positive correlation is found, with r=0.16.

Figure 5-6 shows one analysis result of entering port condition. The classification period is 7:00-11:16, lasted 256 seconds. No linear correlation is found between estimated MWL and VACP score.

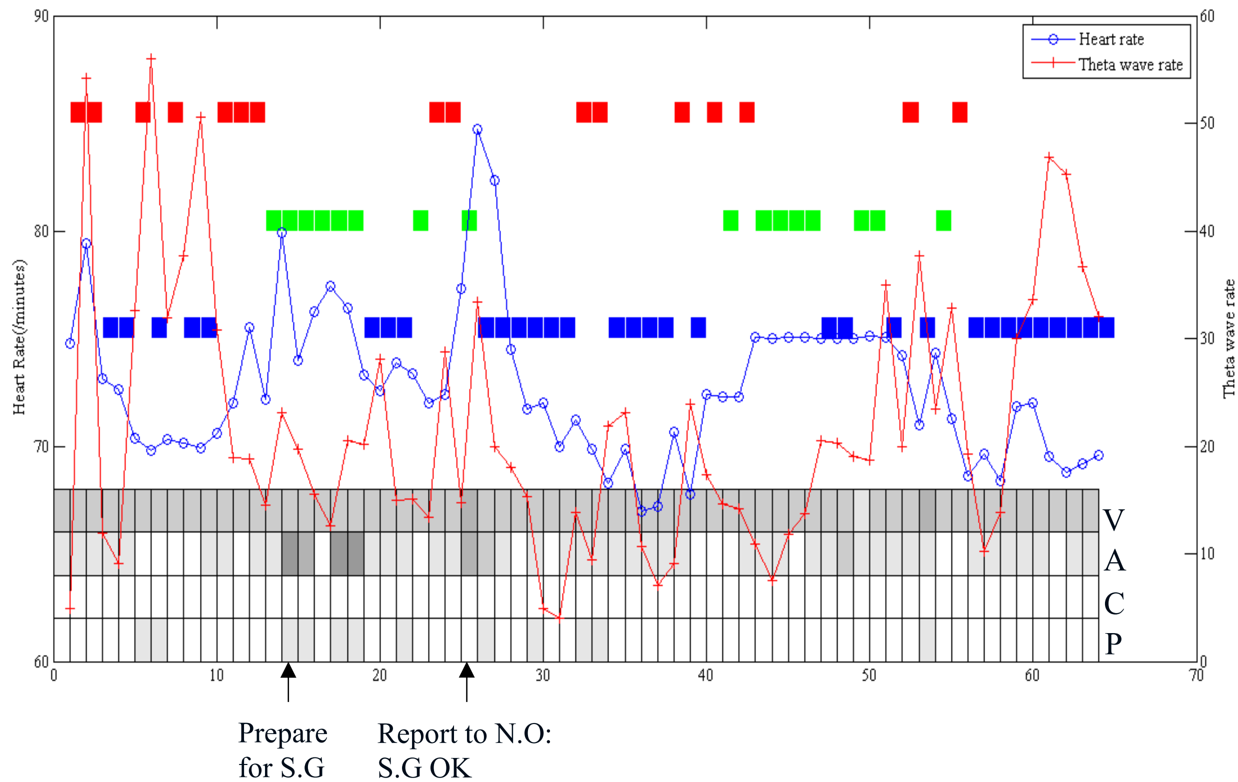


Figure 5-6 Estimated mental workload based on Euclidean distance: Entering port

5.4.2 Classification based on Mahalanobis distance

Based on the Mahalanobis distance method, the choice of significant features were chosen according to the flowchart shown in Figure 5-3. The cluster of 2-back and 3-back is found with minimum cross cluster distance, the I-index of merging 2-back and 3-back cluster is 0.60. 3-back task cluster is found to have the maximum within cluster distance. By deleting 3-back cluster, the I-index of three level clustering is 0.80. Therefore, 4-level clustering is changed to 3-level clustering by deleting the cluster of 3-back task. The feature combination is [r_α , r_β , MHBI, sdHBI, LFHF].

Mahalanobis distance based classification is to calculate the mahalanobis distance between a test data point and a sample, and assign the point to the cluster with minimum distance.

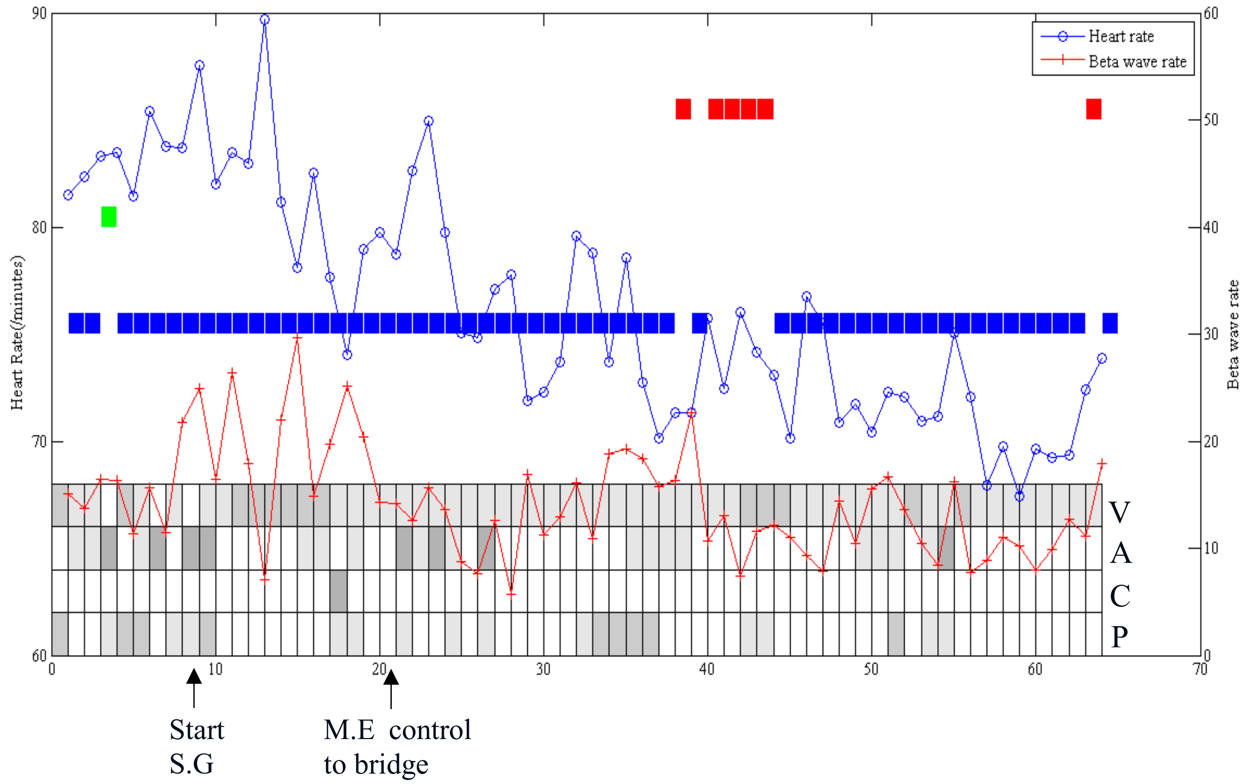


Figure 5-7 Estimated mental workload based on Mahalanobis distance: Leaving port

In leaving port condition, no significant correlation was found between VACP and mahalanobis based estimated MWL. Most of the points are assigned to low MWL situation because the covariance of 0-back cluster is obviously bigger, making the mahalanobis distance between a point and 0-back cluster smaller.

Figure 5-8 shows the mahalanobis distance based analysis of example of entering port condition. No linear correlation between VACP score and estimated MWL level is found.

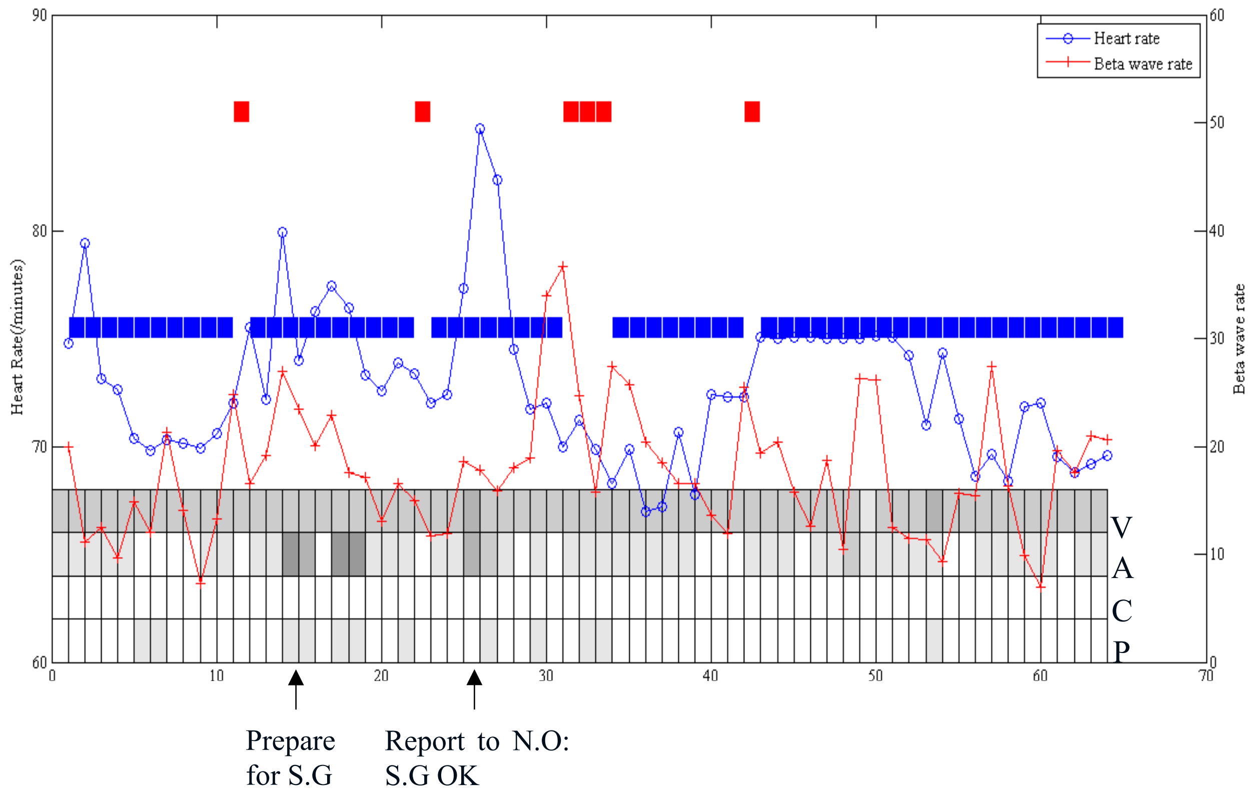


Figure 5-8 Estimated mental workload based on Mahalanobis distance: Entering port

5.5 Discussions

The subject was asked to answer NASA-TLX after each level of n-back task, leaving port, and entering port situation. From 0-back to 3-back task, the weighted NASA-TLX monotonously increases. Meanwhile, mental demand of 3-back task decreases compare to the increasing tendency from 0-back to 2-back, which corresponds with the finding of ceiling effect. The first item of the six items in the NASA-TLX scale is mental demand, representing the degree of mental and perceptual activities required in the task. Compare to the weighted TLX, mental demand is a pure measure of subjective mental stress. Therefore, the rating of mental demand is also examined in the analysis. As shown in Figure 5-9, both mental demand and weighted TLX are lower in entering port compared to leaving port, indicating that the subject felt lower MWL in entering port condition. This is corresponded with the fact that there were less events happened recognized by VACP model in entering port.

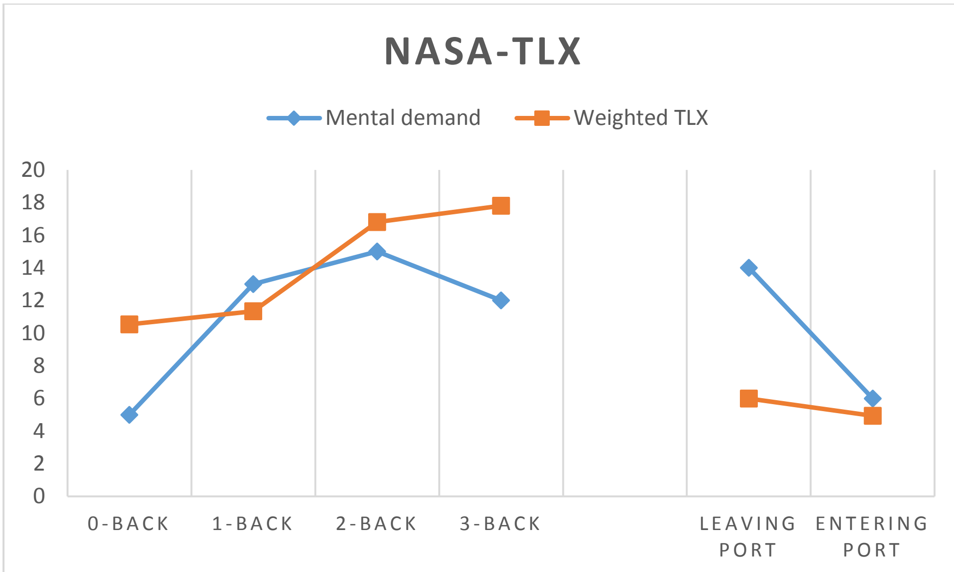


Figure 5-9 Mental demand and NASA-TLX

Based on k-NN classifier, the estimated NASA-TLX in leaving port condition is 10.4 and in entering port condition it is 9.5. Based on mahalanobis distance classifier, the estimated NASA-TLX in leaving port condition is 6.1 and in entering port condition it is 5.9. The estimated NASA-TLX corresponds with the reported NASA-TLX that the operator felt more mentally demanded in leaving port condtion than in entering port condition.

Both the subjective measure and the physiological features based estimation shows that the subject's MWL in leaving port condition is higher than in entering port condition. This might be explained by the following two reasons. First, start the engine system and maintain it work normally requires more mental resources than simply switching off the engine system. Second, after leaving port condition, the engine system continues to work and the subject has to continue his on watch duty, while the subject has free time after entering the port. This provides information that MWL should be noticed more in leaving port operation and the authority should try to reduce the quantity of work during leaving port operations and tasks with lower priority should be arranged during ocean going condition. Although it is found that leaving port condition is more demanding than entering port condition, it is also true that the subject's degree of fatigue may be higher after the voyage compared to before the voyage.

By using physiological features as the input of classification algorithms, the operator's MWL can be continuouly monitored. The operation periodes marked as high MWL are more demanding and countermeasures should be taken to avoide prolonged high MWL operations.

Reference

[1] Wickens, C. D. (2008). Multiple resources and mental workload. Human Factors: The Journal of the Human Factors and Ergonomics Society, 50(3), 449-455.

[2] McCracken, J. H., and T. B. Aldrich. Analyses of selected LHX mission functions: Implications for operator workload and system automation goals. No. ASI479-024-84. ANACAPA SCIENCES INC FORT RUCKER AL, 1984.

[3] Bierbaum, Carl R., Sandra M. Szabo, and Theodore B. Aldrich. Task analysis of the UH-60 mission and decision rules for developing a UH-60 workload prediction model: Summary report. Vol. 1. US Army Research Institute for the Behavioral and Social Sciences, 1989.

[4] Pfeffer, Stefan, Patrick Decker, Thomas Maier, and Eric Stricker. "Estimation of operator input and output workload in complex human-machine-systems for usability issues with iflow." In International Conference on Engineering Psychology and Cognitive Ergonomics, pp. 167-176. Springer Berlin Heidelberg, 2013.

[5] Maulik, U., & Bandyopadhyay, S. (2002). Performance evaluation of some clustering algorithms and validity indices. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(12), 1650-1654.

[6] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE transactions on information theory, 13(1), 21-27.

Chapter 6 Conclusions, summary, and future research

The effectiveness and safety of many real-world complex human-machine systems rely on the normal functional states of both human operators and the machinery systems, and their cooperative interactions. This paper focuses on the research of modelling and quantitatively evaluating the human operator functional state and its interaction with machinery systems. The findings of this paper provide a potential solution to reduce human error and to improve human operator's comfort in the maritime domain.

To manipulate the subjects' cognitive state, standard psychological experiment schemes (e.g. auditory/visual n-back) were developed by using E-prime Software. Simulator based operation scenario, on-board field study, and quasi real-world tasks were designed to ensure the successful manipulation of physical and mental state in various environment. In Chapter 5, human functions were decomposed into four channels: Visual, Auditory, Cognitive, and Psychomotor (VACP), whose quantitative integration was used as one of the ground truth measure. Subjective questionnaires (NASA-TLX, Borg's RPE scale etc.) were also collected as reference information.

Body movement, vehicle vibration, loose electrodes contact, and verbal communication can easily contaminate EEG and HRV signals. A combination of recursive percentage filter and median filter was used to detect and replace outliers of RR interval series. In my attempt to eliminate artifact of single-channel EEG, which is always contaminated across all relevant power bands, an accelerometer was directly attached to EEG electrodes to measure electrodes vibration. A linear model that based on calculating covariance and maximizing independence has been proved effective in reducing artifact of small amplitude across wide range of power bands (1-40 Hz). In an experimental study, the subject's EEG signal was measured in three situations, keep still, verbal speaking, and walking around. The linear model eliminates body movement artefact caused by verbal communication.

Both time-domain and frequency domain (FFT based) features are extracted and their effectiveness are firstly studied by using ANOVA, T-test, and Correlation analysis. In choosing proper combination of these features, clustering quality is determined by I-index, which is based on the ratio between cross clusters distance and within cluster distance. The distance between cluster centres are calculated by Euclidean distance and Mahalanobis distance. The best combination of features are chosen when I-index reaches its maximum value. In using RR interval data to evaluate operator fatigue, Hurst exponent is extracted by Detrended fluctuation analysis (DFA), it is found that RR interval shows more autocorrelation structures during work situation than in rest situation, representing a larger Hurst exponent in work situation. Furthermore, average RR interval highly correlates with the subjective fatigue scale, Borg's RPE scale.

With the rapid development of Interne of Things (IoT) technology, the proposal of connecting the offshore machinery systems and onshore management centre has appeared. Sensors are used to monitor the status of main engine and auxiliary machineries, and these data are transformed to onshore centre to make high-level support for shipping decisions such as whether routing and condition based maintenance. Meanwhile, the OFS evaluation methods developed in this thesis provide the potential to extend this proposal by further transforming the physiological data of offshore ship operator, which is measured by wearable sensors, to the onshore centre. The OFS of key operator can be modelled and evaluated based physiological features. As the abnormal function of key operator is the major precursor of human error accident, the onshore centre and the offshore authority thus can take measures to avoid human error from the root as long as the foreboding such as high fatigue level appears.

Using neurophysiological signals to estimate cognitive or affective states and designing applications that make use of this information require multidisciplinary expertise such as neurophysiology, machine learning, experimental psychology, and human factors. As a researcher with an engineering background, I am highly motivated to overcome the obstacles of applying this promising technology to a wide range of real-world scenarios in my future research. The basic concept of application is Operator Functional State based Adaptive Automation (OFS-AA). Adaptive automation refers to the idea of an automated system that can adapt to a changing environment, herein the operator's functional states. Based on my previous research experience, I aim to design an OFS-AA system and plan to solve the following specific problems in future research (short-term). In current lab-level studies, one single human state (i.e. workload) is usually targeted because it is easy to control and we can avoid other factors that may affect neurophysiological signals. However, according to my previous research, when task demand exceeds the capacity of mental resources, the operator will tend to fall behind with the pace of the task and show a lower activation level, and this is generally associated with a degree of performance degradation. The neurophysiological features during lower activation caused by over-load or fatigue are mixed with the features extracted from proper lower workload situations. Features of high diagnostic quality, namely the ability of one feature (i.e. Alpha band) to infer a specific cognitive state (i.e. boring) and remain unaffected by related factors (i.e. fatigue), should be developed. From the viewpoint of data driven methods, we can directly delete the cluster showing ceiling effect by calculating maximum within-cluster distance. From the viewpoint go mechanism study, it is promising to build a trigger mechanism by studying the transitional variance of activation in different brain areas (i.e. frontal to post cerebral). Physiological activity and its effect on measured electrical signals are affected by the complex interaction between external stimulus and internal physiological compensation mechanisms. Regarding the process as a black box, this paper used data-driven methods to check the temporal delay between stimulus and signal fluctuation. For example, by using RR interval series to evaluate fatigue, and setting the time lag as [-10:5:40] seconds, goal-driven optimization objective functions were defined. However, the solutions oscillated greatly under situations with different evaluation goals. an experiment that is more elaborate and utilize methods that are more complicated (i.e. Granger Causal Relation Test) should be designed to unravel the causal inference issue, thus to improve prediction accuracy. In detail, the findings of this paper would become more applicable to real world scenarios if the following aspects can be improved. First, in EEG measurements, eliminating artefacts caused by body movement is still a troublesome issue. According to the finds in this paper, I plan to focus on developing decomposition methods that utilize information of electrode vibration measured by accelerometer. Second, wearable sensors are rapidly being improved; a possible future direction may be to integrate these sensors into objects of daily use, e.g. EEG measuring safety helmet, heart rate monitoring undergarments, and view tracking glasses. Third, gain high-quality training data by having subjects conduct standard psychological task schemes that are able to elicit physiological responses close to real-world tasks.

Appendix

Original NASA Task Load Index and Japanese translation

Title 項目	Description 説明
Mental Demand 知的・知覚要求	How much mental and perceptual activity was required (e.g. Thinking, deciding, calculating, remembering, looking, searching.)? Was the task easy or demanding, simple or complex, exacting or forgiving どの程度の知的・知覚的活動を必要とするか. （例えば，思考する/判断する/計算する/思い出す/見る/検索する） 課題は，簡単/厳しい．単純/複雑．正確さが求められる/大雑把でよかったか.
Physical Demand 身体的要求	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious? どの程度の身体活動を必要としたか。（例えば、押す/引く/回す/制御する/活性する） 仕事は，簡単/厳しい，ゆっくり/きびきび，緩い/きつい，安らか/骨の折れる，でしたか.
Temporal Demand タイムプレッシャー	How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic? 仕事のペースまたは課題が発生する頻度のために感じる時間的な切迫感がどの程度か. ペースは，ゆっくりで余裕のあるもの/速くて余裕のないもの，でしたか.
Performance 作業成績	How successful do you think you were in accomplishing the goals of the task set by the experimenter? How satisfied were you with your performance in accomplishing these goals? 実験指示者によって設定された課題の目標をどのくらい到達していると考えるか. 目標の達成に関して自分の作業成績にどの程度満足しているか.
Effort 労力	How hard did you have to work (mentally and physically) to accomplish your level of performance? 作業成績のレベルを達成・維持するために，精神的・身体的にどの程度いっしょうけんめいに作業しなければならないか.
Frustration Level フラストレーション	How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task? 作業中に，不安，落胆，イライラ，ストレス，悩みをどの程度感じたか。あるいは逆に安心感，満足感，楽しさ，リラックスをどの程度感じるか.

9/14(水)
13⁰⁰〜

●以下の設問について，適切なレベルに○を記して下さい．

・Mental Demand (MD)　How mentally demanding was the task?
知的/知覚的な要求(MD)　課題の知的・知覚的な要求値はどのくらいですか．

Very Low　Very High

・Physical Demand (PD)　How physically demanding was the task?
身体的な要求(PD)　作業の身体的な要求値はどのくらいですか．

Very Low　Very High

・Temporal Demand (TD)　How hurried or rushed was the pace of the task?
時間的な要求(TD)　仕事のペースの 慌ただしさ/忙しさ はどのくらいですか．

Very Low　Very High

・Performance (OP)　How successful were you in accomplishing what you were asked to do?
成績(OP)　依頼された仕事に関して，自分の作業成績にどの程度満足していますか．

Very Poor　Very Good

・Effort (EF)　How hard did you have to work to accomplish your level of performance?
努力(EF)　あなたの成績レベルに到達するための努力は，どのくらい困難なものでしたか．

Very Low　Very High

・Frustration (FR)　How insecure, discouraged, irritated, stressed, and annoyed were you?
不満(FR)　作業中に不安感，落胆したり，イライラしたり，緊張や悩んだり，どの程度感じましたか．

Very Low　Very High

●Select the one of each pair that provided the most significant source of workload variation in these tasks.
これらの課題において，以下に挙げたペアから、作業負荷の変動に最も有意な要素を，どちらか１つ選んで下さい．

知覚/身体　知覚/時間　知覚/成績　知覚/不満　知覚/労力
~~身体~~/時間　身体/成績　身体/不満　身体/労力　時間/成績
時間/不満　時間/労力　成績/不満　成績/労力　労力/不満

Reference: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (OP), Effort (EF), Frustration (FR)
参　考：知的要求量(MD), 身体的要求量(PD), 時間的要求量(TD), 成績(OP), 労力(EF), 不満度(FR)

Borg's Rating of Perceived Exertion and Japanese translation

[疲労評価の試み]

作業における疲労の程度について、次の疲労評価にご協力をお願い致します。

Q. 次の 6 から 20 までの 15 段階から、疲労程度を、主観的に評価し、該当箇所へ"○"を記して下さい。(数字の中間はありません。)

フランジ 2 まで

20	19	18	17	16	15	14	13	12	11	10	9	8	7	6
	非常にきつい		かなりきつい		きつい		ややきつい		楽である		かなり楽である		非常に楽である	

フランジ 4 まで

20	19	18	17	16	15	14	13	12	11	10	9	8	7	6
	非常にきつい		かなりきつい		きつい		ややきつい		楽である		かなり楽である		非常に楽である	

フランジ 6 まで

20	19	18	17	16	15	14	13	12	11	10	9	8	7	6
	非常にきつい		かなりきつい		きつい		ややきつい		楽である		かなり楽である		非常に楽である	

フランジ 8 まで

20	19	18	17	16	15	14	13	12	11	10	9	8	7	6
	非常にきつい		かなりきつい		きつい		ややきつい		楽である		かなり楽である		非常に楽である	

フランジ 10 まで

20	19	18	17	16	15	14	13	12	11	10	9	8	7	6
	非常にきつい		かなりきつい		きつい		ややきつい		楽である		かなり楽である		非常に楽である	

お疲れ様でした。ご協力ありがとうございます。

船舶機関管理学研究室

Research Participation Consent Form (In Japanese)

研究参加に関する同意書

　このたび、MEPS（Marine Engine Plant Simulator）を使用した人間反応計測に関する研究に参加するに当たり、担当者より以下の項目につきまして、十分に説明を受けました。

① 研究の目的。（別紙）

② 実験の方法。(別紙)

③ 実験への参加は協力者の自由意思によるものであり、研究への参加を随時拒否、撤回出来ること。また、これによって協力者が不利な扱いを受けないこと。

④ データの管理には細心の注意を払うこと。

⑤ 結果の公表の仕方について。また、結果が公表される場合であっても、協力者のプライバシーは確保されること。

⑥ 研究責任者の氏名、職名、連絡先。

本実験は人間反応を計測するためのもので、個人の能力を測るためのものではございません。上記の内容を十分に理解し、承知した上で、自ら本研究に参加することに同意します。

説明日：　　　　　年　　　月　　　日

説明者：

説明者所属：

同意年月日：　　　　　年　　　月　　　日

研究協力者：

本同意書は、研究協力者と研究責任者が一部ずつ保管する。

ACKNOWLEDGEMENT

In 2012, I graduated from Dalian Maritime University and was offered a Government scholarship by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT), under the International Priority Graduate Program: Leadership Development Program for Maritime Sciences in Asia. First of all, I would like to express my deepest gratitude to the MEXT, to the Graduate School of Maritime Sciences of Kobe University, for providing me the opportunity and the financial support to pursue a Doctoral degree in Maritime Sciences.

I am grateful to my supervisor, Professor Makoto Uchida. I would like to thank you very much for your support and understanding over these past five years.

I would also like to show my sincere gratitude to the kind lecturer, Dr. Miwa and the technical staff Mr. Aburagi, you help me a lot when dealing with the problems in conducting experiments, especially for kindly helping me translate English documents, including NASA-TLX, into Japanese. Thank you very much.

In March 2012, I came to Propulsion laboratory in Kobe University and began to work with the colleagues in the propulsion-engineering laboratory. We did many experiments together. I would like to express my gratitude to all of you. Thank you very much for your kind cooperation and understanding.

Most importantly, none of this could have happened without my family. Your understanding and encouragement is the biggest resource for me to fulfil my doctoral study.