Fan, S, Zhang, J, Blanco-Davis, E, Yang, Z, Wang, J and Yan, X

Effects of seafarers' emotion on human performance using bridge simulation

Article

Fan, S, Zhang, J, Blanco-Davis, E, Yang, Z, Wang, J and Yan, X (2018) Effects of seafarers' emotion on human performance using bridge simulation. Ocean Engineering, 170. pp. 111-119. ISSN 0029-8018

EEG-based emotion recognition for seafarers using bridge simulation

Shiqi Fan[a,b,c], Jinfen Zhang[a,b,*], Eduardo Blanco-Davis[c], Zaili Yang[c], Jin Wang[c], Xinping Yan[a,b]

[a] Intelligent Transport Systems Research Centre (ITSC), Wuhan University of Technology, China

[b] National Engineering Research Centre for Water Transport Safety (WTSC), Wuhan, China

[c] Liverpool Logistics, Offshore and Marine (LOOM) Research Institute, Liverpool John Moores University, Liverpool, UK

ABSTRACT: 75-96% of maritime accidents are caused by human and organisational factors. Seafarers' emotion may degrade the effectivity of human behaviour when tasks in onboard environment are complex and demanding. This study was concerned with the relationship between seafarers' emotion and occurring events in navigation. The Electroencephalogram (EEG) and Self-Assessment Manikin (SAM) scale rating are used to investigate the occurrence and impact of seafarers' emotions on their performance using a bridge simulator. The study was conducted and described in two sections: emotion calibration and test recognition. In the first section, two types of emotions are induced by the sound clips of the International Affective Digitized Sounds (IADS), developed by the National Institute of Mental Health Center for the Study of Human Emotions. In the second section, emotion is recognised by the Support Vector Machine (SVM) classifier, as well as self-rated after the crew-qualified test in a bridge simulator. The results indicate that SVM can identify the emotions by EEG feature extraction, with an accuracy of 77.55%. The results concerning officers' emotion in a bridge simulator test reveal that seafarers' emotion in maritime operations, relating to events exposure, affects their behaviour and decision-making. In addition, negative emotion has a higher likelihood of contributing to human errors than positive emotion. Less negative emotion is the most dangerous emotion state during navigation, followed by extreme positive emotion.

KEYWORDS: Human errors, Bridge simulation, Maritime operations, Emotion

1 INTRODUCTION

The ship operation system is a system based on people behaviour, and about 75-96% of marine accidents are caused, at least in part, by human errors (Hanzu-Pazara *et al.*, 2008). The activities onboard or off-board related to seafarers or mariners are influenced by internal and external factors. A study that analysed the specific onboard duties and off-board entities involving Greek-flagged ships, during 1993–2006, found that 57.1% of all accidents were attributed to the human element (Tzannatos, 2010). Among them, 75.8% were detected onboard and 80.4% of the onboard human-induced accidents were related to errors and violations of the ship's master. As the ship's master is responsible for onboard decisions making, it was evident that the master's errors or violations would affect other crews' working procedures, manoeuvring behaviours, and emergency responses. However, problems in international maritime training became obvious, that is the experiential learning gap of entry-level officers or "lost apprenticeship". In addition, the declining of experienced crewmembers and the pressure of fast promotion into responsible positions increased the "experiential learning gap' of officers (Hanzu-Pazara *et al.*, 2008). Therefore, human errors existing within maritime operations are complicated and worth being further investigated.

In this regard, it is meaningful to investigate human factors in a ship bridge from an operational perspective, as it is closer to the root causes of maritime accidents. One of the earliest initiatives was fired up by accidents caused by a typical radar-assisted collision (Grech *et al.*, 2008). In 1956, the collision between the two passenger ships Andrea Doria and the Stockholm was one illustrative example. The root causes of the accident were related to the ship bridge. It was demonstrated that more attention should be paid to human factors and the bridge. Consequently, it caused some interest in the area of bridge design and cognition. Nowadays, the bridge has become more automated. Automation is often highlighted because it has been overwhelmingly understood that it would reduce the involvement of crew, so as to reduce human-related problems, and increase safety and efficiency. However, as demonstrated by the grounding of the Royal Majesty (the Panamanian passenger ship, which grounded on the Rose and Crown Shoal, 10 miles to the east of Nantucket Island, Massachusetts on June 10, 1995), as well as evidenced by other research findings (Lutzhoft and Dekker, 2002), automation has a prospecting expectation of human work which cannot be simply replaced completely. There is no evidence that fewer crew members lead to less individual mistakes in bridge. As increased mental workload onboard affecting situation awareness (Aguiar *et al.*, 2015), emotion, as an individual factors, in bridge operations might contribute to human behaviours in accident chains. In this regard, automation in the bridge creates new error pathways, especially resulting from human errors, deficiencies in mission shifts, and postponed chances to correct errors further into the future in the system. It is noteworthy that bridge operations plays an essential role in the success or failure of navigation.

The machine learning technology and signal processing method have been developing rapidly given the mature of physiological equipment and device to obtain objective data. The investigations of human factors based on physiological data have become an emerging subject. The main contents of human factors in the maritime sector usually compose the following aspects: mental workload, emotion, attention, pressure, and fatigue (Hou *et al.*, 2016, Fan *et al.*, 2017). The emotion factor of the crew is sensitive to tight working spaces, inaccessible information sources, and the single gender in some countries. Roidl *et al.* (2014) pointed out that behavioural patterns, *e.g.* aggressive driving and delayed reactions, could be influenced by strong emotions in the driver. For example, anger leads to stronger acceleration and higher speeds even beyond the emotion-eliciting event. In addition, anxiety and contempt had weaker effects, showed the same negative driving pattern as anger. Fright was related to stronger braking momentum and lower speeds. Moreover, the negative emotions are also related to irritability, tension, instability, depression and burnout with periodic changes (Lafont *et al.*, 2018, Scott-Parker, 2017, Liu and Sourina, 2014). Fairclough *et al.* (2014) found that cardiovascular reactivity to negative mood may be affected by the emotional properties of music in simulated driving. Therefore, studying the emotion associated with accidents would benefit the crew training in navigation and improvement of the watch-keeping operations.

In this paper, the approach to the identification of seafarers' emotion during operations is studied, using a bridge simulator and the EEG device. Based on this, the relationship between operators'' emotion and their performance is investigated. The remainder of paper is organised as follows. In Section 2, the literature review of the relevant studies is presented. The experiment design with the detailed procedures and method is described in Section 3. The results are illustrated in Section 4, including the feature extraction of EEG data, emotion classification, and relationship between emotion and events. The discussions are presented in Section 5. Finally, the conclusion is given in Section 6.

2 LITERATURE REVIEW

2.1 *Human errors in maritime operations*

In the amendments of Seafarers' Training, Certification and Watchkeeping (STCW) Code in 1995, human error was classified into three major taxonomies: operational-based, management-based, and the combination of the two. Human Reliability Analysis (HRA) is one of the most widely used methods, which focuses on the quantification of human operations (Precondition of human and contexts error). HRA is developed from engineering risk analysis aiming to predict likely failure event sequences quantitatively, and analyses human factors in maritime accidents. Error frequency and expert opinion are used to predict the underlying reasons (Kirwan, 1994).

At the early stage of modelling human errors, some studies tried to assign a probability to the failure of a human operator in performing tasks (Zio, 2009), including the Technique for Human Error Rate Prediction (THERP) (Swain and Guttmann, 1983), Accident Sequence Evaluation Program (ASEP) (Swain, 1987) and Human Cognition Reliability (HCR) (Hannaman *et al.*, 1985). However, neither of these studies went beyond individual human errors by considering personnel, situational or organisational factors. Consequently, HRA has been further developed. First, the situational influence on human errors with local conditions and task-specific factors is taken into account to categorize errors, including the Cognitive Reliability and Error Analysis Method (CREAM) (Hollnagel, 1998). Secondly, A Technique for Human Error Analysis (ATHEANA) (Cooper *et al.*, 1996) tried to model the relationship between the context and the probability of a human failure (Zio, 2009). In this way, cognitive failures are traced back to the psychological and situational precursors with relatively less organisational conditions involved.

In more recent research, Celik and Cebi (2009) applied a Human Factors Analysis and Classification System (HFACS) initially from the aviation transportation (Wiegmann and Shappell, 2017) to identify human errors in shipping accidents using a Fuzzy Analytical Hierarchy Process (FAHP). In line with HFACS, as well as Reason's Swiss Cheese Model and Hawkins' SHEL (Software, Hardware, Environment, Liveware) model, Chen *et al.* (2013) proposed HFACS for a Maritime Accidents (HFACS-MA) model to measure the Human and Organisational Factors (HOFs). Studies on the estimation of human failure probabilities include Yang *et al.* (2013), Yoshimura *et al.* (2015), and Yang and Wang (2012). Soner *et al.* (2015) combined Fuzzy Cognitive Mapping (FCM) and HFACS to develop onboard fire prevention modelling for ships. Akyuz and Celik (2015) adopted CREAM to assess human reliability under a cargo loading process. Akhtar and Utne (2015) investigated the common patterns of interlinked fatigue factors. It was illustrated that “inattention”, “inadequate procedures”, “observation missed”, and “communication failure” were related to fatigue factors that influence the human cognitive processes in accidents. Moreover, Hetherington *et al.* (2006) divided human factors into fatigue, stress, health, situation awareness, teamwork, decision-making, communication, automation, and safety cultural diversity.

2.2 *Seafarers' emotion identification*

The investigation on historical data (Barsan *et al.*, 2007, Luo and Shin, 2016) is one of the most popular approaches to identify the causes of maritime accidents. Most of such studies are unable to measure the specific factor changing, especially the quantitative data of psychological and physiological characteristics of the human. Relevant studies (Xi *et al.*, 2017, Akyuz and Celik, 2014, Chen *et al.*, 2013) focus on the concepts of HOFs, HRA, and human errors, human failure, etc. Physiological signals (Hou *et al.*, 2016) are collected to quantify human factors using sensors like Electroencephalograph (EEG), Electrocardiograph (ECG), Electromyography (EMG), blood volume pulse, skin electrical response, and eye movement. Moreover, other studies on angry driving in road transportation (Yan *et al.*, 2015, Zhang *et al.*, 2014, Lafont *et al.*, 2018) have been conducted to find the emotional connection between drivers and behaviours.

The emotion factor of the seafarers in watchkeeping is relevant to working space conditions, inaccessible information sources, and communication. Although there are some studies focused on the road or railway (Lucidi *et al.*, 2010, Read *et al.*, 2012, Morales *et al.*, 2017, Scott-Parker, 2017, Zimasa *et al.*, 2017) emotional factors and human errors quantification, relatively rare researchers study this in maritime operations. In order to identify the negative emotions, Liu and Sourina (2014) started to use an EEG (Electroencephalogram) system in bridge simulators to monitor officers' workload and pressure. It was one of the earliest studies on seafarer's psychological response using bridge simulators. However, the relationship between psychological response and seafarers' performance was not fully demonstrated. For the quantification of crew emotion, a system took into account monitoring emotion, emotional stress, and environmental stress (Liu *et al.*, 2016). It identified the emotion (three-dimensional description) of cadets in the bridge simulator by extracting features of EEG data, but not related to human errors yet. The researchers found that activity of emotional states was localized in relatively non-overlapping brain regions, spanning cortical and subcortical areas (Kragel and LaBar, 2016). The ventral striatum activities are associated with music evoking joy and happiness (Menon and Levitin, 2005), whereas sad music activates the hippocampus, amygdala, and neighbouring medial temporal lobe areas that distinct negative affective states and anxiety (Mitterschiffthaler *et al.*, 2007). Geethanjali *et al.* (2017) detected and recognised human emotion using SAM rating by pleasure, arousal, and dominance. The statistical analysis revealed the emotion identification differences between several groups. Hence, seafarers' emotion identification can be further studied by better incorporating psychological knowledge.

In summary, it is imperative to study the influence of seafarers' emotion in maritime from the perspective of physiological behaviour of seafarers, which is of great significance for identifying the leading causations of human errors and direct causes of accidents. This study is conducted to identify the emotion in the bridge using EEG, and to classify the emotion in a SVM model by use of bridge simulators.

3 MATERIAL AND METHOD

3.1 *Test subject selection*

Seafarers from different companies who were taking the captain and first officer qualification examinations were recruited to be involved in the experiments. There were 11 exams scheduled in two days. Each exam tested one participant who acted as a captain in a four-person exam group. All the test subjects were in good health without head injuries. They had 7.7 years of experience at sea on average, as they presented a typical emotional response during sailing when compared to beginners or cadets. The test subjects ranged from 26-38 years old, with the average of 31.9 years old. They were all males. These seafarers attended the experiments as volunteers. They were also informed that they could quit the experiments whenever they changed their minds. Based on this agreement, the calibration part of this study was conducted before the crew-qualified exam, and the test part was carried out after the whole exam. The test subjects were operating in a bridge simulator room (Figure 1a), while the staffs were in a separate control room (Figure 1b) providing scenarios to subjects.

(a) Test subjects in simulator room

(b) Staff in control room

Figure 1. The test subjects and staff in control room

3.2 *Stimuli selection*

The role of "captain" in the four seafarers during the exam was selected as an independent sample. The rating of their perceived emotion for each stimulus presented uses a SAM scale. In view of this, International Affective Digitized Sounds (IADS) database, developed by the National Institute of Mental Health Center for the Study of Human Emotion, was used as the stimulus with two categories (pleasant and unpleasant). It was presented to them for the first time, and all the test subjects in this study were not aware of the clips prior to the experiment, and may reflect facial avoiding effects on the subjective rating from the questionnaire.

3.3 *Experiment device*

This study utilised a low-cost wireless EEG headset – NeuroSky Mindwave to collect the brain wave signals of test subjects. NeuroSky Mindwave is a general public single-channel (electrode) device, with dry active sensor technology that eliminates the use of gel for electrode placement.

The test subjects were not allowed or willing to use the gel of normal EEG devices in this qualified test. The mobility during their test was highly required so that the wireless device was preferred. For this reason, NeuroSky Mindwave, a wireless single-channel (electrode) EEG headset, was selected to use in this study.

3.4 *Experimental protocol*

The experiment was conducted by EEG technology and SAM scale rating questionnaires received separately within two sections, which are emotion calibration and recognition respectively. In calibration, two types of emotions were induced by the IADS methodology. Every test subject was given by listening to sound clips from IADS with eyes closed in case of blink interrupts. In this section, emotion 1 began with 5 seconds silence to calm down, and 10 seconds for one category of emotion stimulus, and then the SAM rating was carried out. After that, another category of emotion 2 was repeated. The objective of doing this is to calibrate emotion of each subject. In other words, the specific feature or standard of personal emotion type was obtained.

In the test part, the subjects filled the questionnaires after at least 30 minutes' exam in the bridge simulator. Figure 2 demonstrates the process of the experiment. All two sections of each seafarer, calibration part and test part in time zone except for the self-rating were conducted by wearing the EEG device.

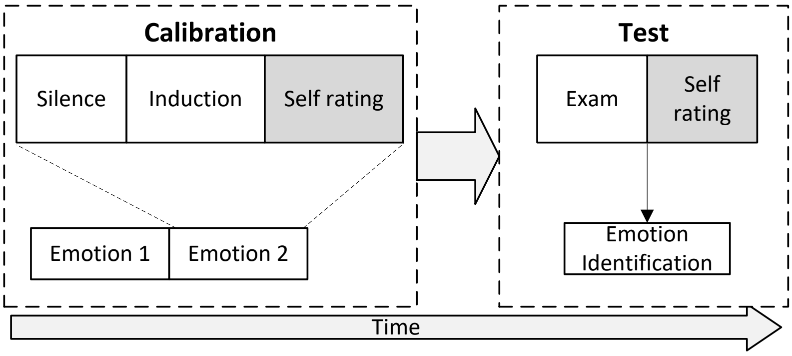


Figure 2. Experimental protocol

4 RESULTS

4.1 *Feature extraction of EEG data*

The EEG device collected 11 test subjects' brainwave signals in both calibration section and test section with the sample rate of 512 Hz. For each test subject, two pieces of calibration data had a duration of no more than 1 minute, and one piece of test data was within 30 minutes. Figure 3 reveals the primary EEG data of 2 seconds from test subject 1.

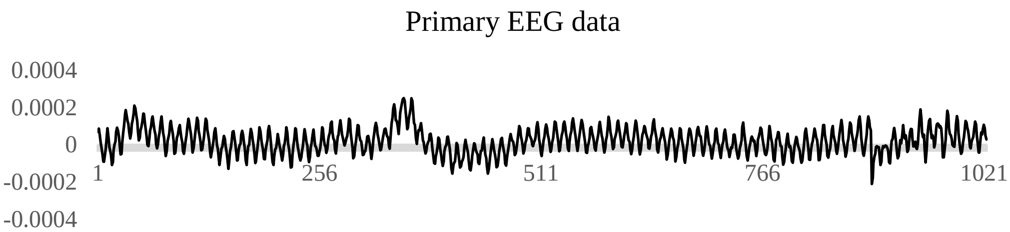


Figure 3. Two seconds primary EEG data from Neurosky Mindwave headset for subject 1- calibration)

In the calibration section, EEG data was extracted by wavelet analysis. A wavelet is a small wave oscillation with an amplitude that increases from zero, and then decreases back to zero. The wavelet transform is a methodology to construct the time-frequency representation of signals, to extract information from many different kinds of data. In this way, the original signal can be represented by a suitable integration over all the resulting frequency components. The Daubechies wavelets are orthogonal wavelets defining a discrete wavelet transform and featured by a maximum number of vanishing moments (Mahmoodabadi *et al.*, 2005). The dbN wavelets are the Daubechies' extremal phase wavelets, where *N* refers to the number of vanishing moments. In this study, Daubechies wavelets db8 was selected to extract features from the EEG data in the model, where 8 Level wavelet decomposition was used to obtain Gamma (40 Hz to 100 Hz), Beta (12 Hz to 40 Hz), Alpha (8 Hz to 12 Hz), Theta (4 Hz to 8 Hz) and Delta (0 Hz to 4 Hz) waveband. These five brainwaves related to different psychological concepts, *e.g*. Gamma waves correlate with anxiety and stress in high levels, depression in low levels; Beta waves are related to inability to feel relaxed in high levels, poor cognitive ability and lack of attention in low levels; Alpha waves usually concern over-relaxed state or an inability to focus in high levels, higher stress levels in low levels; Theta waves reveal hyperactivity or poor emotional awareness; Delta waves is associated with learning problems and poor sleep.

In order to obtain the feature matrix, features of the signal data were extracted with 512 Hz sample rate, where window size was 512, and windows increment was 32. Specifically, there was an input of 10510×1 matrix for test subject 1 in calibration part – negative emotion, then "datasize" equaled to 10510, "winsize" was 512, "wininc" represents 32, and the output was "313×5" matrix, where "313"=floor((datasize - winsize)/wininc)+1 and "5" represents five features: Gamma wave, Beta wave, Alpha wave, Theta wave, and Delta wave. The output matrix formed the classifier of feature extraction.

4.2 *Emotion classification*

In this study, emotion was classified into two categories: positive and negative. In the test section, EEG data was extracted by wavelet analysis, and then a classified by the SVM methodology.

SVM is used to identify the emotion category for the tested seafarers. SVM is a supervised learning model with associated learning algorithms that analyse data used for classification and regression analysis. It finds an optimised hyperplane, calculating the parameters constructing the hyperplan to maximise the margin between two sets while still separating the sets.

For EEG data analysis, it reveals the real-time emotion identification. There are five features describing every two kinds of emotion: Gamma wave, Beta wave, Alpha wave, Theta wave, and Delta wave. In the calibration part, the features matrix extracted from EEG data was used to train the SVM classifier. Then emotion in the test part of seafarers was identified by the classifier training by SVM.

In the questionnaire analysis, the classifier distinguishes the emotion describing the subjective feeling of whole examination, which is the overall emotion identification. These points were defined in three dimensions illustrated in SAM as pleasure, arousal, and dominance. As the emotion was a subjective variable, the SVM used the feature of a specific emotion in calibration to generate the classifier. Using the classifier training by SVM, emotion in the qualified test of seafarers was identified by the three-dimensional description questionnaire. After normalisation, the optimal parameters in the SVM were searched by cross-validation. The kernel function of the model was calculated. The result of identification of emotion taxonomy can be calculated.

4.2.1 *EEG data analysis*

Negative emotion and positive emotion were described in three-dimensional space of pleasure, arousal, and dominance. After extracting the EEG features in calibration section, given negative emotion and positive emotion, emotion classification was carried out by SVM model, where “1” represents negative emotion, and “2” means positive emotion.

For every test subject, there were two piecies of EEG data: calibration EEG data induced by IADS sound clip database, and test EEG data driven by operation process during the mission. The sample rate of emotion identification was 512 Hz, while the instantaneous emotion value was identified as two kinds of emotion. Then the average emotion value was calculated during a certain period, figuring out that the emotion (average emotion) value is between 1 and 2. Figure 4 shows the emotion identification of test subject 2 every 5 seconds. Figure 5 depicts the mean emotion value of single test subjects every 60 seconds, where the emotion value is defined between 1 and 2.

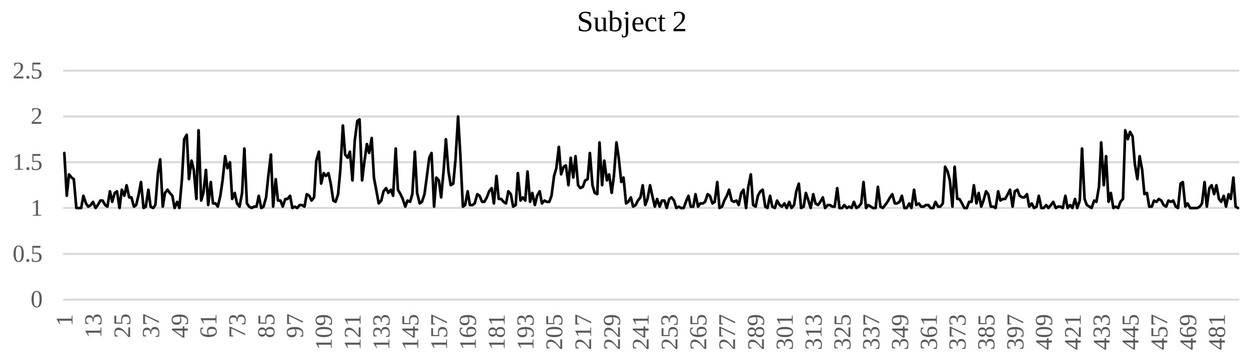


Figure 4. Emotion identification of subject 2 in test (every 5s)

For test subject 2, the SVM model extracted the features from EEG data in the calibration section to establish classifier with accuracy of 91.55% (390/426). It recognised the emotion value in the test section with the classifier. In Figure 4, it demonstrates the emotion value of test subject 2 every 5 seconds. Similarly, the approach can also be used to identify the emotion of other subjects given their exams in Figure 5.

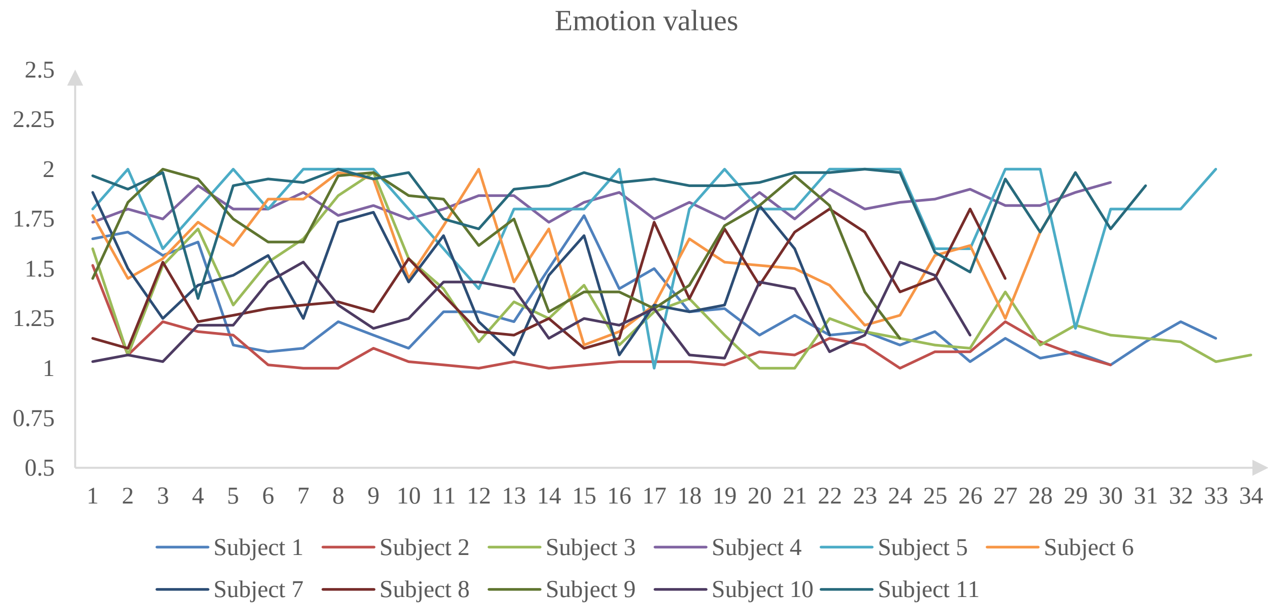


Figure 5. Emotion identification of subjects in the test (every 60s)

From the results, it shows that the emotion identification values of subjects fluctuate with time during the examination. Given the SVM model, the accuracy of classifiers are stated in Table 1, and the average accuracy is 77.55%. According to individual differences among the test subjects, emotion identification reflects various characteristics. Assuming that the emotion state can be described by a given emotion value, there are four levels emotion: extreme negative emotion within value [1, 1.25], less negative emotion within value (1.25, 1.5], less positive emotion (1.5, 1.75], and extreme positive emotion (1.75, 2]. The changes in the emotion value are associated with several events in scenarios during the test.

Table 1. Accuracy of the classifying method

Number	1	2	3	4	5	6
Accuracy	67.3704% (351/521)	91.5493% (390/426)	82.4773% (273/331)	66.3551% (284/428)	73.7226% (202/274)	70.7006% (222/314)
Number	7	8	9	10	11	Average
Accuracy	87.0712% (330/379)	69.6296% (188/270)	93.7269% (254/271)	80.6154% (262/325)	69.8225% (236/338)	77.55%

4.2.2 *Questionnaire data analysis*

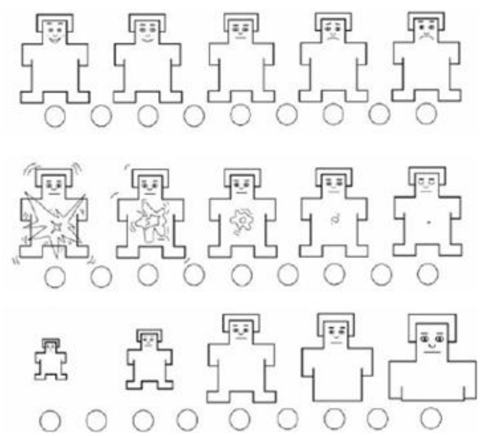
In this paper, the nine-point scale in SAM (Bradley and Lang, 1994) (Bradley and Lang, 2007) was used to describe pleasure, arousal, and dominance in response to the stimuli. Figure 6 shows the questionnaire that the test subjects need to complete after the experiments, reflecting on their subjective feelings during the assessment.

1. SAM rating

Happy — Unhappy

Excited — Calm

Control (not important) — In-control (important)



2. Word rating

Joyful	Surprised	Satisfied	Protected
Angry	Fear	Unconcerned	Sad
Or give your own descriptive word:			

Figure 6. The questionnaire of emotion with SAM scale on a nine-point rating (Liu *et al.*, 2016)

The scoring measures the degree of pleasure, arousal, and dominance associated with the stimuli. The first SAM is the happy/unhappy scale, which ranges from a smile to a frown. The second one is the excited/calm scale, which ranges from left to right. The last dimension is the controlled/In-control dimension. The left end of the scale represents the feeling of completely controlled and influenced whereas the right end of the scale is the feeling of completely in-control, important, and dominant.

The SAM methodology reveals the specific feature of a test subject's certain emotion, as it is a subjective variable. This method quantifies the emotion in a specific time and condition. After the qualified test, comments on the performance of seafarers from the experts is recorded by audio, and the test subjects are given a result of pass or failure.

This study collects 22 (11×2) calibration questionnaires and 11 test questionnaires reflecting 11 seafarers' emotions. Table 2 demonstrates descriptive statistics of seafarers in the experiments, while Table 3 presents the statistics in the IADS (2^{nd} edition) database. The clip sounds 105 represents negative emotion, while 220 represents positive emotion. Letters "p", "a", and "d" represent "pleasure", "arousal", "dominance" respectively while "t" means test emotion. The majority of the mean value in the test is, at large, consistent with the mean value of the IADS, except for the pleasure dimension in negative emotion.

Table 2. Statistics of seafarers in the questionnaires

	Min.	Max.	Mean	SD
105p	1	9	4.82	2.601
105a	1	7	4.18	2.272
105d	1	8	5.18	2.523
220p	3	9	8.09	1.814
220a	1	8	5.27	2.195
220d	3	9	6.36	1.912
tp	3	9	5.73	1.679
ta	1	7	4.64	2.063
td	1	9	6.00	2.449

*SD - Std. Deviation, p – pleasure, a - arousal, d – dominance.

Table 3. Statistics in the IADS (2^{nd} Edition)

	Mean	Std. Deviation
105p	2.88	2.14
105a	6.40	2.13
105d	3.80	2.17

220p	7.28	1.91
220a	6.00	1.93
220d	5.99	1.88

After collecting the emotion data from seafarers by SAM questionnaires, SVM was used to identify the emotion category during watch-keeping. Overall, 11 samples consisting of 33×3 matrix of emotion description, and 33×1 matrix of emotion labels were compiled. The former 22 pieces were from the calibration part as a training set for SVM. The later 11 pieces are from the test part as a test set. From these perspectives, the SVM model was constructed to find a hyperplane that divided the test set into two kinds of emotion categories. Figure 7 is the result of the test classification with the accuracy of 72.73% (the training accuracy of 95.45%), where “1” represents negative emotion, and “2” means positive emotion. The kernel function of this model is calculated in the way that “-$t = 2$” represents a kernel type radial basis function: exp $(-\gamma\times|x\text{-}x'|^2)$; “-$c$ = 776.0469” represents cost parameter C; “-g = 0.0068012” represents γ in the kernel function.

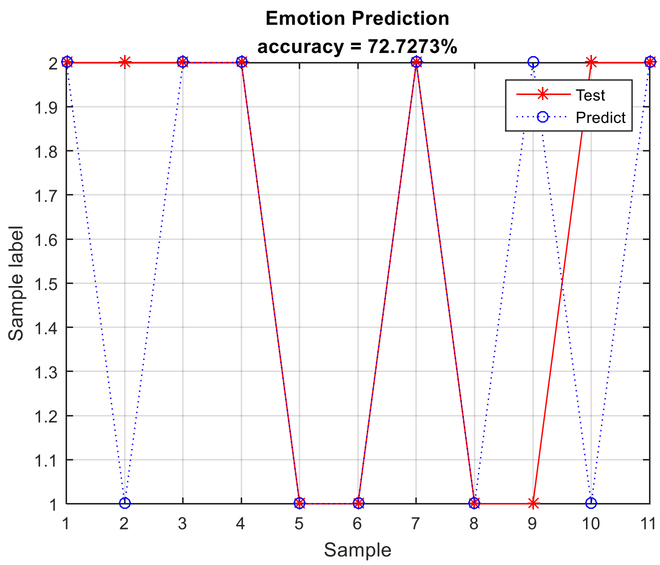


Figure 7. Emotion identification by using the SVM: Accuracy = 95.4545% (21/22) (training); Accuracy = 72.73% (8/11) (test)

The emotion identification by the questionnaire from both the test subjects and the SVM methodology are presented in Table 4, where “P” represents positive and “N” represents negative. More specifically, the self-rating emotions of subjects 2 and 10 are positive but were predicted as negative. The self-rating emotion of subject 9 is negative while it was predicted as positive. All the others have the same results between self-rating and SVM.

Table 4. Comments from self-evaluation and third party

ID	Emotion		Self-evaluation	Third party
	SR	SVM		
1	P	P	Untimely watch keeping in poor visibility Wrong operation sequence	Operate in incorrect sequence when stopping
2	P	N	Too late to realise poor visibility Speed control problem Inaccurate report in time	unconcerned watch keeping
3	P	P	Anxious when collision Wrong decision making (collision at ship body instead of bow)	Not fulfil the Convention on the International Regulations for Preventing Collisions at Sea (COLREGs)
4	P	P	Tension during ship encounter Response too late Unfamiliar with navigation device	Mistake for sail against the current Not fulfil COLREGs Too panic when stranding
5	N	N	Speed control problem Not enough communication Not stop timely	Wrong decision making of the captain Inappropriate manoeuvring
6	N	N	Speed control problem Course deviation	Not enough communication Not enough cooperation not enough
7	P	P	Late report in emergency Unconcerned Inappropriate manoeuvring	Wrong manoeuvring Too high speed Course deviation
8	N	N	Not familiar with rudder failure	Slow speed affecting steering Failure to meet a contingency
9	N	P	Not switch on navigation lights when starting fog	Not on-time watch keeping Too large deflection angle
10	P	N	Unfamiliar with navigation environment Not report the collision on time	Unfamiliar with navigation device Ignore environment when reporting Failed to fulfil COLREGs
11	P	P	Anxious when getting hurt	Speed control problem Irregular language

4.3 *Relationship between seafarers' emotion and events*

The scenarios of the test were not exactly same, as the questions in the exam database that test subjects chose before the qualifying exam were different. The events induced in the scenarios were commanded in the control room without specific or fixed time, so that the performance analysis given events relied on the marks in the examination and comments by the experts/examiners.

4.3.1 *Performance comments*

The comments on the examination for each test subject were further analysed to investigate if negative emotion identified by the SVM model affected human errors and human performance. Meanwhile, the comments from experts as an inevitable process of the qualified exam were collected by audios. It took place after the whole experiment, beginning with the summarised comments from self-evaluation and third party, and ending with experts' comments.

According to the self-evaluation from the subjects and experts, it is common to demonstrate that the human emotion emerging from watch-keeping affects ship-manoeuvring, concentration, response to an emergency, and decision-making. For example, test subject 1 was not able to concentrate on watch-keeping in poor visibility when sailing, which made him incapable of observing the crew onboard falling into the water. Moreover, a further step was supposed to stop in accurate and timely operation sequence. The test subjects 2 and 7 had the same result as unconcerned when encountering collision scenarios in poor visibility, resulting in a delayed report and operational problem. As a result, test subject 2 reported inaccurately in the collision scenario and subject 7 made an unnecessary course deviation. There was evident anxiety when the collision occurred as subject 3 demonstrated, causing not fulfilling COLREGs (International Regulations for Preventing Collisions at Sea). Subject 11 just became anxious when the crew got hurt, causing the irregular use of language and inappropriate manoeuvring. Test subject 4 had tension emotion when the encounter happened and panic emotion during stranding, which caused several mistakes, as shown in Table 5. Also, subjects 4 and 10 had physiological problems because they were unfamiliar with the device. They were not fulfilling COLREGs.

According to the above emotion problems existing in test subjects 1, 2, 3, 4, 7, 10, and 11, all of them rated overall positive emotion after the sessions. However, the subjects who rated a negative emotion did not reveal apparent emotion interruption on performance. Emotion rating through subjective judgement presents the overall feeling after the examination, whereas human errors occur at certain instant moments.

4.3.2 *Real-time relation to events*

From the scenarios of the test, several typical events are mainly considered: ship meeting/multi-ships encounter; emergency events such as stranding, collision, overboard or sudden illness of crews; reduced visibility in the condition of dense fog. The relationships between seafarer’s emotion identification and the occurrence of events are presented in Figure 8 and Figure 9.

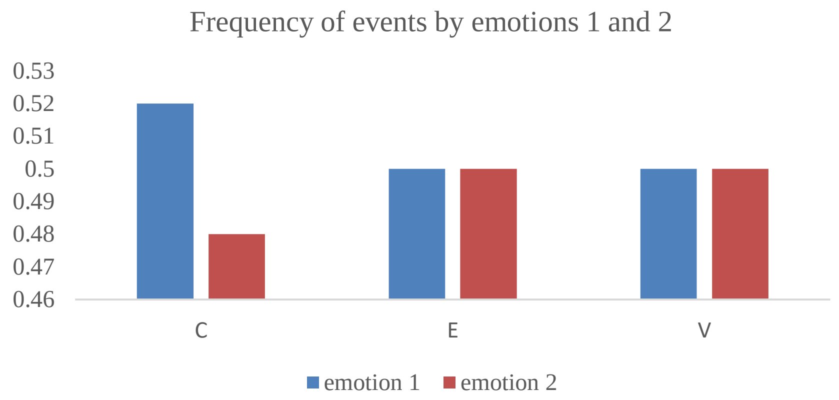


Figure 8 Frequency of events by emotions 1 and 2

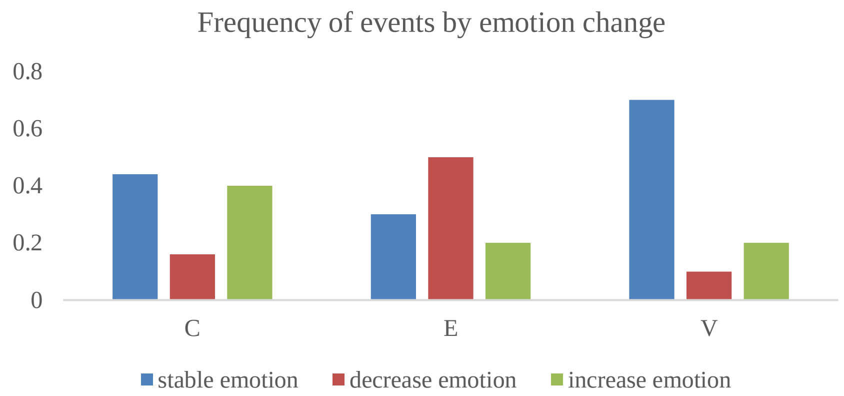


Figure 9 Frequency of events by emotion change

The events in scenarios for test subject 5 are lost due to the recording processes in the experiment. Therefore, the result of 10 subjects is demonstrated above. “C” represents ship meeting/multi-ships encounter; “E” stands for emergency events such as stranding, collision, overboard or sudden illness of crews; “V” means poor visibility in the condition of dense fog.

In ship encounter scenarios, test subjects tended to have both negative and positive emotion, and a subject may indicate two different trends on separate encounter process in the same test. Subjects 1, 2 and 10 reflected relatively smooth or stable emotion, while the other subjects showed differences. Subjects 4 and 6 showed decreased tendency of emotion in the first meeting condition, but increased emotion value on the second meeting condition. In addition, subject 11 revealed a falling emotion value at the first situation, then a stable state of emotion in a later situation. While subjects 8 and 9 demonstrated positive changes of emotion during the first encounter, but negative changes in the later ship encounter process. Subject 7 showed a positive tendency emotion in the condition all the time.

In emergency events, test subjects 2, 4, 11 had relatively stable emotion changes in an emergency; others showed obvious emotion dropping in emergency responses. From the experts' comments, they had problems with poor watch-keeping or were unfamiliar with devices onboard to some extent. Subjects 1 and 6 showed negative emotion and evidence-decreased emotion values to negative emotion. This was confirmed with manoeuvring, lookout or communication problems among their groups. Moreover the subjects 3, 8, 9, 10 demonstrated a sharp reduction of emotion values at the point of the emergency event and revealed to be incapable of fulfilling the regulation as well as committing errors.

In the condition of poor visibility, only test subject 3 showed a decreased change rate of emotion. Others had relatively steady or a slightly increased emotion state.

5 DISCUSSION

Overall, there are 13 cases which account for 8.07% likelihood of human errors happening within 161 negative emotion points, and 9 cases accounting for 7.20% likelihood of human errors existing in 125 positive emotion moments. As shown in Figure 9, the emotion values between 1.25 and 1.5 (where "1" represents negative emotion and "2" represents positive emotion) have the highest frequency (10.17%) of human errors, followed by the emotion values between 1.75 and 2 (frequency of 8.33%).

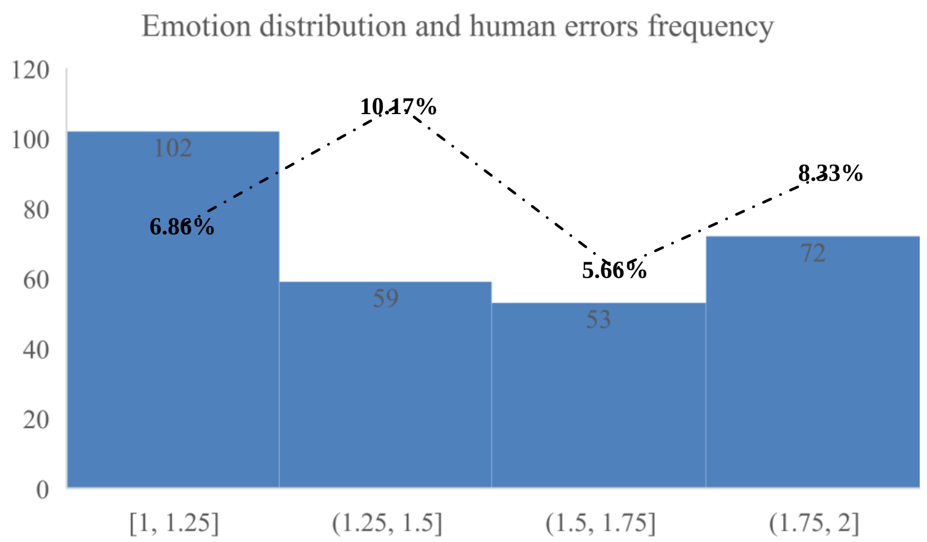


Figure 9 Emotion distribution and human errors frequency

From the questionnaire analysis, there is no definite correlation between overall emotion modes identified and behavioural consequences. As the rating is done after the examination, some seafarers may hide or ignore their true feelings in the questionnaire after the exam if emergency problems are adequately solved in scenarios. However, there is a link between the real-time emotion and events. It is evident that the seafarers' emotion changed along with the scenarios during the simulations. In this study, some subjects behaved better in a repeated situation, due to familiarity with the situation and readiness for the same condition, while others did not behave as good as the previous performance, due to over-confidence with the previous response and posibly due to a "too late" response for an emergency.

From the real-time physiological responses analysis, the link between seafarers' emotion and their performance is tied up to the factors contributing to the errors. It is evidenced that less negative emotion (1.25, 1.5] is more likely to contribute to human errors in this study, followed by extreme positive emotion (1.75, 2]. It is also derived from the accident report (MAIB, 2015) that overconfidence on duties or underestimation of severity of the condition during the navigation leads to errors. Thus, the relations between emotion and human errors are complex, and need to be further analysed considering the factors associated with human errors.

Moreover, this study incorporates an effect delay or advance in the experiment, as the response time and expected procedure of seafarers in the ship is different from it on the road or railway. For example, it is typical for the seafarers to follow a procedure or a checklist to deal with a collision situation instead of taking instant measures (*e.g*. brake hard to avoid collision on the road). Consequently, the psychological reaction of people may be prior to events exposure or postponed for executing an emergency plan after accidents.

6 CONCLUSION

Seafarers' emotion associates with sailing safety. It emerges during watch-keeping and could jeopardise their performance and decision-making. When an emergency happens, there are requests for a timely report and accurate operation of ships. This study utilises SVM as a classifier to extract features of EEG data with an average accuracy rate of 77.55%. The results concerning officers' emotion in a bridge simulator test reveal that seafarers' emotion from maritime operations affects their behaviour, and negative emotion has a higher likelihood of contributing to human errors than positive emotion. In addition, less negative emotion is the most dangerous emotion state during navigation, followed by extreme positive emotion.

Seafarers tend to be in a sensitive position when manoeuvring in a bridge simulator. The difference between bridge simulation and realistic navigation results in the change of emotional state of seafarers, which reveals the limitation of this study. Conducting psychophysiology research in a bridge simulator is significant on human error in maritime operations. In addition, the bridge simulation benefits research on human factors, especially for crew training purpose. In this regards, further studies will involve psychophysiological methods to design human error-oriented scenarios affecting seafarers' performance and measure their mental state in association with these factors.

ACKNOWLEDGEMENTS

The research was sponsored by the China Scholarship Council and the Double First-rate Project of WUT. The project was also supported by the EU project RESET (H2020-MSCA-RISE-2016, 730888), National Key Technologies Research & Development Program (2017YFC0804900, 2017YFC0804904) and National Science Foundation of China (51609194). The authors would like to thank the University of Florida for sharing the IADS database. The authors also appreciate the support of China MSA, the Bridge Simulator Centre in Wuhan University of Technology, and all bridge officers who participated in this experiment.

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

REFERENCES

AGUIAR, Y. P. C., VIEIRA, M. D. Q., GALY-MARIE, E. & SANTONI, C. 2015. Analysis of the user behaviour when interacting with systems during critical situations. *In:* MERCANTINI, J. M. & FAUCHER, C. (eds.) *Risk and Cognition.* Berlin: Springer-Verlag Berlin.

AKHTAR, M. J. & UTNE, I. B. 2015. Common patterns in aggregated accident analysis charts from human fatigue-related groundings and collisions at sea. *Maritime Policy & Management,* 42**,** 186-206.

AKYUZ, E. & CELIK, M. 2014. Utilisation of cognitive map in modelling human error in marine accident analysis and prevention. *Safety Science,* 70**,** 19-28.

AKYUZ, E. & CELIK, M. 2015. Application of CREAM human reliability model to cargo loading process of LPG tankers. *Journal of Loss Prevention in the Process Industries,* 34**,** 39-48.

BARSAN, E., ARSENIE, P., PANA, I. & HANZU-PAZARA, R. 2007. Analysis of workload and attention factors on human performances of the bridge team. *Pomorstvo,* 21**,** 57-67.

BRADLEY, M. M. & LANG, P. J. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry,* 25**,** 49-59.

BRADLEY, M. M. & LANG, P. J. 2007. The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual. *Technical report B-3.* University of Florida, Gainesville, Fl.

CELIK, M. & CEBI, S. 2009. Analytical HFACS for investigating human errors in shipping accidents. *Accident Analysis & Prevention,* 41, 66-75.

CHEN, S. T., WALL, A., DAVIES, P., YANG, Z. L., WANG, J. & CHOU, Y. H. 2013. A Human and Organisational Factors (HOFs) analysis method for marine casualties using HFACS-Maritime Accidents (HFACS-MA). *Safety Science,* 60, 105-114.

COOPER, S. E., RAMEY-SMITH, A., WREATHALL, J. & PARRY, G. 1996. A technique for human error analysis (ATHEANA). Nuclear Regulatory Commission, Washington, DC (United States). Div. of Systems Technology; Brookhaven National Lab., Upton, NY (United States); Science Applications International Corp., Reston, VA (United States); NUS Corp., Gaithersburg, MD (United States).

FAIRCLOUGH, S. H., VAN DER ZWAAG, M., SPIRIDON, E. & WESTERINK, J. 2014. Effects of mood induction via music on cardiovascular measures of negative emotion during simulated driving. *Physiology and Behavior,* 129, 173-180.

FAN, S., YAN, X., ZHANG, J. & WANG, J. 2017. A review on human factors in maritime transportation using seafarers' physiological data. 4th International Conference on Transportation Information and Safety, ICTIS 2017 - Proceedings, 2017. 104-110.

GEETHANJALI, B., ADALARASU, K., HEMAPRABA, A., KUMAR, S. P. & RAJASEKERAN, R. 2017. Emotion analysis using SAM (Self-Assessment Manikin) scale. *Biomedical Research*.

GRECH, M. R., HORBERRY, T. & KOESTER, T. 2008. Human factors in the maritime domain.

HANNAMAN, G., SPURGIN, A. & LUKIC, Y. A model for assessing human cognitive reliability in PRA studies. Conference record for 1985 IEEE third conference on human factors and nuclear safety, 1985.

HANZU-PAZARA, R., BARSAN, E., ARSENIE, P., CHIOTOROIU, L. & RAICU, G. 2008. Reducing of maritime accidents caused by human factors using simulators in training process. *Journal of Maritime Research,* 5, 3-18.

HETHERINGTON, C., FLIN, R. & MEARNS, K. 2006. Safety in shipping: The human element. *Journal of Safety Research,* 37, 401-411.

HOLLNAGEL, E. 1998. *Cognitive reliability and error analysis method (CREAM),* Elsevier.

HOU, X., LIU, Y., WEI, L. L., LAN, Z., SOURINA, O., MUELLER-WITTIG, W. & WANG, L. 2016. *CogniMeter: EEG-Based Brain States Monitoring,* Springer Berlin Heidelberg.

KIRWAN, B. 1994. *A guide to practical human reliability assessment*, CRC press.

KRAGEL, P. A. & LABAR, K. S. 2016. Decoding the Nature of Emotion in the Brain. *Trends in Cognitive Sciences,* 20, 444-455.

LAFONT, A., ROGÉ, J., NDIAYE, D. & BOUCHEIX, J.-M. 2018. Driver's emotional state and detection of vulnerable road users: Towards a better understanding of how emotions affect drivers' perception using cardiac and ocular metrics. *Transportation Research Part F: Traffic Psychology and Behaviour,* 55, 141-152.

LIU, Y., HOU, X., SOURINA, O., KONOVESSIS, D. & KRISHNAN, G. 2016. *EEG-based human factors evaluation for maritime simulator-aided assessment: Proceedings of the 3rd International Conference on Maritime Technology and Engineering (MARTECH 2016, Lisbon, Portugal, 4-6 July 2016).*

LIU, Y. & SOURINA, O. 2014. *Real-Time Subject-Dependent EEG-Based Emotion Recognition Algorithm*, Transactions on Computational Science XXIII.

LUCIDI, F., GIANNINI, A. M., SGALLA, R., MALLIA, L., DEVOTO, A. & REICHMANN, S. 2010. Young novice driver subtypes: Relationship to driving violations, errors and lapses. *Accident Analysis & Prevention,* 42, 1689-1696.

LUO, M. & SHIN, S.-H. 2016. Half-century research developments in maritime accidents: Future directions. *Accident Analysis & Prevention*.

LUTZHOFT, M. H. & DEKKER, S. W. A. 2002. On your watch: Automation on the bridge. *Journal of Navigation,* 55, 83-96.

MAHMOODABADI, S., AHMADIAN, A. & ABOLHASANI, M. ECG feature extraction using Daubechies wavelets. Proceedings of the fifth IASTED International conference on Visualization, Imaging and Image Processing, 2005. 343-348.

MAIB 2015. Report on the investigation of the collision between the passenger vessel Millennium Time and the motor tug Redoubt with 3 barges in tow on the Kings Reach, River Thames, London.

MENON, V. & LEVITIN, D. J. 2005. The rewards of music listening: Response and physiological connectivity of the mesolimbic system. *Neuroimage,* 28, 175-184.

MITTERSCHIFFTHALER, M. T., FU, C. H. Y., DALTON, J. A., ANDREW, C. M. & WILLIAMS, S. C. R. 2007. A functional MRI study of happy and sad affective states induced by classical music. *Human Brain Mapping,* 28, 1150-1162.

MORALES, J. M., DÍAZ-PIEDRA, C., RIEIRO, H., ROCA-GONZÁLEZ, J., ROMERO, S., CATENA, A., FUENTES, L. J. & DI STASI, L. L. 2017. Monitoring driver fatigue using a single-channel electroencephalographic device: A validation study by gaze-based, driving performance, and subjective data. *Accident Analysis and Prevention,* 109, 62-69.

READ, G. J. M., LENNÉ, M. G. & MOSS, S. A. 2012. Associations between task, training and social environmental factors and error types involved in rail incidents and accidents. *Accident Analysis & Prevention,* 48, 416-422.

ROIDL, E., FREHSE, B. & HOGER, R. 2014. Emotional states of drivers and the impact on speed, acceleration and traffic violations-A simulator study. *Accident Analysis and Prevention,* 70, 282-292.

SCOTT-PARKER, B. 2017. Emotions, behaviour, and the adolescent driver: A literature review. *Transportation Research Part F: Traffic Psychology and Behaviour,* 50, 1-37.

SONER, O., ASAN, U. & CELIK, M. 2015. Use of HFACS-FCM in fire prevention modelling on board ships. *Safety Science,* 77, 25-41.

SWAIN, A. D. 1987. Accident sequence evaluation program: Human reliability analysis procedure. Sandia National Labs., Albuquerque, NM (USA); Nuclear Regulatory Commission, Washington, DC (USA). Office of Nuclear Regulatory Research.

SWAIN, A. D. & GUTTMANN, H. E. 1983. Handbook of human-reliability analysis with emphasis on nuclear power plant applications. Final report. Sandia National Labs., Albuquerque, NM (USA).

TZANNATOS, E. 2010. Human Element and Accidents in Greek Shipping. *Journal of Navigation,* 63, 119-127.

WIEGMANN, D. A. & SHAPPELL, S. A. 2017. *A human error approach to aviation accident analysis: The human factors analysis and classification system*, Routledge.

XI, Y. T., YANG, Z. L., FANG, Q. G., CHEN, W. J. & WANG, J. 2017. A new hybrid approach to human error probability quantification-applications in maritime operations. *Ocean Engineering,* 138, 45-54.

YAN, L., ZHENG, K., WU, C., WEN, J., ZHU, D. & SHI, J. A Special laboratory method for inducing driving anger: Based on the virtual scene. Transportation Research Board 94th Annual Meeting, 2015.

YANG, Z. & WANG, J. 2012. Quantitative retrospective analysis of CREAM in maritime operations. *Advances in Safety, Reliability and Risk Manag*.

YANG, Z. L., BONSALL, S., WALL, A., WANG, J. & USMAN, M. 2013. A modified CREAM to human reliability quantification in marine engineering. *Ocean Engineering,* 58, 293-303.

YOSHIMURA, K., TAKEMOTO, T., MITOMO, N. & IEEE 2015. The support for using the Cognitive Reliability and Error Analysis Method (CREAM) for marine accident investigation. *2015 4th International Conference on Informatics, Electronics & Vision Iciev 15*.

ZHANG, H., YAN, X., WU, C. & QIU, T. Z. 2014. Effect of circadian rhythms and driving duration on fatigue level and driving performance of professional drivers. *Transportation Research Record Journal of the Transportation Research Board*, 2402, 19-27.

ZIMASA, T., JAMSON, S. & HENSON, B. 2017. Are happy drivers safer drivers? Evidence from hazard response times and eye tracking data. *Transportation Research Part F: Traffic Psychology and Behaviour*, 46, 14-23.

ZIO, E. 2009. Reliability engineering: Old problems and new challenges. *Reliability Engineering & System Safety,* 94, 125-141.