Neural Network Driven Eye Tracking Metrics and Data Visualization in Metaverse and Virtual Reality Maritime Safety Training

Evangelos Markopoulos\textsuperscript{1,2}, Mika Luimula\textsuperscript{1}, Calin Calbureanu-Popescu\textsuperscript{3}, Panagiotis Markopoulos\textsuperscript{1}, Pertti Ranttila\textsuperscript{1}, Sami Laukkanen\textsuperscript{1}, Niko Laivuori\textsuperscript{1}, Werner Ravvyse\textsuperscript{1}, Juha Saarinen\textsuperscript{1}, Tran Nghia\textsuperscript{1}
\textsuperscript{1}Turku University of Applied Sciences, Turku Finland
\textsuperscript{2}Queen Mary University of London, London GB
\textsuperscript{3}University of Turku. Turku Finland
\texttt{evangelos.markopoulos@turkuamk.fi; mika.luimula@turkuamk.fi; calin.m.calbureanu-popescu@utu.fi; panagiotis.markopoulos@turkuamk.fi; pertti.ranttila@turkuamk.fi; sami.laukkanen@turkuamk.fi; niko.laivuori@edu.turkuamk.fi; werner.ravvyse@turkuamk.fi; juha.saarinen2@turkuamk.fi; nghia.tran@edu.turkuamk.fi.}

Abstract—Understand the human brain, predict human performance, and proactively plan, strategize and act based on such information initiated a scientific multidisciplinary alliance to address modern management challenges. This paper integrates numerous advanced information technologies such as eye tracking, virtual reality and neural networks for cognitive task analysis leading to behavioral analysis on humans that perform specific activities. The technology developed and presented in this paper has been tested on a maritime safety training application for command bridge communication and procedures for collision avoidance. The technology integrates metaverse and virtual reality environments with eye tracking for the collection of behavioral data which are analyzed by a neural network to indicate the mental and physical state, attention and readiness of a seafarer to perform such a critical task. The paper demonstrates the technology architecture, data collection process, indicative results, and areas for further research.

Keywords: Neural Networks, Eye Tracking, Performance Management, Virtual training, Virtual Reality, Metaverse, Cognitive Science, Neuroscience, Behavioral science, Maritime safety training, Shipping.

I. INTRODUCTION

One for the major post millennial challenges that drives applied research globally is the introduction and integration of cognitive and behavioral sciences in business decisions support systems and operations. The advancements of technology in the creation, collection and analysis of user oriented big data, offers unlimited opportunities for human behavioral analysis, management, and improvement.

Cognitive task analysis aims to evaluate the cognitive workload that is being used in a task performance. Cognitive workload is bound by the available processing resources of the brain, it is coupled with task demands, and it fluctuates based on the quantity of concurrent information being processed. There have been numerous methods in use for this aim. The most common methods include the identification of the time spent on a task and the steps required to complete the task. However, a more novel approach to measuring cognitive workload has been through measuring different features of the eyes, also called eye metrics, and determine how they change according to the task at hand. Eye metrics have been identified as promising unobtrusive indices of cognitive workload. Such metrics include eye fixation duration, saccade, blink rate, pupil dilation, task performance, visualization, and scan path. Therefore, by tracking the eyes we can get more information about the cognitive processes that a person goes through when performing a task. One interesting part about eye tracking is that it has many benefits, relative to other approaches, is that it provides continuous, objective measurements with high temporal and spatial resolutions and without burdening or interrupting the user. [1]

II. VIRTUAL REALITY AND EYE TRACKING

Virtual Reality and eye tracking are utilized in several fields as several companies and organizations start to see their benefits. Varjo and Tobii demonstrate many case studies with clients ranging from medical professionals to astronauts, aviators, construction companies and many others that used this combination [2]. Virtual reality headset has nowadays high resolution and in-built hand and eye tracking features. In this paper we present research conducted on eye and hand tracking technologies in VR applications and introduce how we have integrated these technologies in the existing MarSEVR application.[3]

Eye tracking was first introduced in VR headsets in 2017 once Fove 0 was launched. In 2021, there were various manufacturers in the markets utilizing eye tracking sensors in their products such as HP Reverb G2 Omniece Edition and HTC Vive Pro Eye, and Varjo VR-3 and XR-3. All these three manufacturers have sub degree gaze accuracy; however the Varjo VR-3 and XR-3 HMDs have tracking frequency of 200Hz, compared to 120Hz with HP Reverb G2 Omniece Edition and HTC Vive Pro Eye. Some important metrics when it comes defining a good eye tracker are accuracy and precision, which are measured in degrees. The sampling rate is also very important. It means the frequency at which eye tracking data is captured.

In our studies we have also considered to use the iMotions software combined with Varjo’s headset within
iMotions platform which enables a use of multimodal biosensor data such as EDA (electrodermal activity), EMG (electromyography), ECG (electrocardiography), and EEG (electroencephalography). This data in the iMotions platform can provide insights but would not be brought inside virtual training episodes and their feedback systems [4]. Furthermore iMotions presents several case studies related to psychological therapy. VR and eye tracking have been used for example in treating social anxiety and post-traumatic stress disorders [5]

III. RESEARCH BACKGROUND

In previous studies, we presented the development of the MarSEVR training episode, and how to use virtual training technology in maritime safety training [6]. This work was extended to the ShipSEVR training episode which focuses specifically on ship engines and engine rooms safety procedures. [7]

Later we have introduced Maritime Immersive Safe Oceans Technology (MariSOT) [8] which will operate in the future as a training jukebox with each scenario (e.g. MarSEVR, ShipSEVR) to be executed upon schedule or demand. Our first experiences of eye and hand tracking integration has been reported in the MarSEVR training episode [9, 10].

Behavioural data in the command bridge scenario ranges from eye movements, finger and hand movements. They can provide relevant information about the user's cognitive state and task assessment. A proper level of concentration and focus is necessary for successful task completion and proper task evaluation, however the use of interface elements such as joysticks can put distance between the user and the environment he interacts with and can create a cognitive overload. [11]

In this paper we continue to investigate the use of eye-tracking in our Maritime Safety Training with VR (MarSEVR) technology. MarSEVR forms part of a larger maritime immersive ocean technology (MariSOT). MariSOT includes, among other packages, a suite of maritime safety training applications that meet the need of being available on-demand and onboard any vessel.

In particular, MarSEVR is a VR training application that aims to equip seafarers with continuous training opportunities in collision avoidance, environmental challenges and other anomalous situations that may arise while navigating maritime vessels. In our previous work with MarSEVR [9], we identified the potential of eye-tracking metrics as a series of numerical descriptors to determine pre- and post-training eye movement patterns among participants in a VR environment. Moreover, we proposed that further research, with data visualizations like dynamic heatmaps based on gaze data would significantly improve our understanding of the cognitive processes that support human learning.

IV. EYE TRACKING METRICS

Eye metrics have been identified as promising unobtrusive indices of cognitive workload. Therefore, by tracking the eyes we can get more information about the cognitive processes that a person goes through when performing a task.

Figure 1 indicates related results from experiments conducted expectations in our lab on how many times the subject looks at certain objects inside the simulation. With our algorithm, we can track the objects that are being looked at and not only the eye fixations on a 2D surface. This feature can help us create a 3D mapping of eye fixations that shows the points of interest for the subject, since it also gives a depth coordinate. Based on our research this feature does not exist in eye-tracking visualization software such as iMotions, and therefore can be considered a step towards a novel eye-tracking data visualization method.

![Figure 1. Table containing number of eye fixations on certain objects inside the simulation](image-url)

An interesting part about eye tracking is that its many benefits, relative to other approaches, provide continuous, objective measurements with high temporal and spatial resolutions and without burdening or interrupting the user. [1] In the context of behavioural data, this acts as an extension to the ability to assess performance. As seen in the previous studies [12], VR exposure, after a brief accommodation period, tends to mimic real-life activity, and can therefore be a good indicator for reliable behavioural data in the context of eye-tracking. This, and the implementation of finger tracking and hand recognition offer benefits to the concentration of the user on the learning tasks and objectives without the added confusion and stress of VR controls which require some time to adjust to. [10]

When focusing on eye movement behaviour, different eye metrics can play a role in assessing cognitive load (fig. 2). Eye-metrics such as fixations are very important in indicating where the subject is looking, and can reveal useful information regarding the stage of task assessment, he is in. Depending on the fixation location we can assess the areas of interest of the subject that can show which point of information was important and/or confusing to the subject. The pattern of the eye fixations can show a specific density of fixations in certain spots which are indicative of focused and efficient searching.[13]
Based on the characteristics of such fixations we can make more assumptions about a subject’s attention and cognitive processes. Therefore, intuitive data visualization is crucial to begin relevant data analysis for cognitive load.

Through the multitude of eye fixations, a scan-path can be created and compared to a standard, which can in turn allow a quantifiable method to assess performance. For indicating cognitive flexibility, we can use measures of pupil size. It has been shown that fluctuations of pupil size with power-law scaling and Long-Range Temporal Correlations (LRTCs) are characteristic to human psychophysical performance (fig. 3). Strong LRTCs parallel optimal cognitive flexibility and indicate a functionally advantageous state.[15] These LRTCs can provide insight into the subject’s optimal cognitive flexibility and can be used as a good performance indicator which can be a useful tool for use in neural networks.

This helps in giving insight into the crucial information necessary for learning, which is a desired outcome of a simulation. Implementation of such data into our neural network would help in designing a simulation that facilitates maritime education.[8]

V. NEURAL NETWORK EYE TRACKING DATA COLLECTION

Our research build on the eye tracking experiments conducted in MarISOT and improves autonomous feedback and evaluation with the utilization of neural networks. Our system can collect and evaluate user actions and behavioral data on real-time.

To create an autonomous system that can judge the user performance requires an analysis of sequential simulation data (fig. 4). The data is recorded 10 times in a second to have reasonable temporal resolution and keep the relationship between cause and effect. The input data consist of user input and the state of the simulation. The user actions are sequence of actions which will change the state of the simulation over the time.

During the simulation, the user purpose is to achieve the given goal such as, avoiding collision with other vessels, etc. It is typical that the order and timing of the user actions can vary for achieving the same state of the simulation. Obviously, there are proper and non-proper ways to achieve the same state of the simulation. Our neural network purpose is to learn from the training data what is the proper end state of the simulation. Furthermore, it should learn what are the proper input patterns for achieving that.

The architecture of the neural network is LSTM (Long Short Term Memory) with four hidden layers (~48 k parameters). LSTM is especially meant for learning patterns from long sequential data. The current neural network is trained with synthetic data [17]. The training data is automatically generated and annotated by running two different specific versions of the simulator. One version of the simulator will always generate “pass” results and the other one will always generate will “fail” results. Both versions use simple rule for judging the result depending how much the vessel deviate from the ideal path.

Our current training data contains 374 separate training sets. Each of the training set contains sequence of 7700 measurements. Each of those measurements contains multiple values. Currently we are using only three-dimensional input vector which contains two-dimensional location and the course over ground value. The initial test was done by dividing the dataset to the training set size of 270 sample and the validation set size of 100 samples. Based on the tests with the synthetic data the current LSTM architecture can achieve accuracy of ~90%. Training the neural network with real data is necessary to evaluate model’s capability to classify human generated data. Also adding more dimension to input vector is required.
VI. BEHAVIOURAL VS STEERING DATA

The data collected and analyzed can be categorized into behavioral data (by means of eye-tracking metrics) and steering data (the actual steering actions the user makes during gameplay). We expect that the utilization of is data comparison (behavioural vs. steering data) will generate various new research questions and same time will also create new business applications and opportunities. Already now we can say that behavioural data tells us what kind of mental state (such as attention or stress level and cognitive load) the trainee has during the training session.

We base our assertions on the work by Krejtz K. [18], who compared eye-tracker data for their aptness as gauges of cognitive load ascribed to task difficulty. The study in this question specifically tested two metrics in response to task difficulty: (a) the change in pupil diameter with respect to trial baselines; and (b) the rate and magnitude of microsaccades. They went on to conclude that task-evoked pupillary and microsaccadic magnitude are reliable and sensitive measures of cognitive load. Further evidence supporting eye-tracking data as a suitable quantification of mental fatigue was reported by Yamada Y., and Kobayashi, M. [19]. In their study, they utilized eye-tracking data to improve a previous model to 90% accuracy in the detection of mental fatigue among younger and older adults in natural viewing situations.

In our previous project called Virdi, we have developed NeuroCar, a virtual reality assessment tool to be used in driving inspection. Results achieved in Virdi proves that we can detect right-side perceptual bias and sleep deprivation in virtual driving simulator [20, 21]. These results and the knowledge of the development of assessment tool can be applied in maritime safety training.

Steering data in turn, tell us what really has happened during the training session. Combined with neural network and machine learning we can analyze how well the trainee was able to steer the vessel and complete all other user actions compared to an ideal way to proceed as a professional in the command bridge.

VII. EYE TRACKING DATA VISUALIZATION

The work developed for this paper illustrates and explains how scan path visualization, derived from eye fixations, together with pupil size can be combined to infer a metric for cognitive load. Cognitive load refers to the amount of information that working memory can hold at a given time [11]. Thereby, through avoiding activities and other distractions that do not contribute to optimal cognitive processing, we can produce VR applications that are more efficient and effective in delivering precise learning outcomes in our maritime training suite.

This paper explores the potential of cognitive load for VR design considerations and introduces the notion of using cognitive load as part of a neural network can assess and correct participant behavior while they undergo training with our MarISOT applications.

VIII. METAVERSE AND REAL TIME GAMEPLAY GUIDANCE

Virtual reality training episodes, especially occupational safety, can be considered more than gamified applications [22]. An important part of their gameplay is to assure the existence of sufficient instructions during the execution of the scenario. A trainee needs to have a clear understanding of the task and he or she needs to know how to use available equipment. In the case of MarISOT, real time guidance is divided into two parts: Guidance for the game scenario and guidance for the player about the execution.

Guidance related to the game scenario is, for example, instructions for using controls or instructions related to the player's task. The user interface and the various controls can be instructed with a separate pregame tutorial part and in-game through the user interface with visual elements.

Guidance relate to the execution of the scenario is based on instructions related to the player's tasks (fig 5). This can be provided in two separate ways: by operating instructions given before the start of the gameplay and by new instructions during the exercise. Instructions can be brought to the user through a user interface in the form of sounds and texts. The task can also change when the game scenario changes through different weather conditions or alarms for example.

![Figure 5. Instructions / guidance in two dimensional visualization showing correct performance in red.](image-url)
One goal in the MarISOT project is to provide player guidance based on an analysis of the player's execution. The purpose of this analysis is to replace the real-time guidance of an instructor or an experienced captain with artificial intelligence and machine learning. In real-time gameplay, the data is collected and analyzed during the exercise using eye tracking, hand tracking, finger tracking, and other cognitive science and artificial intelligence technologies (speech recognition, natural language processing, avatars, etc.). This analysis is used to predict the outcome of a particular decision made in the exercise. This prediction can then be used to provide additional information to the trainee and teach the player the impact of his/her decision right at the time of the decision.

IX. POST TRAINING FEEDBACK ANALYSIS

In any virtual training scenario, post-training debriefing and performance analysis are key to meeting the learning outcomes of the scenario. Moreover, the debriefing phase helps users to process and consolidate their VR learning events [25]. For successful feedback analysis to take place, it requires a monitoring report of in-training player actions and an expert facilitator to use the report as an enabler for deeper analytical discussions [26, 27].

Although our system is currently a work in progress, we have envisioned a dual reporting system that leverages the neural network timestamps and recording of the user actions where they have not made ideal choices. That is, for the first part of the report, the neural network would tabulate the player decisions versus ideal choices together with the cognitive load (derived from eye-tracking data) for the decision-making phase. The second set of reporting comes from video snapshots, taken at the time of the user transgression. The practical application of this system requires several cameras, over and above the first-person camera of the player view, to record the scene while the player is working through it. When a user transgression occurs, the system would retain video footage from each camera for a certain amount of time before and after the transgression. We intend to research the ideal video time retention and preferred number of cameras once the system has been completed. The first-person camera view will have additional visualizations from the eye-tracking system that include gaze heat maps (where the player was focused) and scan-pathing (the sequence of focus points) during the transgression event. Once again, the exact length of video and visualization detail will be researched at the completion of the system.

The debriefing reports will be stored per-episode on a secure remote repository that can be accessed by the user for that specific episode and the trainer. This allows the user to reflect on their performance at any time, but more importantly, it affords opportunity for the trainer and user to use the report as a discussion enabler at any time that suits them both, either remotely or in-person.

X. IMPACT TO THE MARITIME SECTOR

The research presented in this paper has been applied specifically on the MarISOT technology due to the impact of the maritime and shipping industry in the global economy but also due to the industry challenges on sustaining the professional qualifications of their human resources in safety operations.

Shipping is a conservative industry with long management traditions, not easy to accept technological advancements that can replace human expertise and judgement. On the other hand, the continuous pressure to reduce operational costs without risking the safety of the seafarers and the vessels with accidents that can have tremendous financial and reputational impact on the companies and environment makes such technologies extremely valuable and important.

Understanding the state of mind of marine officers in a collision avoidance scenario, as demonstrated in this research, can determine and reduce cost factors such as the insurance costs that impact the company’s credibility. The marine cargo insurance cost, for example, is calculated with 3% on the 110% of the cargo value [28] and covers only cargo damages, excluding environmental damages from any type of shipping accident. Shipping insurance determinants are related with the vessel seaworthiness, scope of coverage, voyage, type of cargo, etc.

There are many demanding aspects of seafaring such as the inability of employees to leave the worksite, extreme weather conditions, long periods away from home, and motion of the workplace. The use of cognitive science in controlled virtual reality environments can increase maritime safety performance and reduce shipping accidents by monitoring and modifying human factor issues such as fatigue, stress, health, situation awareness, teamwork, decision-making, communication, automation, and safety culture [29].

XI. AREAS OF FURTHER RESEARCH

The distance from theory to practice is what determines applied from theoretical research. This is achieved with extended tests that involve industry experts in the fields the research intends to be applied. In this case, it is within our research plan to involve more experienced and international seafarers from different backgrounds, ethnicities, educational levels, professional expertise and other characteristics that will help us determine social factors that impact the trainee’s behavior when virtually delivering a safety training scenario under real circumstances.

Similar research is being conducted using hand tracking and finger tracking technologies in virtual reality safety training scenarios. The integration of these technologies...
with the eye tracking will allow us to study the coordination between eyes, hands, and thoughts. Neural networks will remain the backend system for behavioral data analytics that will be further supported with expert systems for decision making and dynamic learning.

Furthermore, this research and the indicated results, especially on the impact of eye tracking in scientific and business applications can be extended and applied in other engineering disciplines such as the cognitive information communications (CogInfoCom), software development and testing. Besides the traditional procedures, the use of eye movement tracking systems can be an alternative method to analyse the effectiveness of different programming technologies [30]. Such Eye tracking applications can identify the readability and comprehensibility of the software code. This can be determined by analysing the heatmap and gaze route besides measuring and evaluating eye movement parameters [31]. The eye tracking metrics indicated in this research can also provide insights on how human cognitive capabilities can be integrated and extended with the cognitive capabilities of the virtual reality spaces and the digital devices surrounding the user, to enable seamless interactions between humans and artificially cognitive agents [32].

XII. CONCLUSIONS

Covid-19 impacted the global business operations, and strategies. The need for immediate adjustments in a new reality created threats and opportunities. It provided the space for new technologies to emerge and demonstrate research results that have been delivered years ago without the actual adaptation and acceptance they deserved [33].

Virtual reality, artificial intelligence and cognitive sciences are scientific disciplines that have been kind of neglected as futuristic, complicated, and expensive.

This paper demonstrated the integration of cognitive science technologies with virtual reality and neural networks to address the increasing demand for digital and remote professional training. The specific approach demonstrated in this research focused on human behavior analysis when the human is physically absent from the real training environment and expert judgment or instructions on the trainees performance cannot be possible or accurate due this this distance.

The maritime and the shipping industries have been strategically selected as the area of applying this research. According to the International Maritime Organization (IMO) “Shipping is perhaps the most international of all the world’s great industries and one of the most dangerous.” [34].

Ensuring that safety training in shipping can be improved with such research can lead into green ocean strategies [35] that can benefit the science, the economy, the environment, and the society.

ACKNOWLEDGEMENT

The authors would like to thank everyone who have participated in the development of this training VR technology. This work was supported by Business Finland and Turku University of Applied Sciences.

REFERENCES


[34] IMO (2021), International Maritime Organization Available at: https://www.imo.org/en/OurWork/Safety/Pages/Default.aspx