TRANSNAV
http://www.transnav.eu
the International Journal
on Marine Navigation
and Safety of Sea Transportation
Volume 10
Number 1
March 2016

DOI: 10.12716/1001.10.01.04

Extended Framework for Usability Testing in e-Navigation Systems

P. Zalewski & B. Muczyński
Maritime University of Szczecin, Szczecin, Poland

ABSTRACT: The paper presents the framework of usability testing of ECDIS equipment based on the IMO's Guidelines on Software Quality Assurance and Human-centred Design for e-Navigation. By incorporating the eye tracking techniques into the procedure it was possible to measure the visual attention distribution and the cognitive work-load. The presented method could be used to evaluate usability of every e-navigation system, which is necessary to ensure that the seafarers are able to successfully perform primary operations of systems upgraded with e-navigation functions, regardless of the type and specifications of the system and users' knowledge and experience with the system. The initial results are presented and discussed as the study is still ongoing.

1 INTRODUCTION

E-navigation systems are expected to enhance safety of navigation and security. With each new generation of navigational equipment, producers are trying to provide new features and extended functionality. However, additional e-navigation functions can make it more difficult to understand systems' primary information and may hamper the operation of the primary functions of the system with poor usability. For example, additional information in an Electronic Chart Display and Information System (ECDIS) may impede the route monitoring function. Therefore, the usability rating methods must be developed to ensure that the seafarers are able to successfully perform primary operations of systems upgraded with e-navigation functions, regardless of the type and specifications of the system and users' knowledge and experience with the system.

2 USABILITY TESTING

The International Organization for Standardization (ISO) defines usability as "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use". When considering usability of a software application the usability means the perception of a target user of the effectiveness (fit for purpose) and efficiency (work or time required to use) of the given interface. ISO and various research public and private organizations like NASA or SA Technologies have developed over the last 30 years several methods of rating and quantifying the task load (TL) (Hart 2006, NASA 1986) and situation awareness (SA) (Endsley 1988, Endsley 2013). Nowadays these methods are common industry standards used while assessing a task, system, or team's effectiveness, or other aspects of performance. They utilize many task-specific techniques to achieve the rating goal. For example, the monitoring of eye tracking in combination with

SAGAT is one of the specific techniques developed in Maritime University of Szczecin (MUS) for evaluation, improvement and general usability testing (UT) of ship's navigation equipment (Muczynski et.al. 2013, Zalewski et.al. 2012)

ISO/TR 16982:2002 provides information on human-centred UT methods, comprising various task-specific rating techniques, which can be used for design and evaluation of e-navigation displays. It details the advantages, disadvantages and other factors relevant to using each UT method. The digest of these methods is presented in the Table 1.

In March 2015, International Maritime Organization has issued the circular on "Guideline on Software Quality Assurance and Human-centred Design for e-Navigation", officially introducing UT methods into future electronic equipment for marine navigation. The appendix 3 of this guideline presents an UT process based on the ECDIS example as a one closely aligned with testing of future e-navigation systems. This UT example aligns with the integration and testing stage of a Human-Centred Design (HCD) process for evaluating the performance of essential tasks by competent users. The selection of test participants is important and has a bearing on the quality of test results. If tasks require operations based on navigational experience or knowledge, then appropriate participants should be selected. Tasks that are generally performed by less experienced or knowledgeable personnel should be similarly tested. The UT activity involves the following steps:

1 planning;
2 preparation;
3 undertaking and controlling tests;
4 evaluation of results; and
5 use of feedback.

A UT plan should be developed by defining scenarios and identifying the most important or critical tasks that users must perform. Users and the test environment are identified at this stage. A goal-based approach should be used when setting the tasks with the aim of facilitating flexible yet practical assessment of the target e-navigation system. The following steps can be a part of the goal-based approach:

1 definition of goals based on the context of use of the system, which may come from functions stipulated in internationally agreed performance standards;
2 specifying functional requirements or the criteria to be satisfied in order to conform to the goals, taking into account the relevant performance standards and user requirements;
3 specifying usability requirements that must be achieved during testing, based on the aspects of effectiveness, efficiency and satisfaction; and
4 preparation of tests that will assist in verifying the extent to which the system conforms with the identified goals.

A care has to be taken to guarantee the reproducibility of the test on different types of equipment and with the same settings and scenarios. This means that it is advised to avoid test scenarios where it is necessary to use the specific functionality available only in a single type of ECDIS.

Table 1. UT methods that can be applied while designing e-navigation products (ISO 2012)

Name of the method	Direct involvement of users	Short description of method
Observation of users	Yes	Collection of information in a precise and systematic way about the behaviour and the performance of users, in the context of specific tasks during user activity.
Performance-related measurements	Yes	Collection of quantifiable performance measurements in order to understand the impacts of usability issues.
Critical incident analysis	Yes	Systematic collection of specific events (positive or negative).
Questionnaires	Yes	Indirect evaluation methods which gather users' opinions about the user interface in predefined questionnaires.
Interviews	Yes	Similar to questionnaires but with greater flexibility involving face-to-face interaction with the interviewee.
Thinking aloud	Yes	Involves having users continuously verbalize their ideas, beliefs, expectations, doubts, discoveries, etc. during their use of the system being tested.
Collaborative design and evaluation	Yes	Methods which allow different types of participants (users, product developers and human factors specialists, etc.) to collaborate in the evaluation or design of systems.
Creativity methods	Yes / No	Methods which involve the elicitation of new products and system features, usually extracted from group interactions. In the context of human-centred approaches, members of such groups are often users.
Simulation	Yes / No	Use of computer simulation modelling tools for initial evaluations.
Document-based methods	No	Examination of existing documents by the usability specialist to form a professional judgement of the system.
Model-based approaches	No	Use of abstract representations of the evaluated product to allow the prediction of users' performance.
Expert evaluation	No	Evaluation based on the knowledge, expertise and practical experience in ergonomics of the usability specialist.
Automated evaluation	No	Algorithms focused on usability criteria or using ergonomic knowledge-based systems which diagnose the deficiencies of a product compared to pre-defined rules.

3 EYE TRACKING

Eye tracking is a set of techniques and methods to measure the position of the subject's eyes in relation to the visual scene. In this way a gaze point is obtained. Eye tracking itself has not been included in IMO's "Guideline on Software Quality Assurance and Human-centred Design for e-Navigation" but it meets criteria for both "Observation of users" and "Performance-related measurements" methods and can be used efficiently in "Simulation" methods. It has been proved as a valid method for usability testing in many previous researches (Ehmke, Wilson 2007, Goldberg, Kotval 1999, Strandvall 2009). Few studies reported usefulness of this technique in the ship's bridge environment (Papachristos et.al. 2012, Lutzhoft, Dukic 2007). (Jacob and Karn 2003) reports four most common eye tracking measures that are used in usability studies, those are:

1. Fixation: a relatively stable eye-in-head position within some threshold of dispersion (typically~2°) over some minimum duration (typically 100-200 ms), and with a velocity below some threshold (typically 15-100 degrees per second).
2. Gaze Duration: cumulative duration and aver-age spatial location of a series of consecutive fixations within an area of interest. Gaze du-ration typically includes several fixations and may include the relatively small amount of time for the short saccades between these fixations. A fixation occurring outside the area of interest marks the end of the gaze. In some studies, this measure is called "dwell", "glance" or "fixation cycle".
3. Area of interest: area of a display or visual environment that is of interest to the research or design team and thus defined by them (not by the participant).
4. Scan Path: spatial arrangement of a sequence of fixations.

Depending on the equipment used and type of a study a number of other measures can be used. A detailed list is given in (Holmqvist et.al. 2011) and includes number, duration and frequency of blinks, saccades direction and velocity and microsaccades. Comprehensive eye tracking study can give insight into search efficiency (e.g. due to poor arrangement of display elements), importance of specific interface element, task difficulty and participant's stress and cognitive work-load.

Usefulness of eye tracking techniques is hampered by several technical difficulties related to both data collection and data analysis. Despite technological advancement still around 10-20% of population cannot be tracked reliably, this is usually the case with older participants that have any kind of visual impairment and have to use either glasses or contact lenses. Another problem is related to the fact that each eye position is given in vertical and horizontal coordinates in a system that is fixed in the eye tracker–head frame. This means that when using a stationary eye tracker, the participant should restrict head movements to a small area (about a cubic foot). When a mobile eye tracker is used, which is the only reliable solution when conducting a study on a ship's bridge simulator, fixation coordinates has to be transformed into a ship's bridge coordinate system. This has to be done manually using frame-by-frame analysis or a dedicated software that allows for fixation-by-fixation mapping. Both methods are very laborious and time consuming and present a serious drawback for any study with a considerable number of participants and long scenarios. Last major difficulty is related to data interpretation. Eye tracking data analysis can proceed either top-down – based on cognitive theory or design hypotheses, or bottom-up – based entirely on observation of the data without predefined theories relating eye movements to cognitive activity (Goldberg et.al. 2002). For a usability study it is important to closely examine the data stream and relate it to the current task and environment. For example, when considering long fixations during an ECDIS' Usability Testing all external factors have to be identified before a statement about higher difficulty of the task can be made.

4 FRAMEWORK FOR ECDIS USABILITY TESTING

In the environment of the Full Mission Ship's Bridge Simulator (FMBS) in MUS the UT process of the two Kongsberg manufactured ECDISes, SeaMap 10 and K-Bridge 7.0, was conducted in accordance to the recommendations set by IMO. The goal was defined as "to plan and display the ship's route for the intended voyage and to plot and monitor positions throughout the voyage", based on SOLAS regulation V/19.2.1.4. Similarly, functional requirements for the ECDISes were defined based on the IMO's ECDIS performance standard (IMO 2006). The following functional requirements related to the nautical data handling necessary for safe navigation, with the following sub-requirements were taken into account:

1. Chart data handling (for instance: change dis-play orientation, mode, etc.);
2. Own ship data handling (for instance: read position, speed, etc.); and
3. Tracked target (TT) and radar data handling (for instance: show TT symbols overlaid on ECDIS chart area, etc.).

In the case of ECDIS, the "usability" can be evaluated in terms of user effectiveness and efficiency for each of the tasks and overall satisfaction of the system (for example through subjective evaluation by TL and SA). As highlighted in the Table 2, the measures of effectiveness were related to the difficulty and completeness of the task execution. The achievement rate was used as a measure of "effectiveness" and quantified by the four levels: "1. Smoothly", "2. Not smoothly", "3. With errors", "4. With suggestions". Usability outcomes were based on the dialogue principles, as identified in (ISO 9421-110:2006), using UT methods based on (IMO MSC.1/Circ.1512, ISO/TR 16982:2002).

The specific scenarios and tests, tasks were created to satisfy the functional requirements. The following are the tasks for a basic display handling scenario:

- Task 1: Adjust display modes and scale to meet operator's needs
- Task 2: Obtain information about a lighthouse
- Task 3: Measure the bearing and distance to a land-mark

- Task 4: Overlay a tracked target symbol and obtain information about the target

Quantitative performance criteria such as time taken to complete tasks and questionnaires which assist with overall subjective system evaluation were included in-line with the criteria set in the Table 2. These were necessary, as one can easily deduct while studying the Table 2 that, for example, to differentiate between "Achieved not smoothly" and "Not achieved" the time limit of the specific task completion must be set.

Table 2. Achievement criteria for the generic usability rating based on (IMO MSC.1/Circ.1512)

Achievement level		Criteria
Achieved	1. Smoothly	1. Participants understood the information correctly and operated properly with confidence. 2. Participants made some mistakes but noticed the mistakes immediately and achieved the goal smoothly.
	2. Not smoothly	1. Participants completed the task properly by themselves, but with some hesitation or confusion. 2. Participants took time to find the first action or to recover from errors but completed the task with a small number of interactions.
Not achieved	3. With errors	1. Participants could not understand the information correctly. 2. It took a large number of interactions to achieve the goal even if they completed the task properly.
	4. With suggestions	1. Participants could not complete the task by themselves and needed suggestions from the instructor or moderator.

Based on the study presented in (Muczynski et.al. 2013) participants were divided into groups with different experience. Two factors were taken into account:

1 General seafaring experience
2 ECDIS experience

Since both ECDIS generic and ECDIS type-specific courses are mandatory, it was considered if the participant took the type-specific course for a given ECDIS type and what was the last time when a participant was working with the same type of ECDIS.

Each participant was given the same set of task with no time limitation. To supplement the IMO recommendations, the eye tracking data were collected for each participant throughout the initial study. The mobile eye tracker "SMI Eye Tracking Glasses" was used for the data collection. The data was analysed using Semantic Gaze Mapping function provided by the SMI BeGaze software. Number, location, frequency and duration of fixations were recorded and mapped on the ECDIS interface. The interface was divided into 3 main Areas of Interest: chart area, alarms and sensor data and menus (Fig. 1). Because most of the menus are displayed on the screen after a specified button has been clicked, the size of Areas of Interests was not constant. When it was relevant and advisable fixations were identified with the accuracy to a single menu, submenu, button or either graphical or numerical information.

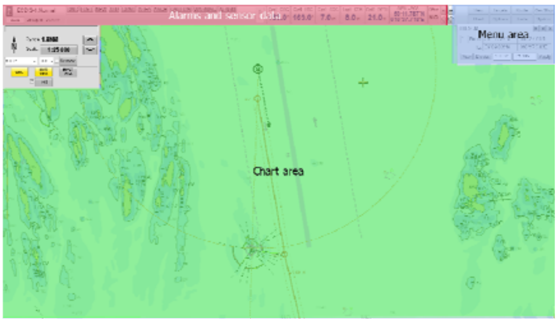

Figure 1. Main areas of interest on the ECDIS interface: chart area, menu area and alarms and sensor data

The initial framework for usability testing includes 4 stages, as given in table 3.

Table 3. Four stages used for ECDIS usability rating

Stage 1	Achievement level	Graded on scale 1 – 4 as shown in Table 2
Stage 2	Time	Total time required to accomplish given task
Stage 3	Eye tracking measures	All relevant data captured with the eye tracker
Stage 4	Scanpath analysis	Detailed analysis of participant's visual attention distribution

In the stage 1 each task is rated in a scale from 1 to 4 according to IMO guidelines (Table 2). During the first iteration of the study each task was evaluated by the experienced instructor. For the next iteration a dedicated computer software is being developed. This will help to make the data analysis faster and more efficient. It will also remove the subjectivity of the evaluation process and allow increase the reproducibility of this study.

The stage 2 is concerned with the time required for accomplishing each task. This is closely related to the first stage, when discriminating between level 1 and level 2. It is suggested to treat the time independently because by itself it provides an indirect measure of number of steps required for each task.

In the stage 3 all relevant eye tracking measures are taken into account and evaluated. This data provides a basis for a cognitive evaluation and should help in identification of those tasks that are the most demanding and result in increased workload for the participant. In the FMBS K-ECDIS study following measures were considered:

- total number of fixations per task,
- fixations frequency,
- fixations duration,
- location of fixations in a given area of the interface,
- gaze duration,
- number of blinks,
- duration of blinks.

Those measure have to interpreted and compared both between the participants and between the baseline scenario. The baseline scenario should be designed to provide low cognitive workload environment for each participant and should be used to collect eye tracking measures. The collected data is later compared to the baseline scenario and used to infer about the changes in cognitive workload, task difficulty and design issues.

Via stage 4 analysis it is possible to identify all distractors and errors during a given task, by close examination of scanpaths. For example, during a task of measuring a bearing and a distance to a given landmark, the scanpath shows precisely where the participant's attention was focused in any given moment (Fig. 2). On a typical scanpath fixations are represented as circles, where size of each circle corresponds to the fixation's duration and colour intensity is used for ordering – more recent fixations are shown with vivid and opaque colour. Without the eye tracking technique, it is only possible to register participant's actions and evaluate if those were either correct or incorrect. By incorporating the eye tracker into the study it is possible to register and evaluate participant's attention distribution. This shows not only actions but also intentions of the participant and makes it possible to recreate the search process on a cognitive level.

Due to the complex nature of the scanpaths it is not feasible to create and analyse a single scanpath for a complex task. For this kind of analysis each complex task should be divided into a set of simple sub-tasks.

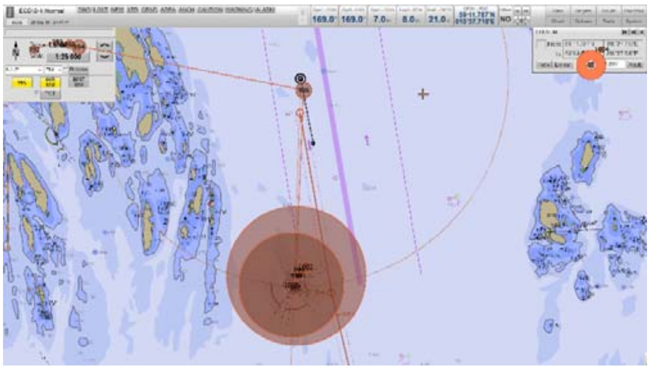

Figure 2. Scanpath for the task: bearing and distance measurement

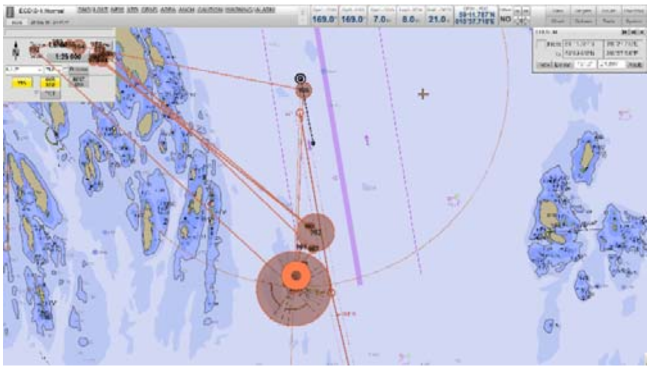

Figure 3. Scanpath for the task: bearing and distance measurement. Situation where participant had problem with setting an appropriate chart scale

Each sub-task should be clearly defined but only if it can be proved to be an indispensable step in a given task. For example, when a route creation task is considered, it can be divided into following subtasks: checking and setting default route parameters, opening a route window, defining a waypoint, saving a route and validating a route.

5 CONCLUSIONS

The developed framework extends the usability test procedure as described in IMO's "Guideline on Software Quality Assurance and Human-centred Design for e-Navigation". By incorporating the eye tracking techniques into the procedure it is possible to measure visual attention distribution and cognitive workload. This allows for a detailed analysis of the interface and identification of the major design flaws.

At this stage of the study the obtained qualitative and quantitative measures are preliminary and cannot be used for reliable estimations of the usability rating of the SeaMap 10 and K-Bridge 7.0 interfaces. The first stage of this study was necessary to develop the described framework and to verify its validity. To draw an unambiguous conclusion, it is necessary to conduct the study on a considerable number of participants so the sample size is large enough to describe the variability of each measure and the measure's significance for the usability rating.

The described procedure requires a considerable amount of time to analyse the eye tracking data, and a specialized research equipment, to obtain a precise and reliable eye movement data.

REFERENCES

Ehmke, C., Wilson, S., Identifying web usability problems from eye-tracking data in Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it - Volume 1 (BCS-HCI '07), Vol. 1. British Computer Society, 2007, Swinton, UK, pp. 119-128.

Endsley, M. R., Situation awareness global assessment technique (SAGAT). Proceedings of the National Aerospace and Electronics Conference (NAECON), pp. 789–795. New York: IEEE, 1988.

Endsley, M. R., Situation Awareness Oriented Design, In Lee, J. D. and Kirlik, A. (eds.), The Oxford Handbook of Cognitive Engineering. New York: Oxford University Press, 2013.

Goldberg, J.H., Kotval, X. P., Computer interface evaluation using eye movements: methods and constructs, International Journal of Industrial Ergonomics, vol. 24, no. 6, pp. 631–645, Oct. 1999.

Goldberg, J.H., Stimson, M.J., Lewenstein, M. Scott, N. & Wichansky, A.M., Eye tracking in web search tasks: design implications. In Proceedings of the Eye Tracking Research & Applications Symposium 2002. 51-58. New York, ACM.

Hart, S. G., NASA – Task Load Index (NASA–TLX) 20 Years Later, NASA-Ames Research Center Moffett Field, CA, 2006.

Holmqvist, K., Nyström, M., Andersson R., Dewhurst R., Jarodzka H., and van de Weijer J., Eye tracking: A comprehensive guide to methods and measures. Oxford University Press, Inc., 2011.

IMO MSC.1/Circ.1512, Guideline on Software Quality Assurance and Human-Centred Design for e-navigation, IMO, London, 2015.

IMO MSC.232(82), Revised ECDIS Performance Standards, IMO, London, 2006.

IMO SOLAS, International Convention for the Safety of Life at Sea, 1974, as amended, Consolidated Edition, IMO, London, 2014.

ISO 9421-110:2006, Ergonomics of human-system interaction – Part 110: Dialogue principles, ISO, 2009.

ISO/TR 16982:2002, Ergonomics of human-system interaction – Usability methods supporting human-centred design, ISO, 2002.

Jacob R. J. K. and Karn K. S., Commentary on Section 4. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises., in The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research, Elsevier, 2003, pp. 573–605

Lützhöft, M., Dukic, T., Show me where you look and I'll tell you if you're safe: Eye tracking of maritime watchkeepers. Proceedings of the 39th Nordic Ergonomics Society Conference, 2007, Lysekil, Sweden

Muczyński B., Gucma M., Bilewski M., Zalewski P., Using eye tracking data for evaluation and improvement of training process in ship's navigational bridge simulator. Scientific Journals Maritime University of Szczecin, 33(105), 2013, 75–78.

NASA, NASA Task Load Index (TLX) v. 1.0 Manual, NASA, NASA-Ames Research Center Moffett Field, 1986.

Papachristos, D., Koutsabasis, P., & Nikitakos, N., Usability evaluation at the ship's bridge: A multi-method approach. 4th International Symposium on Ship Operations, Management and Economics., 2012, Athens, Greece.

Strandvall T., Eye Tracking in Human-Computer Interaction and Usability Research, in Human-Computer Interaction – INTERACT 2009, vol. 5727, T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. Prates, and M. Winckler, Eds. Springer Berlin Heidelberg, 2009, pp. 936–937.

Zalewski P., Tomczak A., Gralak R., Simulation Analysis of ECDIS' Route Exchange Functionality Impact on Navigation Safety, The European Navigation Conference 2012, Gdańsk, 25-27.04.2012, Annual of Navigation 19/2012/part