Risk-Based Information Integration for Ship Navigation

Boris Gauss and Matthias Rötting

Berlin University of Technology, Department of Human-Machine Systems,
Sekr. 2-7/1, Franklinstr. 28-29, 10587 Berlin, Germany
{gauss, roetting}@mms.tu-berlin.de

Abstract. The Navigational Risk Detection and Assessment System (NARIDAS) is an approach to risk-based information integration on the ship's bridge. The purpose of this novel system is to reduce data overload and to support situation awareness of the bridge team. This paper focuses on the evaluation of NARIDAS during the development process. Evaluation is performed with system prototypes and practitioners. Three levels of evaluation are addressed: risk model validity, graphical user interface (GUI) design, and system usability. In two evaluation studies, positive results were obtained on all three levels. These results suggest that NARIDAS provides a valid model for the risks of ship navigation, a well-designed GUI, and a high usability for enhancing situational risk awareness of the bridge team.

Keywords: Ship Navigation, Risk, Support System, Evaluation, Situation Awareness, Information Integration.

1 Introduction

Approximately 80% of maritime accidents are attributed to "human error". Analyses show that many of these accidents occurred because the bridge team had lost situation awareness [1]. Situation awareness can be defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [2]. We assume that an important reason for loss of situation awareness on the ship's bridge is data overload, caused by ill-designed human-computer interaction. Data overload is considered a significant problem in many domains of human-computer interaction [3]. Modern ships are equipped with extensive technological aids for navigation (e.g., Automatic Radar Plotting Aids, Electronic Chart Display and Information System, Automatic Identification System). Due to the prevailing technology-centred approach to system development, usability of these systems is often low. Practitioners find the design of modern ships bridges an "ergonomic nightmare" [4].

In consequence, support provided by existing navigation aids is limited. If we consider the function of existing systems for the information processing of the human operators, we find an ever-growing number of 'information acquisition systems' on the bridge (Stage 1 of human information processing, Fig. 1). More and more information is acquired by technological systems, but the tasks to assess the information from multiple sources (i.e., to achieve and maintain situation awareness),

D. Harris (Ed.): Engin. Psychol. and Cog. Ergonomics, HCII 2007, LNAI 4562, pp. 668–677, 2007.

and to decide what to do next remain with the human operators. The bridge team cannot profit any more from the very fast and accurate numerical description of the ship navigation process, because the overabundance of data presented by technological systems exceeds their cognitive capacity.

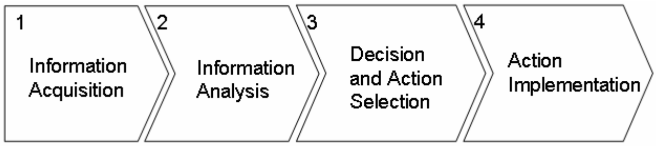

Fig. 1. Four stages of human information processing [5]

A possible solution to the problem of data overload, caused by too many "information acquisition systems", is the development of support systems for the cognitive processing stage of information analysis. On this stage, the information acquired on the first stage is integrated by relating it to the current goals. The operators extract the *meaning* of the information in their task environment for decision and action selection. An important semantic category at this stage is the concept of *risk*. Risk can be defined as the anticipation of an event with negative consequences. In dynamic human-machine systems, subjective risk assessments are directly linked to decision making and action. If subjective risk is too high, the operator will change his or her plan and take adjusting actions to reduce risk to an acceptable level [6]. Of course, it is crucial for adequate decision making that risk is assessed correctly, i.e., that the operators' subjective risk reflects the situation's actual or "objective" risk. Thus, an "objective" risk assessment system could support the cognitive processing stage of information analysis in order to overcome the data overload problem. In addition to the raw sensor data of information acquisition systems, a risk assessment system offers a task-oriented integration of the acquired information.

The rest of this paper is organised as follows: In section 2, we provide a short description of the Navigational Risk Detection and Assessment System (NARIDAS). Section 3 is dedicated to the evaluation framework within the NARIDAS development process. Section 4 presents the procedure and the results of two evaluation studies. In section 5, we discuss our results and outline some perspectives for further research and application.

2 The Navigational Risk Detection and Assessment System

For ship navigation, the *Navigational Risk Detection and Assessment System* (NARIDAS) is an approach to support integration of nautical data by dynamic risk assessments. The basis of NARIDAS is the breakdown of the bridge team's navigation task into eight task dimensions [7]:

- COLLISION AVOIDANCE (COL): pass other ships or objects safely
- ANTI-GROUNDING (GRD): adjust own ship's speed to the natural conditions

- TRACK KEEPING (TRA): keep track and consider manoeuvering area
- TRAFFIC (TRF): account for characteristics and density of traffic
- BRIDGE MANNING (MAN): consider the condition of the bridge crew
- ENVIRONMENT (ENV): account for the meteorological and hydrological conditions
- ENGINE/WHEEL (ENG): consider the state of propulsion and rudder engines
- ECONOMY (ECO): comply with the economic criteria of the voyage

For each of these task dimensions, NARIDAS calculates the corresponding risk by means of knowledge-based and rule-based procedures. In a first step, approximately 100 technical or physical input parameters – that are continuously updated from various sources (e.g., radar, electronic chart, integrated navigation system) – are processed by crisp mathematical algorithms for nautical calculations. In doing so, the input parameters are integrated into 24 higher-order variables. These higher-order variables are further processed with fuzzy algorithms comparing their current values with standard values for "good seamanship" to obtain the eight navigational risk values on a scale from 0="No Risk" to 1="Accident".

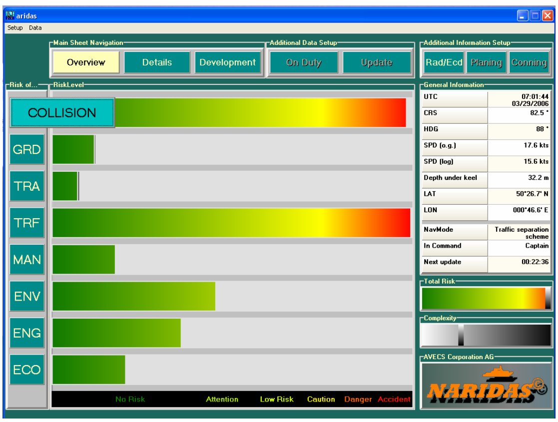

Fig. 2. NARIDAS graphical user interface

On the graphical user interface (GUI) of NARIDAS, the eight situational risk values are displayed in a bar graph (Fig. 2). This comprehensive display allows for an assessment of the situational risks of the navigation process at a glance. In addition, the system offers access to more detailed explanations, so the users can check the reasons behind the system's risk assessments.

Since navigational risks are context-specific, the NARIDAS knowledge-base is customised on three different levels: (1) long term: to the particular ship (manoeuvring properties, engine characteristics etc.), (2) medium term: to the voyage plan (way points, estimated time of arrival etc.), and (3) short term: to the current sea area. For the latter, NARIDAS distinguishes between six different "navigation modes" ("coastal waters", "approach", "traffic separation scheme", "fairway", "open sea", "at anchor"). For each navigation mode, a specific knowledge base is activated.

3 Evaluation Framework

To avoid the problems resulting from technology-centred development ("ergonomic nightmare", see above), NARIDAS is developed in a parallel-iterative process. From early stages of the development process, we work on technological and human aspects of the system in parallel. The match of these aspects is controlled in iterative evaluation loops with prototypes and the participation of practitioners. The main objective of the evaluation is formative, i.e., to gather information about how to improve the system. Evaluation in the NARIDAS development process can be assigned to an 'evaluation pyramid' of three levels (Fig. 3). On the basic level, the validity of the NARIDAS risk model is verified. Secondly, the design of the graphical user interface (GUI) is reviewed. Finally, the usability of the complete system is evaluated.

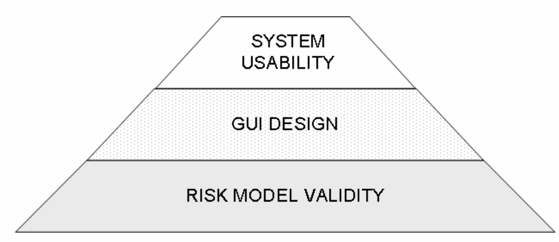

Fig. 3. Evaluation pyramid

3.1 Risk Model Validity

NARIDAS was invented by a domain expert, Dr. D. Kersandt, on the basis of his vast experiences on board ships as Master and Nautical Officer as well as in the academy as nautical instructor and accident researcher [7]. He designed and adjusted the NARIDAS knowledge base, i.e., the algorithms for risk calculation, in several years of development work. The evaluation objective at the basic level is to check how well the algorithms represent the risk assessments of other nautical experts. Of course, it is an essential prerequisite for the usability of NARIDAS that the risk algorithms reflect the common view on risk, and not just the personal opinion of a single expert. Key criteria on this level are sensitivity and selectivity of the model's risk assessments. Sensitivity refers to the degree to which the model distinguishes between different states of risk, in particular, the degree to which it detects states of "objective" high risk. Selectivity is the degree to which the model is sensitive only to changes in "objective" risk.

3.2 Graphical User Interface Design

The most important question at this level is how the risk values should be displayed to provide an optimal overview of the situation. Also, the presentation of the additional information (e.g., details of risk calculations, explanation components), the menu structure, and the GUI's conformity with general dialogue principles according to ISO

9241-111 [8] (e.g., controllability, error tolerance, suitability for learning) have to be evaluated.

3.3 System Usability

Usability is defined 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" [9]. The two basic levels of the evaluation pyramid can be considered necessary conditions for system usability. But a valid risk model and a usable GUI are not sufficient for usability of the complete system. To evaluate usability, the three criteria of effectiveness, efficiency and satisfaction have to be specified and measured. While satisfaction can only be assessed with "soft" subjective judgments on the system by its users, effectiveness and efficiency should be confirmed by "hard facts" where possible. However, effectiveness and efficiency of a risk assessment system are not easy to prove. In particular, the economic benefits (i.e., efficiency) are hard to calculate prospectively. As a first step, we focus on the effectiveness of the system in terms of its effects on situational risk awareness and navigation performance of the bridge team.

4 Evaluation Procedure and Results

Until now, two studies were conducted in the NARIDAS evaluation process. Study I addressed the two basic levels of the evaluation pyramid. Study II investigated the top level.

4.1 Study I: Evaluation of Risk Model and GUI Design

Study I was carried out in two rounds with a functional NARIDAS prototype, which presented the GUI, and contained the nautical data and risk values for several pre-defined static traffic scenes. The objectives of the study were (1) to compare the NARIDAS risk values with risk judgments of nautical experts and (2) to enquire the experts' opinions about the GUI and the overall concept of this nautical risk assessment system.

Procedure. Participants were 16 nautical experts (masters, mates, final-year students) of German nationality. All of them were between 25 and 60 year-old men with nautical experience on board of large vessels world-wide. The study was conducted in individual trials. After an introduction to NARIDAS, 14 static traffic scenes were presented to the expert. These scenes represented a broad range of different navigational requirements (e.g. passing Straight of Gibraltar; approaching port of Livorno; open sea) and environmental conditions. For each scene, the experts received data about own ship characteristics (pilot card), traffic situation and sea area (screenshots of radar and electronic chart), and environmental data (wind, waves, visibility etc). Experts were instructed to judge the navigational risks of the traffic scene on the eight dimensions. After the risk assessment, a computer screen with the functional prototype was switched on, so that the experts could explore the system and compare their own risk assessments with the NARIDAS values. During risk assessment and system exploration, experts were asked to think aloud. Verbal data was recorded, transcribed and analyzed qualitatively. After completion of the risk judgments, a detailed usability questionnaire with rating questions (5-points Likert-Scale) was administered. At the end of each trial a short structured interview was held on the expert's opinion about NARIDAS.

Results. Over all traffic scenes and risk dimensions, experts' judgments and NARIDAS values were highly consistent (Cronbach's Alpha between .89 and .94). For analysis of sensitivity and selectivity, rates of "*misses*" (sensitivity) and "*false alarms*" (selectivity) were determined.

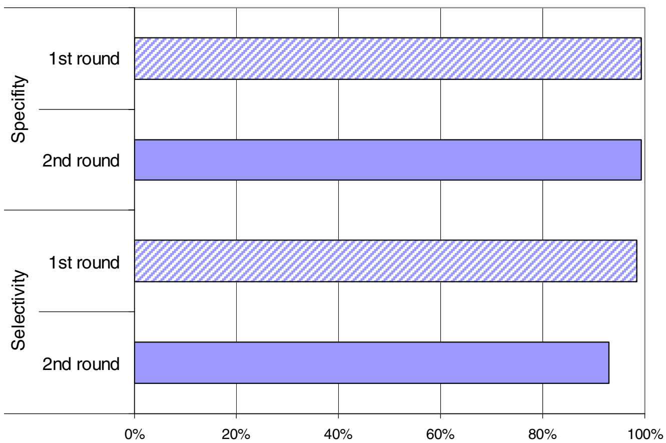

Fig. 4. Sensitivity and selectivity in both rounds of study I

A *miss* was defined as a case if >50% of the experts assessed a risk as "dangerous" (>.80) and NARIDAS assessed the risk as "not dangerous" (<.60). A false alarm was defined as a case if NARIDAS assessed a risk as "dangerous" and >50% of the experts assessed the risk as "not dangerous". With 0.9% of misses (both rounds), and 1.5% (1^{st} round) and 7.1% (2^{nd} round) of false alarms for a total number of 112 cases (14 scenes*8 risk values), sensitivity and selectivity of NARIDAS were high (Fig. 4).

In the questionnaire, the GUI was rated very positive. Participants judged the NARIDAS interface as clearly designed and easy to use. Overall usability of the system, assessed on a 10-items-scale (e.g., "NARIDAS is a reliable system", "NARIDAS would enhance the safety of navigation"), achieved 73.2 of 100 points. Also the qualitative data (think-aloud protocols, interviews) showed that the experts considered NARIDAS as a useful support to ship navigation.

4.2 Study II: Experimental Evaluation of System Usability

For study II, a highly-developed NARIDAS prototype was implemented in the full-mission ship-handling simulator in Elsfleth (Lower Saxony, Germany) (Fig. 5). The Elsfleth simulator provides four interconnected, fully equipped ship's bridges, two of

Fig. 5. NARIDAS in the Elsfleth Simulator

them with visual simulation system. These two were used for the present study. The objectives of the study were to investigate the effects of NARIDAS on situational risk awareness and navigation performance of the bridge team during a simulated voyage.

Procedure. NARIDAS was connected to the simulator network, so it was calculating the risks online during the whole voyage. Participants were 23 nautical students in the final year of their studies (all men; age between 21 and 48 years, mean=28 years). They were grouped into 11 bridge teams, each team consisting of one "Master" and one or two "Watch Officers". A traffic scenario in the English Channel of 80 minutes was constructed with high traffic density and rather unpleasant environmental conditions (4m swell from 220°, 30kn wind from 180°, 2.5kn current from 50°). Own ship was a container vessel traveling from Cadiz to Rotterdam.

A simple one-factor experimental design was realized with "NARIDAS support" as independent variable, which was varied within teams. Each team traveled one 40-minutes section of the trip with NARIDAS, the other 40-minutes section without NARIDAS. The sequence of sections with and without NARIDAS was balanced between teams. Dependent variables were assessed with a combination of different methods. After each section, rating questionnaires were applied to assess situational risk awareness (SRA) and navigation performance (self-ratings by the subjects, and assessment of the teams by an experienced instructor). Furthermore, SRA was measured with an online-test, 3 times during each voyage section (after 15, 25 and 35 minutes). For this test, the 'Master' received a phone call from the experimenter. He was asked to report the three most dangerous risks at the particular moment, and to rate these risks on a scale from 0 to 100. The answers were recorded, and categorized ex post to the NARIDAS risk dimensions for analysis. As an additional indicator for navigation performance, NARIDAS risk values were recorded during the whole trip.

Results. In the SRA online-test, subjects had higher risk awareness in the sections traveled with NARIDAS support (Figure 6). In particular, more collision risks (i.e., dangerous radar targets) were reported by the participants. The difference between the sections with and without NARIDAS is statistically significant (Wilcoxon-Test, p<.01). Results also show that only three of the eight risk dimensions (collision, environment, and traffic) were rated "dangerous" during the test. This indicates that overall complexity of the traffic scenario was rather low. The teams were able to handle the requirements of this simulator exercise without major problems.

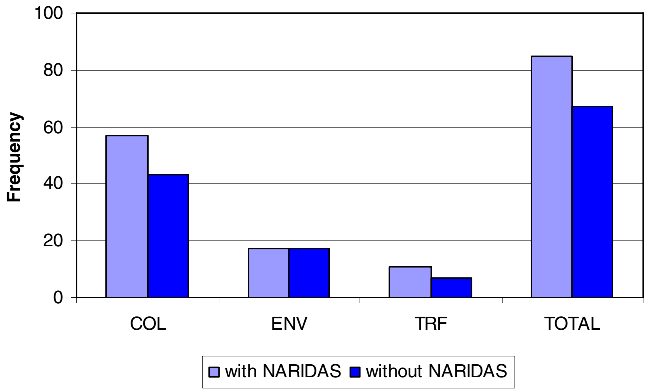

Fig. 6. Reported risks in the SRA online-test

Analysis of navigation performance showed that with NARIDAS, a higher risk of collision (the most important risk dimension in the scenario) was associated with better SRA and navigation performance ratings by the instructor. Without NARIDAS, a higher risk of collision was associated with a poorer instructor rating (Table 1). This result suggests that NARIDAS can contribute to a better handling of high risks. If a high risk is taken consciously (i.e., with a high SRA, supported by a risk assessment system), navigation performance is good, and the situation remains under control. In contrast, if the navigators take a high risk without recognizing it (lower SRA, no support), navigation performance becomes unstable.

Table 1. Correlations between instructor ratings and recorded COLLISION risks

Correlations (Spearman-Rho, *p<.05)		COLLISION risk With NARIDAS	COLLISION risk Without NARIDAS
Instructor Ratings	Navigation Performance	.48	-.59*
	Situational Risk Awareness	.68*	-.45

Usability of NARIDAS was rated positive by the participants, and their satisfaction with the system was high. In an overall judgment, 19 participants rated NARIDAS as "good" or "very good", the other 4 participants as "neither good nor bad". There were no negative judgments on this novel system.

5 Discussion

In the two empirical studies, results were encouraging on all three levels of evaluation. In study I, the risk values calculated by NARIDAS matched very well with the risk judgments of nautical experts. Sensitivity and selectivity were high. These results indicate that the NARIDAS risk model is valid. Furthermore, these findings imply that there is a common view on the navigational risks among nautical experts, and this common view can be modeled by a combination of mathematical and fuzzy-set algorithms. However, it should also be noted that consistency of risk assessments between the experts and NARIDAS, as well as inter-individual consistency between the different experts, is high but not perfect. If we use more abstract concepts like risk, we will be confronted with a higher degree of uncertainty than with crisp technical or physical parameters (e.g., ship's speed, course, position). In complex, dynamic processes like ship navigation, human operators will always have to cope with uncertainty. The concept of risk makes uncertainty measurable and visible. The positive expert ratings on user satisfaction in both studies suggest that practitioners believe they will profit from the display of risks by NARIDAS, despite the residual fuzziness of the risk concept.

For study II, NARIDAS was successfully implemented in the full-mission ship-handling simulator Elsfleth, so the system's operational capability could be demonstrated online in a dynamic setting. Experimental comparison showed positive effects of NARIDAS on situational risk awareness and navigation performance, even though the voyage scenario realized for the simulator study resulted to be not extraordinarily challenging for the well-trained participants. In the future, NARIDAS effectiveness should be tested under more tricky conditions, e.g. a slowly evolving emergency scenario in a simulator exercise of several hours. We assume that the benefits of NARIDAS should appear even clearer if the bridge team had to switch unexpectedly from operational routine to a peak workload situation. In study II, workload was rather moderate without major variations during the exercise, reflecting an everyday's working scenario.

Furthermore, the measurement of navigational risks by NARIDAS offers perspectives for various applications beyond the use as support tool for the bridge. In the ship-handling simulator, NARIDAS could provide online training feedback for the students as well as standardized assessments of navigation performance. Last but not least, dynamic risk assessments could be integrated into voyage data replay systems. So, incident and accident analyses would profit from risk profiles of critical situations, e.g. for a quantitative determination of the 'point of no return'.

References

1. Baker, C.C., McCafferty, D.B.: Accident Database Review of Human-Element Concerns: What Do the Results Mean for Classification? In: Proceedings of Conference on Human Factors in Ship Design, Safety & Operation. RINA, London, pp. 65–71 (2005)
2. Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors 37(1), 32–64 (1995)

3. Woods, D.D., Patterson, E.S., Roth, M.: Can We Ever Escape from Data Overload? A Cognitive Systems Diagnosis. Cognition, Technology and Work 4, 22–36 (2002)
4. The Nautical Institute: Alert! The International Maritime Human Element Bulletin 3 (April 2004) [visited 6 February, 2007] http://www.he-alert.org/
5. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A Model for Types and Levels of Human Interaction with Automation. IEEE Transactions on Systems, Man, and Cybernetics – Part. A: Systems and Humans 30(3), 286–297 (2000)
6. Näätänen, R., Summala, H.: Road-User Behavior and Traffic Accidents. North-Holland, Amsterdam (1976)
7. Kersandt, D.: ANRIS – Automated Navigation Risk Identification System. A Knowledge-based Early Warning System for Avoiding the Point of No Return. In: The Influence of the Man-Machine-Interface on Safety of Navigation. Proceedings of ISHFOB ’95 International Symposium of Human Factors on Board (1995)
8. Ergonomics of human-system interaction – Part 110: Dialogue principles
9. Ergonomic requirements for office work with visual display terminals – Part 11: Guidance on usability