Grootjen, M., Neerincx, M.A., Weert J.C.M. van, Truong, K.P., "Measuring Cognitive Task Load on a Naval Ship: Implications of a Real World Environment", *Proceedings ACI/HCII*, Beijing, 2007.

Measuring Cognitive Task Load on a Naval Ship: Implications of a Real World Environment

Marc Grootjen[1,2], Mark A. Neerincx[2,3], Jochum C.M. van Weert[3], Khiet P. Truong[3]

[1]Defense Materiel Organization, Directorate Materiel Royal Netherlands Navy, P.O. Box 20702, 2500 ES The Hague, the Netherlands
[2]Technical University of Delft, P.O. Box 5031, 2628 CD Delft, the Netherlands
[3]TNO Human Factors, Kampweg 5, P.O. Box 23, 3769 ZG Soesterberg, the Netherlands
Marc@Grootjen.nl, {mark.neerincx, khiet.truong}@tno.nl, jochum@gmail.com

Abstract. Application of more and more automation in process control shifts the operator's task from manual to supervisory control. Increasing system autonomy, complexity and information fluctuations make it extremely difficult to develop static support concepts that cover all critical situations after implementing the system. Therefore, support systems in dynamic domains should be dynamic as the domain itself. This paper elaborates on the state information needed from the operator to generate effective mitigation strategies. We describe implications of a real world experiment onboard three frigates of the Royal Netherlands Navy. Although new techniques allow us to measure, combine and gain insight in physiological, subjective and task information, many practical issues need to be solved.

1 Introduction

Designing highly complex interactive man-machine systems in process control has been subject of many studies. Due to ongoing automation, major changes in information technology have taken place, causing a radical change in the position of the operator from monitoring and control to supervision. Grootjen et al. [1] defined six main problem areas: (1) Increasing complexity; (2) Changing of information type and volume; (3) Increasing system autonomy causing 'out of the loop' operator problems; (4) Task integration from different domains; (5) Decreasing personnel and training budgets; (6) Increasing legislative constraints.

Human-centered design methods are often proposed to establish for the human a central and distinct position from other aspects of the system [2,3]. In the design of a new system, different configurations are tested and evaluated. Finally, the most efficient configuration which fits on a generic set of constraints will be implemented. However, a different configuration could be more efficient for a specific situation or subset of constraints, although it would be less efficient generally. Even if we follow a human-centered design method, the six defined problem areas make it extremely difficult to develop static support concepts that cover all critical situations after implementing the system. Therefore, support systems in dynamic domains should be dynamic as the domain itself. For example Niwa and Hollnagel [4] use dynamic mapping of alarm support of to enhance operator performance in nuclear power plants. Another example is the alarm handling support system of Dutch navy ships, which has different levels of automation [5].

A way of optimizing human-machine interaction by dynamic or adaptive automation is the use of augmented cognition. Augmented cognition symbiotically integrates man and machine in a closed-loop system. The machine's task is to detect both operator cognitive state and the operational context and to then dynamically adapt in real time in response, with the goal being to improve total man machine performance [6]. Taking the limitations of human information processing as a starting point, measurements of the cognitive state of the operator are used to trigger mitigation strategies to bypass or overcome the bottleneck e.g. [7-9]. Next to cognitive state measurements, a wide variety of data is proposed [10,11]. Our approach is multimodal sensing for complementary and more robust measurements. In Grootjen et al. [1] we describe a high level framework for HMC consistsing of four models (operator, task, system and context model), and an allocator which generates a work plan (i.e. mitigation strategy). This paper gives an overview which permanent and dynamic operator characteristics should be sensed and recorded in the operator model, classified in five categories (Section 2). Section 3 describes an experiment on board of three frigates of the Royal Netherlands Navy and shows which state-of-the-art sensing and fusing tools were applied. Section 4 presents implications and future directions, Section 5 the conclusion.

2 Operator model

An essential source of information for effective and efficient mitigation is the operator model. The operator model contains a large variety of information (see [1] for a literature overview), we propose classification in five categories (Fig. 1).

1. Permanent characteristics. This category contains fixed personal characteristics, which can be used to uniquely identify an operator. Characteristics like age and gender can affect a person's cognitive abilities and can be used in classification of operator groups.

2. Dynamic characteristics. This category contains characteristics that can change gradually over time. Like with some permanent characteristics, this information can be used for classification of operator groups.

3. Baseline state. This category contains baseline values of the operator's state.

4. Momentary state. This category contains real-time operator state measurements and task information (i.e. on which task the operator is working). Note that this information is also used to update the baseline state information.

5. Critical load areas: This category describes critical load areas using the CTL model of Neerincx [3].

The allocator uses information from the operator model to determine critical load areas, which are stored in the operator model itself. Furthermore, the actual real-time CTL is also calculated by the allocator. Comparing actual CTL and critical areas enables determination of a work plan.

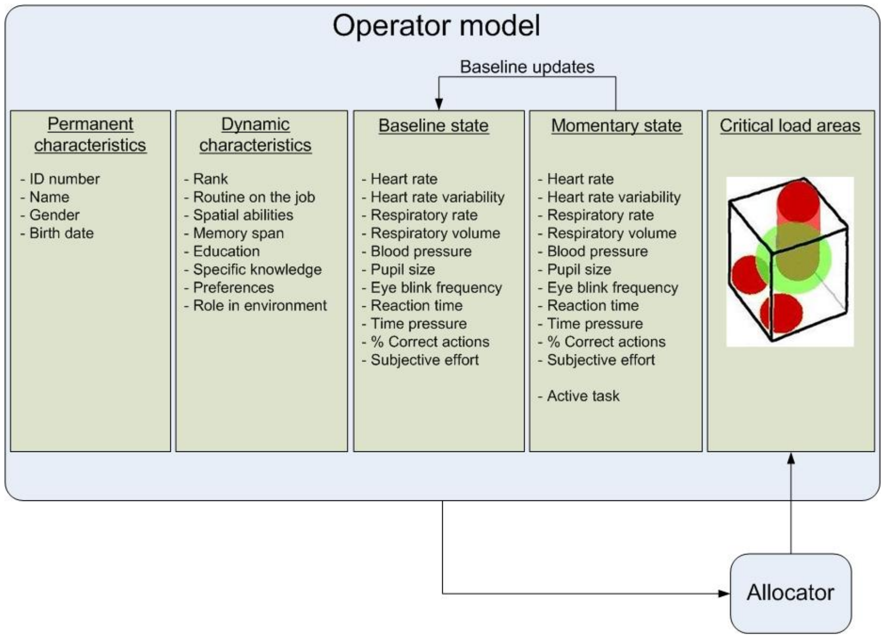

Fig. 1. Classification of operator information in five categories.

3 Real World Experiment

The Royal Netherlands Navy started in 1991 with the development of a new type of frigate: the Air Defense and Control Frigate (ADCF). In 2001 the first of four ADCFs was commissioned. For monitoring and control of all platform, energy and propulsion systems the ship has been equipped with an IMCS (Integrated Monitoring and Control System). During multiple sailing sessions on board of the ADCFs, the ship control centre (SCC) was evaluated using machinery breakdown drills. An important part of this evaluation was the collection of a wide variety of data for development of the operator model. Section 3.1 will focus on the method used to collect this data, the complete experimental setup used for the total SCC evaluation is out of the scope of this paper. Section 3.2 describes the organization of data afterwards.

3.1 Method

Three ADCFs participated in the experiment, each contained four active teams responsible for around the clock SCC operation. All teams consisted of four persons, a manager (sergeant), an operator (corporal), and two engine room operators (sailors). The participants had to deal correctly with the emergencies that appeared. In our experiment we focused only on the manager and the operator (12 of each in total, all males). The manager and operator used an IMCS console consisting of three screens, a keyboard and trackball for system operation. For communication purposes they had a voice terminal, several telephones and a separate push-to-talk engine room communication system. Fig. 2 shows the experimental setup.

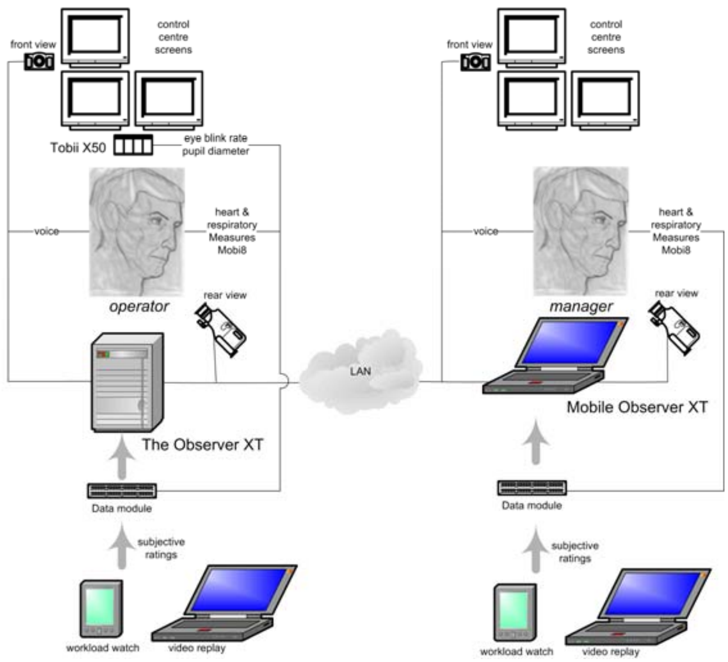

Fig. 2. The experimental setup of the operator (left) and manager (right).

For the operator and for the manager we installed a small spy camera (1/3" Sony super HAD color CCD) between the SCC screens to record a front view. On the ceiling we installed two digital cameras for the rear view. After each scenario, the participants had to do an individual video evaluation of the tasks just performed, using a laptop and handheld device. Each minute they were asked to rate task complexity and subjective effort on a five point scale (Fig. 3, based on the workload watch of Boer [12]. For heart and respiratory measures a Mobi8 recording device was used [13]. During the experiment we had one eye tracker available [14], which was used for the operator (Fig. 4, right). High quality voice recordings were made using 2 wireless headsets (Img Stageline HSE-210/SK), a receiver (Img Stageline TXS-890/TXS-890HSE) and amplifier (InterM PP-9214/R150Plus). Two Observer XT systems [15] were used to record all video and audio data. The external data module of the Observer XT system was used to import and synchronize data afterwards (Section 4). Individual systems were connected by a Netgear Fast Ethernet Switch FS608 v2.

The scenarios took place during the day or night, depending on the availability of the ship and working schedules of the participating SCC teams. The total time needed for one team was about 4 hours.

Fig. 3. Operator scoring complexity and effort during video replay of the just performed scenario (left and middle). Measuring pupil dilation (right).

3.2 Organizing Data

After a challenging period of collecting data at sea, the next exercise was organizing the data and prepare it for interpretation. To do this, we used the Observer XT of Noldus. This section will give an overview of the steps followed in organizing the data for the *operator*. Data for the manager was organized in a similar way, except for the eye tracker data. Fig. 4 shows part of the data in a screen dump of the Observer software.

Video and audio. With the Observer we recorded rear and front images of the operator. The rear recordings included low quality audio, the front high quality. Using this high quality audio we measured four raw speech features that are likely to correlate with stress (F0, SLOPE, COG and HI). See Section 5 for a short description. All features were extracted with Praat, a tool for speech analysis [16]. Furthermore, the front images were analyzed using Vicar Vision's FaceReader software [17]. To get the best results, we configured the FaceReader to do two frame-by-frame analyses. The first time we configured the FaceReader as if the spycam was positioned under the IMCS screen. The second time we took the default position. See Section 5 for a short description. Each analysis gives 7 output variables: happy, sad, angry, disgusted, surprised, scared and neutral.

Physiological Mobi data. We measured the expansion of the abdomen and the thorax due to breathing, the electrocardiogram and a sync signal. This data was downloaded from the flash card and analyzed with external software. Six variables were imported into the Observer: instantaneous respiratory rate (Hz), instantaneous heart rate (bpm), heart rate variability (0.04-0.15 Hz and 0.15-0.4 Hz), the sync signal and an artifact signal.

Subjective data. There are 3 subjective variables: complexity, effort and expert performance rating. All three variables were scored every minute. To import this data into the Observer, the variables had to be converted from single values every minute to a continuous block signal.

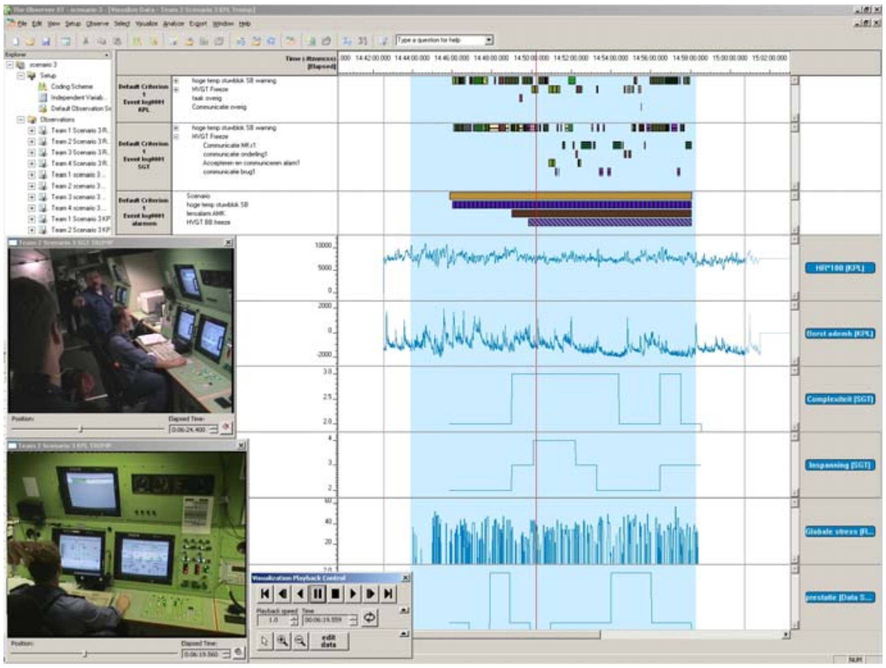

Fig. 4. The Observer XT (Noldus, 2006 192 /id). From top to bottom: three rows of task data, heart rate, respiratory rate, complexity, effort, voice stress and performance.

Tobii data. The eye tracker data was recorded on a separate computer. This data was exported with using Clearview [14], analyzed with MatLab [18] and imported into the Observer.

Task data. To obtain task data, videos were analyzed by experts using the observer software. Behavioral scoring schemes were composed according to the CTL method of Neerincx [3].

4 Implications & Future Directions

Observer XT. During and after the experiment the Observer XT was intensively used. We linked two Observer systems and the eye tracker so we could start recording all audio, video and the eye tracker by pressing one button. This saved us a lot of work synchronizing data streams after the experiment. Imported data using the external data module had to be synchronized manually. Having all data on the same timeline gives great insight, for example one can easily detect that missing eye tracker data often occurs simultaneously with high subjective effort. Clicking on an interesting part of a variable directly shows accompanying video images. However, such a big amount of data requires datamining techniques. We planned to import all data into the Observer, synchronize it and then export in again in one datasheet.

Unfortunately the export option was not working in this specific way, and will be implemented in a next version of the software. Future research will hopefully reveal the real meaning of this data.

Voice analysis. Many studies have tried to determine a general acoustic profile of stressed speech. In order to find a general acoustic stress profile, Scherer et al. [19] suggest different types of stress should be defined (e.g., emotional stress and cognitive stress). Although some acoustic features seem to correlate with stressed speech, a general "acoustic stress profile" has yet to be found. Most of the studies report higher values for speech rate, mean F0, mean intensity/energy, and the energy in the higher frequency regions of the spectrum in speech under stress [19,20]. We have measured four raw speech features that are likely to correlate with stress. First, we measure F0 which is expected to increase in speech under stress. The other three measures are based on the observation that the proportion of energy in the higher frequency regions increases in stressed speech. This can be measured in several ways: 1) by measuring the slope of the long-term averaged spectrum (SLOPE), 2) by measuring the centre of gravity (COG) and 3) by measuring the Hammarberg Index (HI) which is defined as the maximum energy below 2000 Hz minus the maximum energy between 2000 - 5000 Hz (HI). Thus, it is expected that F0, SLOPE and COG increase, while HI decreases in stressed speech. We experienced some difficulties during feature extraction that were due to the real world nature of the data: e.g., background noise like alarm signals, clipping of the signal by the ship's communication system and crosstalk.

Facial emotion. Automatic facial expression analysis usually consists of 3 tasks: 1) detection of the face, 2) feature detection and extraction and 3) classification of emotions. The best known method for facial data extraction is the facial action coding system (FACS, see [21]). The FACS system describes facial movements in terms of Action Units (AU). The system consists of a taxonomy of 44 AUs with which facial expressions can be described, and has attracted many researchers from the field of computer vision to develop automatic facial expression analyzers based on AUs. The automatic facial expression analyzer we have used, is based on Active Appearance Models and can classify emotions in six basic universal emotion categories. The FaceReader can only give output when the image quality is at an acceptable level, which is difficult given the real world conditions under which the data was recorded. Bad illumination, different poses of the head and background noises are some of the difficulties that the recognition software has to deal with. Fig. 5 shows classification of one image using the model where the camera is positioned under the screen. In contrast with this result, image quality is too poor for classification if we use the default model. Also due to background clutter, image quality can decrease making a classification impossible [22]. For future research we would recommend to reduce background clutter as much as possible, for example with a curtain. Furthermore, increase the amount of light in the face with a evenly distribution and zoom in as much as possible.

We did not yet analyze the seven output variables of the FaceReader for a stress related component due to export problems. As was stated earlier this Section, future research should reveal correlations.

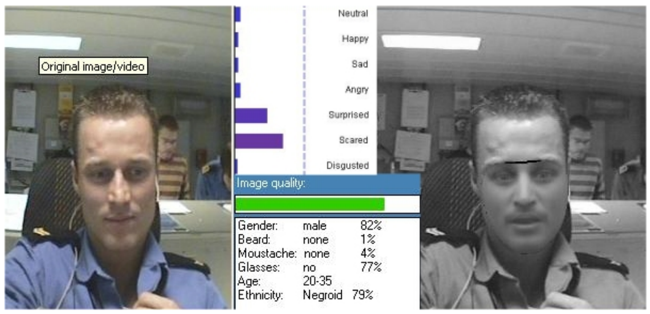

Fig. 5. Classification by the FaceReader [17] with a "below screen model". Using a default model for the camera position decreases image quality and makes classification impossible.

Heart and respiratory rate. During the experiment, the operator and manager wore the Mobi8. This is a small and light 8 channel physiological recording system. Data can be recorded on a flash card, or can be directly transferred using Bluetooth to a computer. Our initial plan was to use Bluetooth and link the system to the Observer so data would be automatically synchronized. Due to unknown problems with Bluetooth connection, we chose for recording with the more reliable flash card. At the start of every scenario the participants gave a synchronization pulse by pressing a button on the device, which could later be used when importing the data into the Observer. In future experiments we strongly recommend to use an automatic synchronizing feature.

Eye tracker. In the SCC, the participants had three screens available. Because the screen on top wasn't used very often, the eye tracker was configured for the two lower screens. Unfortunately, the operator moved around a lot. Communication devices, talking to colleagues or simply moving from the screen to relax or to the screen to read something, caused missing data. Because of the missing data, we were not able to subtract eye blinks from the pupil dilation information. In future real world experiments we would recommend a mobile, wearable eye tracking system. A multiple camera system could also work.

Follow up. We are planning the next experiment for spring 2007. For this experiment, experiences of real world experiments like this one and Grootjen et al. [23] and more controlled experiments [24,25] will be combined into an augmented cognition Closed Loop Integrated Prototype (CLIP). Although a lot preparation has already been done, some challenges are:

The allocator. The allocator should construct critical (and optimal) CTL areas, be able to generate a real time CTL operator signal and generate mitigation strategies. After analyzing data described in Section 3, we will select measurements and develop a method to fuse this data into one or two variables. With SOWAT [1] we will create CTL areas. Using information from other models (e.g. task model) should prevent us from an oscillating system [26].

Trust, situational awareness (SA) and locus of control. Being able to know when to mitigate is just one side of the story. Harkness-Regli et al. [27] state: "Adjusting non-optimal cognitive states can be at least as difficult as recognizing them." Our experiment will include a full usability evaluation. To improve trust and SA, we will evaluate usage of visual load feedback and a special SA display. We expect to increase trust and SA by giving the operator insight in the adaptive process. Furthermore, we will compare a situation in which the user is in charge of mitigation with one where the system is in charge.

5 Conclusion

Application of more and more automation, increasing system autonomy, complexity and information fluctuations make it extremely difficult to develop static support concepts that cover all critical situations after implementing the system. Augmented cognition offers a possible solution: symbiotic integration of man and machine in a closed-loop system. Although theoretical frameworks and methods are being developed [1,28,29] and controlled experiments have proven to increase performance [8], experience with real world environments is crucial in the development of new augmented cognition systems. This paper showed the implications of a real world environment during measurement and processing of data. Although new techniques allow us to measure, combine and gain insight in physiological, subjective and task information, many practical issues need to be solved.

References

1. Grootjen, M., Neerincx, M. A., and Weert, J. C. M. v., "Task Based Interpretation of Operator State Information for adaptive support", Foundations of Augmented Cognition 2nd Edition, 2006, pp. 236-242.
2. Stanton, N. A. and Baber, C., "The ergonomics of command and control",Ergonomics, Vol. 49, No. 12-13, 2006, pp. 1131-1138.
3. Neerincx, M. A., "Cognitive task load design: model, methods and examples," Handbook of Cognitive Task Design, edited by E. Hollnagel Handbook of Cognitive Task Design, Lawrence Erlbaum Associates, Mahwah, NJ, 2003, pp. 283-305.
4. Niwa, Y. and Hollnagel, E.. Enhancing operator control by adaptive alarm presentation. International Journal of Cognitive Ergonomics 5[3], 367-384. 2001. Lawrence Erlbaum Associates.
5. Mulder, J. M. F., "Supporting the internal battle", Thirteenth International Ship Control Systems Symposium (SCSS), 2003.
6. Kruse, A. A. and Schmorrow, D. D., "Session overview: Foundations of augmented cognition," Foundations of Augmented Cognition, edited by D. D. Schmorrow Lawrence Erlbaum Associates, Mawah, NJ, 2005, pp. 441-445.
7. Dickson, B. T., "Closed Loop Systems - Stability and Predictability," Foundations of Augmented Cognition, edited by D. D. Schmorrow Lawrence Erlbaum Associates, Mawah, NJ, 2005, pp. 617-620.
8. Dorneich, M. C., Whitlow, S. D., Mathan, S., Carciofini, J., and Ververs, P. M., "The communications Scheduler: A task scheduling mitigation for a Closed Loop Adaptive System," Foundations of Augmented Cognition, edited by D. D. Schmorrow Lawrence Erlbaum Associates, Mawah, NJ, 2005, pp. 132-141.
9. Diethe, T., "The Future of Augmentation Managers," Foundations of Augmented Cognition, edited by D. D. Schmorrow Lawrence Erlbaum Associates, Mawah, NJ, 2005, pp. 631-640.

10. Fuchs, S., Berka, C., Juhnke, J., Hale, K. S., and Levendowski, D., "Physiological Sensors Cannot Effectively Drive System Mitigation Alone", Foundations of Augmented Cognition 2nd Edition, 2006, pp. 193-200.
11. Schnell, T., Macuda, T., and Poolman, P., "Toward the "Cognitive Cockpit": Flight Test Platforms and Methods for Monitoring Pilot Mental State", Foundations of Augmented Cognition 2nd Edition, 2006, pp. 268-278.
12. Boer, L. C., "Workload-watch as an element of human engineering testing and evaluation", Eleventh Ship Control Systems Symposium, Vol. 2, Computational Mechanics Publications, Southampton, Boston, 1997.
13. TMS International. Mobi 8. 200. TMS International, Oldenzaal, the Netherlands.
14. Tobii Technology AB. Tobii. 2006.
15. Noldus. Observer XT. [6.0]. 2006.
16. Boersma, P., "Praat, a system for doing phonetics by computer", Glot International 5, Vol. 9/10, 2001, pp. 341-345.
17. Uyl, M. J. d. and Kuilenburg, H. v., "FaceReader: an online facial expression recognition system", Proceedings of the 5th international conference on methods and techniques in behavorial reseearch , 2005, pp. 589-590.
18. Mathworks. Matlab 7.1. 200.
19. Scherer, K. R., Grandjean, D., Johnstone, T., Klasmeyer, G., and Bänziger, T., "Acoustic correlates of task load and stress", Proceedings of ICSLP2002, 2002.
20. Rothkrantz, L. J. M., Wiggers, P., Wees, J. v., and Vark, R. J. v., "Voice stress analysis", International conference on text, speech and dialogue, Vol. 3206, 2004, pp. 449-456.
21. Ekman, P. and Friesen, W. V., "Facial action coding system: A technique for the measurement of facial movement", Consulting Psychologists Press, 1978.
22. Truong, K. P., van Leeuwen, D. A., and Neerincx, M. A., "Unobtrusive Multimodal Emotion Detection in Adaptive Interfaces: Speech and Facial Expressions", This Volume, 2007.
23. Grootjen, M., Neerincx, M. A., and Veltman, J. A., "Cognitive Task Load in a Naval Ship Control Centre: from identification to prediction",Ergonomics, Vol. 49, No. 12-13, 2006, pp. 1238-1264.
24. Grootjen, M., Bierman, E. P. B., and Neerincx, M. A., "Optimizing cognitive task load in naval ship control centres: Design of an adaptive interface", IEA 16th World Congress on Ergonomics, 2006.
25. Grootjen, M., Neerincx, M. A., and Passenier, P. O., "Cognitive task load and support on a ship's bridge: Design and evaluation of a prototype user interface", INEC 2002 Conference Proceedings, 2002, pp. 198-207.
26. Wilson, G. F. and Russel, C. A., "Psychophysiologically Versus Task Determined Adaptive Aiding Accomplishment", Foundations of Augmented Cognition 2nd Edition, 2006, pp. 201-207.
27. Harkness-Regli, S., Tremoulet, P. D., Hastie, H., and Stibler, K., "Mitigation Strategy Design for Optimal Augmented Cognition Systems", Foundations of Augmented Cognition 2nd Edition, 2006, pp. 208-214.
28. Nicholson, D., Stanney, K., Fiore, S., Davis, L., Fidopiastis, C., Finkelstein, N., and Arnold, R., "An Adaptive System for Improving and Augmenting Human Performance", Foundations of Augmented Cognition 2nd Edition, 2006, pp. 215-222.
29. Greef, T. E. d., Dongen, C. J. G. v., Grootjen, M., and Lindenberg, J., "Augmenting Cognition: Reviewing the Symbiotic Relation Between Man and Machine", Proceedings of ACI 2007, This Volume, 2007.