Nationaal Lucht- en Ruimtevaartlaboratorium
National Aerospace Laboratory NLR

NLR-TP-2004-272

Detection of workload elevation using different types of physiological measures for use in adaptive automation: some practical implications

P.J. Hoogeboom and L.J.M. Mulder*

* University of Groningen

This report has been based on a paper presented at HCI Aero 2004, Toulouse (France), 28 September - 1 October 2004.

Customer: National Aerospace Laboratory NLR
Working Plan number: AT.1.F.1
Owner: National Aerospace Laboratory NLR
Division: Air Transport
Distribution: Unlimited
Classification title: Unclassified
July 2004

Approved by author:	Approved by project manager:	Approved by project managing department:
6/8	6/8	11/8

Abstract

One of the problems for Air Traffic Control (ATC) operations is the maintenance of a sufficient level of Situation Awareness by the controller. On-line detection of vigilance decrements and provisions to help to overcome temporary detriments are necessary to maintain the current ATC safety record with increasing traffic density. In this respect a desktop experiment has been conducted with a simplified controller working position to verify real-time vigilance decrement detection algorithms. The algorithms combine controller performance and physiological information to derive the controller functional state, represented in the so-called Operator Status Model. This study indicates that real-time feedback of controllers' momentary state can be beneficial to increase overall ATC safety and the controllers' effectiveness.

Contents

(15 pages in total)

1 Introduction

With continuously increasing levels of air traffic (up to three times more aircraft movements in 2020 [Anon. 2001], the controllers' task-load, as measured by the number of aircraft handled over a given period of time, is expected to increase also. The "standard" approach to distribute the task-load over several controllers by assigning responsibilities for given sectors is reaching its limit due to the trade-offs like minimum sector-size and involved overhead in the required co-ordination between controllers. Therefore, alternative ways have to be found to cope with increased traffic demands. The continuous technological development (e.g. for computer and sensor capabilities) has almost reached a stage where work environments can be adapted to the human needs in real-time based on momentary operator status estimates from both task performance and physiological information [Hockey, Gaillard & Burov, 2003, Hoogeboom & Mulder, 2003, Wilson, 2003]. Given technology readiness and operational need, it is expected that the future Air Traffic Controller working positions will use adaptive automation to optimise system interactions in order to enhance efficiency, effectiveness and safety of the Air Traffic Management process. One of the key elements for the implementation of adaptive automation is the reliable estimation of the momentary status of the controller. This paper addresses the development of such human state estimators, and some of the encountered practical problems.

2 Adaptive Automation

Research in the field of adaptive automation started already in the late fifties, e.g. in the military domain, with the requirement to provide the pilot with the right information in the right format at the right time. Several attempts have been made since that time to progress the state of the art [Greenberg & Witten, 1985, Rouse, 1988, Parasuraman, Sheridan & Wickens, , 2000, Bennet, Cress, Hettinger & Stautberg, 2001, Haas & Hettinger, 2001]. Most were dealing with specific task-settings and were consequently highly specific. As an alternative, the approach chosen in the COMPANION project, co-financed by the Dutch Ministry of Economic Affairs, is generic, meaning that only the more generic attentive and physiological state of the user is estimated in real time. The automation processes verify the suitability of user involvement using information from this "functional state". Envisaged control items are the delay of information displays (wait till the user is ready to conceive the information), the display location (provide visual information at the location where the user is looking), attention getting (e.g. blinking fields and supporting auditory messages), the information modality (e.g. visual, auditory or haptic) and information format (e.g. graphical or textual). However, before being able to generate adequate feedback, the momentary state of the operator has to be estimated. This paper therefore addresses the development of real-time algorithms to detect small lapses and decrements in operator attention, and tries to categories the applied control strategy for a highly simplified air traffic control task.

3 Method

A group of 24 paid naïve participants (15 male, 9 female, age between 18 and 31) performed an experiment using a simplified Air Traffic Control (ATC) approach task [Hoekstra, 2003]. Each participant was trained with the task for 1 hour using an increasing level of task difficulty. None of them had problems with colour vision (e.g. discrimination between red and green colours).

The task itself consists of directing aircraft towards an extended runway centre-line (fig. 1). To this end, the participant selects an aircraft by mouse clicks. Heading directives can be issued by clicking at the appropriate point within the 'heading circle' surrounding the selected aircraft. Whenever the aircraft was within the intercept area ("topped-cone" area), an intercept clearance could be given after which the aircraft proceeded on its own towards the touch down point. Although provided by the simulation, speed and altitude clearances were not used in the experiment.

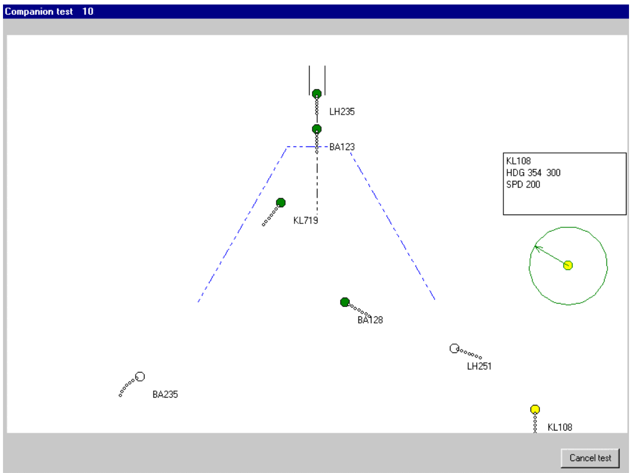

Figure 1: ATC task screen layout: aircraft entered the screen from the bottom (left and right) and were to be directed to the runway at the top of the screen. Heading directives could be given using the selection circle as shown on the right hand side of the screen. The relation between selected aircraft and the 'heading select circle' was colour coded (yellow). Imminent collisions (STCA) were coded red. Aircraft cleared for intercept and landing were coloured green.

In the experiment, each participant performed four scenarios: two hard and two easy ones. Each scenario lasted for 12 minutes. The difficulty of each scenario was controlled through the number of aircraft which entered the 'control area' including the entry point and initial heading. Two entry points - left and right at the bottom of the screen - were used.

The 'heading-select' circle could either be located surrounding the selected aircraft ("direct communication"), or could be in a fixed position at the right-hand side of the screen ("indirect communication"). The latter forced the participant to project the intended heading onto the circle away from the aircraft, and was therefore thought to lead to additional cognitive workload. The new heading ('heading clearance') could be selected by pointing within the circle at the appropriate location. Therefore, no cognitive translation of the intended heading into a numerical value was necessary.

In order to prevent aircraft flying off the screen, graphical feedback on the imminence of this condition was provided in half of the experimental conditions. The feedback consisted of graying the whole screen except for the location of the relevant aircraft, trying to get the focus of the participant towards the appropriate area.

Two control strategies could be used:

1. "Direct": half of the participants was given freedom to invent their own control strategy, which normally consisted of trying to get the aircraft within the intercept cone as fast as possible, followed by issuing an intercept clearance.
2. "Merging": the other half of the participants was instructed to merge aircraft coming from each entry point into a linear stream parallel to the extended runway centre line ('down-wind'), followed by merging the two resulting streams at the extended centre line including an 180 degrees heading change towards the touch down point. Finally, the intercept clearance could be issued. This 'double merging' procedure reflects to a certain extent the normal ATC control operation and allows a structured approach to the sequencing problem.

Participants were given also feedback in all experimental conditions on the imminence of a collision ('Short Term Collision Alert' or STCA) by highlighting the relevant aircraft with the colour red. In this case approximately ten seconds remained to de-conflict the aircraft through issuing the proper heading instructions. Each aircraft was simulated by a point-mass model with a heading change limit of 3 deg/s. Therefore, during a heading change the aircraft followed a more-or-less circular path until the new heading was reached. Thereafter the aircraft follows a straight line. This behaviour is comparable to the behaviour of real aircraft utilising a "rate-one turn" with approximately 30 degrees of bank. Note: in the experiment, no wind conditions were simulated.

During the experiment the location of all aircraft, status information like occurrence of warnings and number of aircraft-on-screen, the mouse position, selections, and physiological information of the participants was collected. Physiological information included ECG, respiration, EOG and eye-tracking based on the head mounted ASL501 (Applied Science Laboratories).

4 Results

4.1 Feedback for aircraft flying-off-the-screen

Feedback for aircraft almost flying-off-screen was perceived positively by the participants (based on verbal comments). No significant differences were found when the feedback was correlated to the number of aircraft 'about-to-leave' the screen or to the applied control strategy (direct or 'merging'). Figure 2, however, indicates for the case of the direct control strategy that the number of aircraft about-to-leave was reduced whenever feedback was provided. In this case 'about-to-leave' was defined as flying off the screen within 15 seconds if no action is taken (based on position, heading and speed).

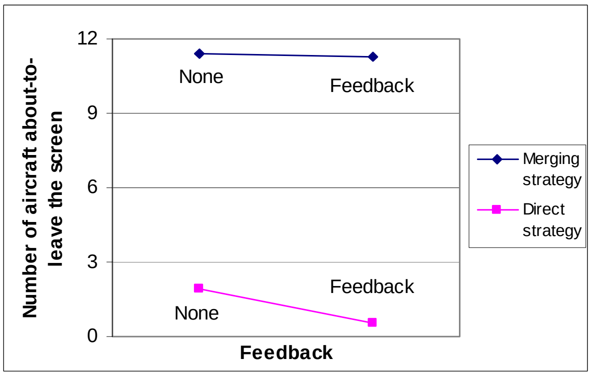

Figure 2: Effect of feedback (horizontal) versus number of aircraft 'about-to-leave' the screen. The lower line represents the direct control strategy.

The availability of feedback slightly increased also the average number of aircraft within the top corners of the screen, and led to a slight reduction of the number of aircraft flying-off-screen ('lost aircraft'). However it should be noted that the total number of lost aircraft was low, rendering the statistics to become unreliable. In summary, feedback for aircraft flying off the screen increased the number of aircraft in the upper screen areas, and simultaneously reduced the number of screen edge conflict warnings.

4.2 Workload

The envisaged workload dimensions were: 1. the traffic level, 2. direct/indirect communication (location of the heading selection circle), and 3. the applied control strategy. Of those dimensions, the Heart Rate Variability (HRV, 0.1 Hz frequency band) depended significantly on the applied control strategy[1] (p = 0.005), see fig. 3.Since the HRV reduces with increased workload, it can be concluded that the merging control strategy took more effort. The same was indicated by the average pupil diameter (p=0.004), even though the ambient light illumination levels differed considerably between participants, see fig. 4. (Please note that the pupil diameter increases with visual workload.) The communication method shows the expected trend (however, statistically not significant) in that direct communication reduces the visual workload.

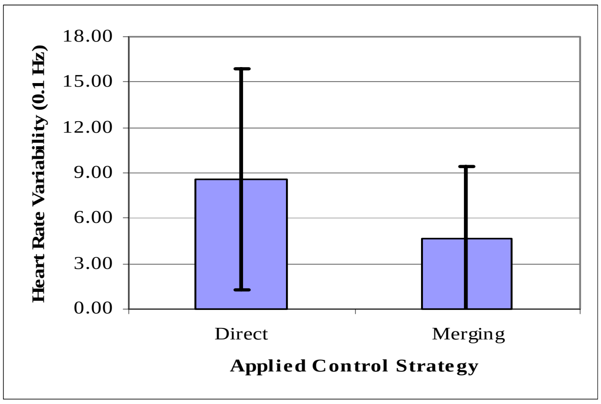

Figure 3: HRV versus control strategy

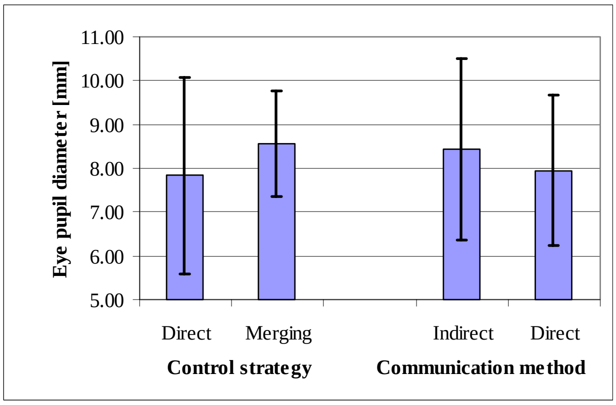

Figure 4: Pupil diameter

The influence of the control strategy was also visible from the average STCA duration, which increased significantly with the applied control strategy (p= <0.001). The average STCA duration was also significantly (p=0.001) influenced by the traffic level workload manipulation, in which higher traffic loads led to more frequently occurring and longer lasting STCAs.

[1] The data has been analysed according to the actually used control strategy, derived from the distribution of the aircraft over the screen area, instead to the instructed control strategy since many participants had difficulty in applying the merging control technique.

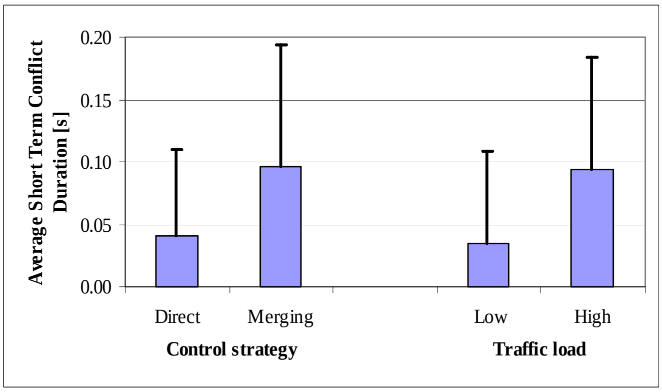

Figure 5: STCA versus control strategy & traffic load

It was observed that - opposed to instructions - some participants deliberately allowed conflicts (STCA, eventually followed by a collision) to reduce the number of aircraft on the screen, and as such reduced their workload.

The viewing distance - between eye and screen - was influenced by the control strategy and communication method, see fig. 6. The effects were statistically not significant, however. As expected, the viewing distance was reduced with increased taskload, as depicted by the control strategy. The direct communication method was thought to lead to a lower taskload (less mouse movements and avoidance of the mental projection of the heading vector), and as such the viewing distance should be larger for the direct communication method compared to the indirect method. However an opposite trend was found. A possible explanation for this discrepancy might be that the direct method allowed the participants to concentrate on a smaller part of the screen, and as such “delved into it”. Therefore, multiple factors may play a role in the interpretation of these measures.

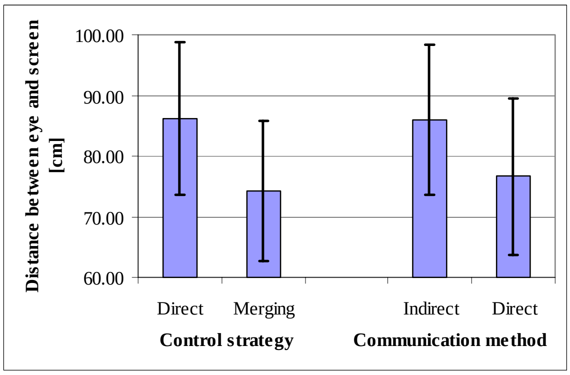

Figure 6: Viewing distance versus control strategy and communication method.

4.3 Eye tracker

As usual it was found that the eye-tracker calibration process required for accurate measurements is critical and requires specific expertise. Since in this experiment the participants had to select moving targets (the aircraft under control), a high correlation between the mouse position and the Eye Point Of Gaze (EPOG) locations can be expected at the moments an aircraft gets selected, see fig. 7 for an example. In this case the participant used the direct control strategy. The correlation between the two different measures is obvious, however, a closer examination reveals that the two measures use different grids. Also some skewness for the translation between the two measurements can be observed. Although this figure represents a typical example, there are subjects with lower correlations between the registered eye movement pattern and the mouse clicks. This stresses the necessity of an adequate calibration procedure of the eye tracking device.

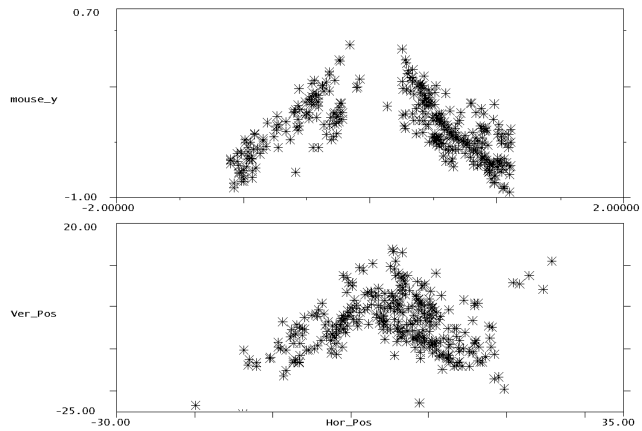

Figure 7: Mouse click positions (top graph) versus the EPOG locations (bottom graph) during the selection of aircraft.

5 Interpretation

Based on the feedback results, it can be expected that the participants were motivated to avoid the feedback message, as opposed to the intention to only respond to the timely warning of a developing situation requiring immediate attention. Therefore, the result of the message was as intended (less aircraft lost), but the interaction method (message avoidance instead of assistance) was opposed to the design intention. On the other hand, the ‘safety-net function’ led to the increase in the number of aircraft in the corners of the screen, which is an indication that the participants used the available space more effectively.

For the traffic load condition, a discrepancy existed between the observed changes in workload indices derived from the pupil diameter and heart rate variability versus results obtained from the average STCA duration. This difference can be explained from the observation that the participants invested all their cognitive energy irrespective of the difficulty level. Therefore, the physiological effects are the same for each workload level, but performance results (e.g. in the form of total STCA duration, or the number of successfully landed aircraft) may differ considerably. This observation has been found also in other experiments in which subjects were able to adapt their own workload (Mulder & Mulder, 1987) and is a clear indication for the motivation of the participants. In day-to-day operations, humans normally strive to acceptable results (instead of the best) and hence the physiological indices are expected to provide a better indication than the task results. The conclusion is that physiological measures should always be combined with performance measures in order to derive a reliable task-load indication.

Looking at the variability of the physiological measures between subjects, it is clear that a calibration procedure should be applied. Within subject variation of the measures provides a clear indication of mental workload changes. Unfortunately, the absolute level of the measures like HRV and pupil diameter can not easily be associated with a “percentage on the workload scale”.

The results from the eye-tracker measurements also lead to the need of having an option of ‘dynamic calibration’, meaning that the EPOG measurements are automatically compared to mouse-click positions or other position-on-the-screen related performance observations. In our experiment a strong relation between the EPOG and observed mouse positions was present for those instances where the participant made a selection (mouse-click). The underlying cause for this relation was the fact that the targets (aircraft) were moving on the screen in combination with a relatively small selection area. However, having such a relation, it can be used in reverse to calibrate the EPOG data to match more closely to the mouse-click positions. Since the calibration information is almost stationary (will only change due to helmet and visor shifts, which normally occur stepwise e.g. due to participants scratching their head), the derived information can also be used for the intermediate points. Since the conversion uses an offset, gain and skewness (together shaped like a trapezoid), a likely transfer function takes the form:

$$EPOG_{hor} = a_h.X + b_h.X.Y + c_h.Y + d_h$$
$$EPOG_{ver} = a_v.X + b_v.X.Y + c_v.Y + d_v$$

In which X and Y are the observed screen co-ordinates, a, b, c and d the calibration constants for the observed EPOG co-ordinate (respectively the horizontal or vertical channel). Each EPOG channel uses its own set of calibration constants.

Unfortunately, the given relationship is non-linear and no analytical solution for the required inverse exists: for the dynamic calibration it is required to compute the screen X and Y positions given the EPOG values. Therefore an iterative approach has to be used to solve the equation for X and Y given the $EPOG_{hor}$ and $EPOG_{ver}$ observations. Since sometimes multiple acceptable X,Y solutions may exist (e.g. both solutions being on the expected surface), the solution closest to the previous determined screen position is selected. This selection is based on the assumption that the eye moves in relative small steps from one fixation point to another and will not jump from one side of the screen to another. However, it is known that this assumption will not always hold since large saccades are known to occur once in a while. The development and refinement of this dynamic calibration method for eye-trackers is still ongoing: initial results are promising.

6 Conclusions

An experiment using 24 naïve participants has been conducted with the aim to obtain performance and physiological data for the further development of algorithms to analyse physiological signals in real-time. In this (and other) experiments a high dependency on participant characteristics has been found for the interpretation of physiological information. This has led to the need to calibrate physiological information before it is to be used to draw conclusions. For example, the interpretation of heart rate variability measures requires either reference periods or interpretation of the measures in a relative way (e.g. the participant is now heavier loaded than in the previous period). This notion has been extended in this experiment towards the use of overlapping information to dynamically calibrate eyetracker Eye Point of Gaze (EPOG) measures. In this case moving target selections with a pointing device (mouse) could be used to pinpoint where the user was most likely looking at. Having such a dynamic calibration facility would reduce (and optionally remove) the burden of having to calibrate the participant before the actual use of the equipment, and would also allow to detect the need for re-calibration whenever the participant touches the head mounted assembly. In total the practicality of using head-mounted eye-tracking equipment would be dramatically improved for day-to-day usage in divers environments. Or more generally, sometimes it is possible to use information from different sources in a dynamic setting to calibrate specific physiologic measures. Moreover, in some cases such task situations may be obtained by producing adequate task scenarios.

Another lesson from this study is that sometimes the behaviour of measures is different than expected due to less obvious causes. For instance, having the heading selection area coincide with the moving aircraft led to a reduction of the viewing distance (between eye and screen).

This reduction is normally taken to be an indicator of increased workload. Simultaneously, the eye-pupil diameter decreased which is an indicator for visual workload reduction under the assumption that the ambient light conditions for the eye remain the same. This latter behaviour was expected since having the heading selection area coincide with the aircraft leads to a lower use of the short-term memory. Note: the reduction in viewing distance may lead to a slightly increased illumination of the eye (the computer screen provides light) and therefore a smaller pupil diameter. In this case, however, the screen contrast with the room background lighting was very small. Therefore, the viewing distance change (approximately 10%) did not significantly increase the amount of light as received by the eye.

The difference in the applied control strategy was reflected in both the spatial distribution of aircraft and in the participants' workload. However, this observation can not easily be applied to real Air Traffic Control since the experiment provided only a simplified set-up, used naïve participants, and, apparently, some of the participants regarded the experiment as a computer game. Therefore, normal safety rules and procedures were not obeyed. An example of the last is that participants allowed aircraft to collide in order to reduce their workload, which would be unacceptable in real-life situations.

The applied control strategy observations also led to the conclusion that the participants' briefing and training was found to be insufficient to ensure correct behaviour: some of the participants used the direct control strategy instead of the requested merging method. Therefore, it is envisaged that in this kind of experiments some inherent motivation should be provided, stimulating required participants' behaviour. An example of the latter could be the inclusion of a performance score counter, which increments with appropriate behaviour (e.g. the number of successfully landed aircraft), and decrements significantly with inappropriate events like aircraft lost in space (flying off the screen), STCAs and aircraft collisions.

7 References

1. Anonymous (2001) European Aeronautics: a vision for 2020. Report of the group of personalities, https://europa.eu.int/comm/research/growth/aeronautics2020/en/
2. Bennet, K.B., Cress, J.D., Hettinger, L.J. and Stautberg, D. (2001). A theoretical analysis and preliminary investigation of dynamically adaptive interfaces. The International Journal of Aviation Psychology, II(2), pp169-195
3. Greenberg, S., and Witten, I.H., (1985). Adaptive personalized interfaces: a question of viability. Behavior and Information Technology, 4. pp32-45
4. Haas, M.W., and Hettinger, L.J. (2001). Current Research in Adaptive Interfaces. The International Journal of Aviation Psychology, II(2), pp119-121
5. Hockey, G.R.J., Gaillard A.W.K., and Burov, O. (2003). Operator functional state: the assessment and prediction of human performance degradation in complex tasks. Nato Science Series. Amsterdam: IOS press.
6. Hoekstra, R. "Een stap richting een adaptieve interface" *(Dutch: One step towards an adaptive interface)*, NLR CR-2003-135, Feb 2003
7. Hoogeboom, P.J, and Mulder, L.J.M., (2003) Physiological indices for the estimation of momentary changes in cognitive workload and mental state. HFES Conference Proceedings, Lund.
8. Mulder, L.J.M. & G.Mulder, (1987) Cardiovascular reactivity and mental workload. In: R.I. Kitney & O. Rompelman (Eds.), The beat-by-beat investigation of cardiovascular function: measurement, analysis, and applications, Clarendon Press, Oxford
9. Parasuraman, R., Sheridan, T.B., and Wickens, C.D. (2000). A model for types and levels of Human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans. Vol. 30, no 3, pp286-297
10. Rouse, W.B. (1988). Adaptive aiding for human/computer control. Human Factors 30, 431-443.
11. Wilson, G.F. (2003) Operator functional state assessment as a critical component of automated system operation. RTO symposium "Critical design issues for the human-machine interface", Prague, Czech Republic

ACKNOWLEDGMENTS

This study is supported by the Dutch Ministry of Economic Affairs under contract number IOP-MMI99002A&B.