Implementing dynamic changes in automation support using ocular-based metrics of mental workload: a laboratory study

Serena Proietti Colonna[1], Claudio Capobianco[2], Simon Mastrangelo[2],
Francesco Di Nocera[1]
[1]Sapienza University of Rome
[2]Ergoproject s.r.l., Rome
Italy

Abstract

Adaptive Automation has been often invoked as a remedy to indiscriminate introduction of automation support. However, this form of automation is difficult to implement without a sensitive and reliable index of the Operator Functional State. In a series of studies we have showed the usefulness of the distribution of eye fixations as an index of mental workload to be used as a trigger of automation. Particularly, the distribution pattern was found to be sensitive to taskload variations and types, thus making it very appealing for designing adaptive systems. This approach seems to be valid and reliable, but a necessary step in this research program would be testing the effectiveness of automation driven by fixation distribution and its capability in reducing the workload. The present study is a first attempt to carry out such validation.

Introduction

In many work domains the introduction of automation can improve complex systems performance and reduce overall costs by limiting the intervention of human operators. This can be accomplished in several ways: for example, through the assignment of routine tasks to computer systems in order to relieve the operator from performing them, as well as by implementing automatic monitoring of a process in order to improve safety (Rouse, 1981). Automation could also be implemented for removing the operators from dangerous work environments and for operating in environments that are inaccessible to the humans (Sheridan, 1992).

However, a major challenge in automation design is function allocation, that is "what needs to be automated" and "to what extent" in order to optimise performance (Inagaki, 2003). Several models have been devised in order to answer those questions and for supporting automation design. Some accounts represent all-purpose taxonomies initially developed in specific research domains (e.g. Sheridan and Verplank, 1978), whereas others attempted to address the issue of function allocation in terms of its relation with human information processing (e.g. Parasuraman, Sheridan & Wickens, 2000). There is, however, a third question that

should be answered in order to properly allocate functions, which is "when to automate". Indeed, albeit technology is aimed at reducing mental workload and errors, human interaction with automated systems also result in paradoxical side effects: automating a task often lead to loss of situation awareness and to higher mental workload when the operator is asked to intervene into the ongoing operation, particularly if the operator has been confined to monitoring functions for prolonged periods. With that in mind, it would be desirable having more flexible forms of automation in which function allocation dynamically varies during system operation and it is matched to what has been recently defined as "operator functional state" (Hockey, Gaillard, & Burov, 2003). That would facilitate a positive trade-off between the benefits and costs of automation itself (Parasuraman and Wickens, 2008). This form of automation is usually called "adaptive" and represents a closed-loop system in which the state of the operator is constantly monitored and support is provided only when it is needed.

The quest of a trigger

Adaptive automation is difficult to implement without a sensitive and reliable index of mental workload. The choice of the index to use for triggering the system when the functional state of the operator significantly deviates from optimal levels is thus one of the most important issues both for research purposes and for effective implementation of dynamic function allocation. Spontaneous psychophysiological activity showing sensitivity to variations in mental workload (e.g. cardiovascular, cerebral and ocular activity) is commonly considered the best choice, because it provides the opportunity for steady monitoring (and control). Many efforts have been devoted by several research groups for finding viable methodologies. Only to name a few, indices of "engagement" obtained from continuous EEG (Pope, Bogart, & Bartolome, 1995) or neural networks integrating data from multiple psychophysiological measures (Wilson, Lambert, & Russell, 2000) have been tested as potential triggers for adaptive systems. However, there is still no agreement in the literature about which indicator to use.

Among the many valuable accounts in the literature, our research group has proposed the use of the distribution of eye fixations as an index of mental workload and a potential trigger for adaptive systems. Particularly, a statistical indicator of spatial dispersion (the Nearest Neighbour Index) has been repeatedly found to vary with taskload (Camilli et al., 2007; 2008; Di Nocera & Bolia, 2007; Di Nocera et al. 2006; 2007; 2014). The index is based on the ratio between minimum distances observed in the distribution of eye fixations and the minimum distance expected by chance. Fixations spreading appear to be associated to mental workload when taskload depends on the temporal demand, whereas fixations clustering would be associated to mental workload when taskload depends on the visio-spatial demand. This index seems to be valid and reliable, but a necessary step in this research program would be testing the effectiveness of NNI-driven automation support in reducing the workload. Particularly, an adaptive system based on NNI should: 1) activate when the index deviates from (a previously computed) baseline; 2) produce a corresponding change in the fixation distribution (back to baseline limits)

indicating a mitigation of workload; 3) deactivate when that state has been reached. The present study is a first attempt to carry out such validation.

Study

Participants. Nineteen individuals (9 females, mean age = 26.52 st. dev. =2.65) volunteered in this study. All participants were right-handed and had normal or corrected to normal vision.

Experimental setup. The X2-30 wide eye tracking system (Tobii, Sweden) was used for recording ocular activity and custom Matlab code has been developed for running this experiment. The Tetris game, a commonly known tile-matching puzzle videogame successfully endorsed in a variety of studies (e.g., Trimmel & Huber, 1998), was used as experimental task. The gaming platform was based on "matlabtetris" by Matt Figg[2]. The entire experimental package was developed using Matlab® 2013a along with Tobii Analytics SDK v. 3.0 and was composed by three modules: the Tetris game, the NNI suite and the NNI monitor. The layout and the graphics of the game were kept as minimalistic as possible in order to reduce spurious saccades. Ocular data sampling frequency was set at the maximum available rate (40Hz). The NNI suite was created after the ASTEF package code (Camilli et al., 2008), performed all the tasks related to spatial statistics and computed them in real-time in order to trigger the automation support. This module can compute NNI based on convex-hull or smallest-rectangle areas, with or without the Donnelly adjustment. The suite can also analyse data and generate time series to be used in successive statistical analyses. The NNI monitor (available to the experimenter for visual inspection during the recording) plots the ongoing NNI value.

Procedure. Participants were seated in front of a 17" display, at a distance of approximately 60 cm. The room was dimly illuminated only by the display. After calibration of the eye-tracking system, they underwent a practice run of the Tetris game at the same velocity of the real game for avoiding context effects in the subjective assessment (see Colle & Reid, 1998).

The version of the game used in this study acted as a common version of the Tetris game with the exception that in this case the game restarted from a blank screen (starting condition) each time the stack of Tetriminos (game pieces) reached the top of the playing area and no new Tetriminos were able to enter. This condition commonly denotes the end of the game, whereas in this very situation the game needed to go on until the end of the entire experimental session (10 minutes each session). The game was therefore immediately restarted when the Tetriminos (reached the top and the restart was scored as a loss: a performance indicator to be used as dependent variable (# of restarts).

---

[2] http://www.mathworks.com/matlabcentral/fileexchange/34513-matlabtetris

Automation support was implemented as a projection of the falling Tetrimonos ("ghost block") over the pieces lying at the very bottom of the game area. This is known to facilitate the proper positioning by providing a time gain to the player. This manipulation has been already used in one previous study (Di Nocera et al., 2006).

During the first 5 minutes of gaming the NNI baseline for each subject was computed in real-time and any NNI value greater than ±1 standard deviation was marked as "out of range". Data collected within this "calibration" epoch was not included in the analyses.

Three automation conditions were implemented: manual control (no automation support), self-paced automation support (subjects could activate/deactivate the ghost block at their ease), and adaptive automation (the ghost block appeared when NNI deviations from baseline occurred and disappeared right after the NNI values returned within baseline limits). Each condition lasted 15 minutes and the presentation order was balanced across participants.

Data analysis and results

Given the scope of this study, the dependent variable to employ should represent the effectiveness of automation support in producing a return to NNI baseline values after deviation. With that in mind, a composite variable (proportion of "inwards" after deviation) has been computed by dividing the number of consecutive minutes within the ±1 standard deviation interval by the number of total minutes within the ±1 standard deviation interval. This measure would represent the effectiveness of the automation support in keeping the individual within acceptable workload levels for a prolonged period. The variable has been computed for all conditions (manual, self-paced, adaptive), thus we should expect a lack of significant differences between conditions if the return to baseline is random and/or "physiological". Two ANOVA mixed designs were carried out using the proportion of "inwards" (system effectiveness) and the number of game restart (individual performance) as dependent variables. Condition (Manual vs. Self-paced vs. Adaptive) and Gender (Males vs. Females) were used as factors. The latter was included in order to control for differences between males and females in computer gaming performance (see American Association of University Women, 1998). Results showed a significant interaction Gender by Condition for the proportion of inwards ($F_{2,38}=3.10$, p=.056). Duncan post-hoc testing showed that the interaction was due to males showing greater proportion of inwards after deviation with the adaptive automation than with manual control and self-paced automation (p<.05; figure 1). Main effects of Gender ($F_{1,19}=6.90$, p<.05) and a tendency towards statistical significance for Condition (p=.08) were found for the number of restart. Females gaming performance was significantly worse than males' and Duncan post-doc testing showed that gaming performance with self-paced automation was worse than that in manual control and adaptive automation.
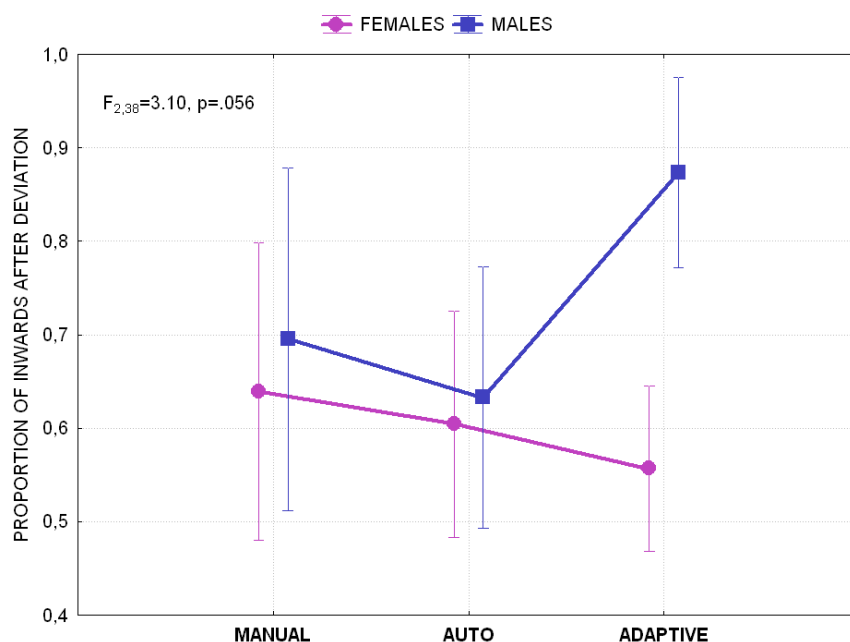
Figure 1. Proportion of inwards after deviation by condition and gender.

Discussion and Conclusions

An adaptive system should provide support for mitigating mental workload when the operator functional state is compromised and it should deactivate when the operator is back into "normal" functioning. In this study we have devised an experimental adaptive system based on the distribution of eye fixations. It was a rather basic system aimed at a laboratory investigation. The system was able to activate the automation support when the ocular index deviated from (a previously computed) baseline and to deactivate it when the index returned to baseline values. Changes in the index values obtained in the adaptive automation condition were compared to those occurring in the same task under manual control and self-paced automation support. Results showed a beneficial effect of the ocular-driven adaptive automation, but limited to male participants. Analyses carried out on gaming performance showed that females performed significantly worse than males in this task, thus suggesting that the gender difference found for the automation support should probably be considered a floor effect. Differential ability of males and females with visuo-spatial gaming and computing in general is well known (American Association of University Women, 1998) and in this case has been exacerbated by the absence of a proper training with the game prior to experimentation.

Interestingly, we found a detrimental effect of self-paced automation on performance. Apparently, performance in the adaptive automation condition

matches that obtained with manual control, although it was characterized by better workload management (as indicated by results on the proportion of inwards, even if limited to males). Self-paced automation appears to "get into the way", neither producing a mitigation of workload nor improving performance.

This was a first attempt in testing the potential of the NNI as a real-time trigger for automation. Moreover, these findings and the potential application of the technique are limited to those settings in which an operator seats in front of a display (e.g. Air Traffic Control). Results are far from being conclusive, but yet encouraging. One of the major flaws of the present study was lack of training with the task that probably affected female participants most. A replication of this study with a trained sample showing homogeneous performance levels is needed to disentangle the effect found.

Acknowledgements

References

American Association of University Women (1998). *Separated by sex: A critical look at single-sex education for girls.* Washington, DC: American Association for University Women Educational Foundation.

Camilli, M., Terenzi, M., & Di Nocera, F. (2007). Concurrent validity of an ocular measure of mental workload. In D. de Waard, B. Hockey, P. Nickel, and K. Brookhuis (Eds.), *Human Factors Issues in Complex System Performance* (pp. 117-129). Maastricht, The Netherlands: Shaker Publishing.

Camilli, M., Nacchia, R., Terenzi, M., & Di Nocera, F. (2008). ASTEF: A simple tool for examining fixations. *Behavior research methods*, *40*, 373-382.

Camilli, M., Terenzi, M., & Di Nocera, F. (2008). Effects of Temporal and Spatial Demands on the Distribution of Eye Fixations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52, 1248-1251.

Colle, H.A., & Reid, G.B. (1998). Context effects in subjective mental workload ratings. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *40*(4), 591-600.

Di Nocera, F., Terenzi, M., & Camilli, M. (2006). Another look at scanpath: distance to nearest neighbour as a measure of mental workload. In D. de Waard, K.A. Brookhuis, and A. Toffetti (Eds.), *Developments in Human Factors in Transportation, Design, and Evaluation* (pp. 295-303). Maastricht, the Netherlands: Shaker Publishing.

Di Nocera, F., & Bolia, R.S. (2007). PERT networks as a method for analyzing the visual scanning strategies of aircraft pilots. In *Proceedings of the 14th International Symposium on Aviation Psychology* (pp. 165-169).

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A random glance at the flight deck: pilot's scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making*, 1, 271-285.

Di Nocera, F., Proietti Colonna, S., Dessì, F., Capobianco, C., Mastrangelo, S., Steinhage, A. (2014). Keep Calm and Don't Move A Muscle: Motor

restlessness as an indicator of mental workload. In D. de Waard, K. Brookhuis, R. Wiczorek, F. Di Nocera, P. Barham, C. Weikert, A. Kluge, W. Gerbino, and A. Toffetti (Eds.), *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference* (pp. 183-191). Available as open source download. ISSN 2333-4959 (online).

Hockey, G.R.J., Gaillard, A.W.K., & Burov, O. (2013), Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks. Amsterdam: IOS Press.

Inagaki, T. (2003). Adaptive automation: sharing and trading of control. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design* (pp. 147–169). London: Lawrence Erlbaum Associates.

Parasuraman, R., Sheridan T.B., 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, 30 (3)*, 286–297.

Parasuraman R., Wickens C.D., (2008). Humans: still vital after all these years of automation. *Human Factors*, *50*, 511–520.

Pope, A.T., Bogart, E.H., & Bartolome, D.S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology*, *40*, 187-195.

Rouse, W.B. (1981). Human-computer interaction in the control of dynamic systems. *Computing Surveys*, 13, 71-99.

Sheridan, T.B., & Verplank, W.L. (1978). *Human and computer control of undersea teleoperators*. Massachuttes Instistute Of Technology. Cambrige: Man-Machine Systems Lab.

Sheridan T.B. (1992). Telerobotics, automation, and human supervisory control. Cambridge: MIT Press.

Trimmel, M., & Huber, R. (1998). After-effects of human-computer interaction indicated by P300 of the event-related brain potential. *Ergonomics*, *41*, 649-655.

Wilson, G.F., Lambert, J.D., & Russell, C.A. (2000). Performance enhancement with real-time physiologically controlled adaptive aiding. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 61-64.