EEG-based Mental Workload Recognition in Human Factors Evaluation of Future Air Traffic Control Systems

Yisi LIU[a, 1], Fitri TRAPSILAWATI[b], Xiyuan HOU[a], Olga SOURINA[a], Chun-Hsien CHEN[b], Pushparaj KIRANRAJ[b], Wolfgang MUELLER-WITTIG[a], Wei Tech ANG[b]
[a] *Fraunhofer IDM@NTU, Nanyang Technological University, Singapore*
[b] *School of Mechanical and Aerospace Engineering*
Nanyang Technological University, Singapore

Abstract. With growing air traffic density, air-traffic controllers (ATCOs) are facing more challenges in interpreting and analyzing air traffic information. As one of the solutions to this problem, automation supports such as tactile human computer interface, interactive 3D radar displays, and conflict resolution aid (CRA) are proposed for the enhancement of the current air traffic control (ATC) systems. To evaluate the proposed ATC systems, questionnaires are commonly used to get the feedback from ATCOs. However, the questionnaires are usually administered upon completion of each ATC simulation task thus provide only overall ratings towards the new ATC systems. In this paper, we propose and implement a novel Electroencephalogram (EEG)-based neurocognitive tools for evaluation of ATC systems. The nature of EEG-based technique is that the brain states can be monitored in a high resolution time, fitting the nature of time-critical ATC tasks. Thus, such EEG-based human factors study allows for real-time monitoring of ATCOs' mental workload during the task performance in ATC systems. We designed and conducted an experiment to evaluate the costs and benefits of the CRA and tactile user interface in future ATC systems. Thirty six participants participated in the experiment and were assigned into three groups of different display modes: Non-Display, Display, and Trajectory Prediction. In each group, three CRA conditions were given: Manual, Reliable and Unreliable. The EEG data were recorded during the tasks, and the traditional workload evaluation method NASA Task Load Index (TLX) was administered at the end of each task. The result shows that the ratings obtained from NASA TLX and from the EEG labeling are highly correlated. Thus, the EEG-based system is reliable to recognize workload during the task performance. With the proposed EEG-based system, we found that 1) relatively high workload was observed at the beginning of the experiment in each group which could be due to participants' unfamiliarity with the interfaces; 2) participants in the trajectory prediction group had much higher workload as compared to vertical display and non-display groups, which could be attributed to its complexity. The changes of workload are due to interaction between the type of display modes and time ($p<0.05$); 3) workload varied slightly with different CRA setting. The results show that the proposed EEG-based system for human factors study can provide better understanding of real-time mental workload changes during the task performance in new ATC systems, therefore enables the evaluation of current and future ATC systems.

Keywords. EEG, human factors, neuroergonomics, mental workload, air traffic control

[1] Corresponding Author. [LIUYS@ntu.edu.sg]

Introduction

Air Traffic Control is one of the key components in ensuring low occurrence of aviation safety incidents and it has remained 'human-centered' even until today. Human factors are an important element in air traffic control as Air Traffic Controllers (ATCOs) are responsible in coordinating aircrafts in the airspace. In addition, the need to provide information such as weather and traffic information as well as the need to provide assistance when it is requested by pilots, have made their job more demanding and stressful. From the statistics provided by Changi Airport Group (CAG), the number of air flights movements and commercial aircraft movements stands at 1.84 million and 341.4 thousand respectively [1]. Moreover, from the statistics, there is an upward trend in the number of aircraft movements in Singapore through the years [1] and we will not expect this trend to change in the near future. This poses a great challenge for air traffic control because safety risks will steeply increase when air traffic doubles.

There has been increase in the use of automation aids such as conflict alert and minimum safety altitude warning which enable controllers to take immediate action to eliminate disastrous accidents [2]. Moreover, more researches are dedicated to the evaluation and implementation of better automation aids such as Conflict Resolution Aid (CRA) [3] and Automatic Identification of Risky Weather Objects in Line of Flight (AIRWOLF) [4] which help to improve performance of controllers. It has been shown that workload contributes to a significant 24% to incidents [5] and workload will become even more relevant when air traffic density increases in the future. There are many different methods available in the evaluation of workload experienced by ATCOs. In this paper, the main focus will be the use of Electroencephalography (EEG) due to real time monitoring and its potential to provide better workload evaluation [6]. We conducted an experiment in the span of four months at Air Traffic Management Institute, Nanyang Technological University, which aims at assessing the conflict resolution automation and tactile user interfaces in future ATC systems. By using EEG, we monitored the brain states such as workload when the ATCOs were performing ATC tasks. Subjective evaluation methods such as questionnaires were also given to ATCOs upon completion of the tasks. However, this kind of method can only provide an overall evaluation instead of continuous monitoring.

The paper is constructed as follows. Section 1 presents review on common workload evaluation methods. Section 2 introduces the real-time EEG based mental workload recognition system. Section 3 gives the details of the experiment. Section 4 shows the results, and Section 5 concludes the paper.

1. Related work

The controller's workload is an important consideration in determining the maximum capacity of enroute air traffic [7]. Thus, there is a need for higher accuracy methods in evaluating ATCOs' workload which is able to provide a comprehensive feedback considering the different factors affecting workload. Usually, the measurements of workload can be subjective and objective.

1.1. Subjective workload evaluation

Subjective workload is based on the judgement and memory of a person's perception of his performance in a task [8]. However, there is a limitation to the accuracy of this method in workload estimation since numerous factors, such as a person's plan greatly affect the selectivity of consciousness [9]. Nevertheless, subjective workload measures are widely accepted within research and industrial domains due to low cost and ease of administrating procedures for such measures [10].

The common use of subjective workload measures for air traffic control research, is based on the use of ratings on unidimensional or multidimensional scale to determine the workload of an individual [11]. NASA-TLX is a well-known method which includes six factors such as mental, physical, temporal demands, frustration, effort, and performance. It assumes that the combinations of these factors can reflect the level of workload experienced when people are performing tasks [12]. NASA-TLX provides the highest sensitivity and obtains the highest acceptance among operators despite on it requires the longest time to complete [13].

1.2. Objective workload evaluation

Objective workload evaluation is often associated with the performance of the tasks including the primary and secondary task measures. The primary task measure estimates workload based on the capability of an individual in performing a primary task, while the secondary task measure estimates workload based on the capability of an individual to perform two tasks concurrently [14]. Better performance of secondary task could indicate that more residual capacity is available to it from the primary task, and the demand of cognitive resources of the primary task is also low [15]. Situational Present Assessment Method (SPAM) is associated with performance-based workload estimation via secondary task measure as well as situational awareness [16].

1.3. Psychophysiological workload evaluation

Psychophysiological methods in evaluating workload involve the analysis of the effect of physiological behavior due to changes in psychological variables [17]. Common psychophysiological methods used in workload evaluation for air traffic control include monitoring of cardiac activity, respiratory activity, eye activity, speech measures and brain activity [11]. EEG possesses many advantages in measuring workload due to its high time resolution, and it is considered to be the best among other workload evaluation methods that use eye tracking, pulse rate, etc. in human-machine interaction [18].

2. EEG-based workload recognition

We presented the real-time EEG-based brain states monitoring system CogniMeter in [19] which can identify brain states such as workload, emotion, and stress. In this study, the same algorithm which employs fractal dimension [20] and statistical features [21] are used to recognize ATCOs' mental workload. The Support Vector Machine (SVM) is used as the classifier. The algorithm is validated in [22] and the best accuracy is

90.39% for 2 levels mental workload recognition and 80.09% for 4 levels mental workload recognition. Since the SVM-based machine learning is a supervised one and the proposed mental workload recognition algorithm is subject-dependent one, a calibration is needed to train the classifier model. During the calibration, stimuli are given to the user to elicit different levels of workload and the EEG data are recorded. Fractal dimension and statistical features are then extracted. Together with the self-assessment labels indicating the workload levels, the features are fed into the SVM classifier to train the model. In the real-time recognition phase, the processing steps also include feature extraction and classification. Same features as in the calibration phase are extracted. Then these features are used as the input to the SVM model obtained from calibration. Finally, the current workload level can be identified.

3. ATC Experiment

3.1. Experiment settings

An experiment was carried out to study human factors in ATC work place. There were a total of 36 ATCO participants from the Civil Aviation Authority of Singapore (CAAS) and the Republic of Singapore Air Force (RSAF) controllers and students with prior knowledge in ATC. The participants were divided into three groups in this experiment: non-display, vertical display and trajectory prediction group. Besides the main radar display and two pseudo pilot display, a secondary touchscreen display, was used for either vertical display or trajectory prediction depending on the set-up. The vertical display provides information on the current and predicted flight level of aircrafts against time. The trajectory prediction provides information on the current and predicted flight level, plan view, speed (knots) and rate of descent and climb against time. No touchscreen display was available in the non-display group. Each group had to perform ATC tasks under three conditions: without Conflict Resolution Aid (CRA), with reliable CRA, and with unreliable CRA. Under the condition of reliable CRA, all conflict resolution provided by the system had perfect accuracy in resolving potential conflict, whereas under the condition of unreliable CRA, conflict resolution advisory provided by the system may not be able to resolve a potential conflict. The experiment session under each condition lasted for 1 hour.

3.2. Data Collection

Two types of data were collected in this experiment: the EEG data recorded during calibration and each session and NASA-TLX data collected after each session.

3.2.1. EEG Data

A wireless Emotiv EPOC is used for the collection of EEG signals in this experiment. The Emotiv EPOC is mounted on the head and has 14 channels (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4) following the international 10-20 electrode system with CMS/DRL references. The sampling frequency of Emotiv EPOC is at 128 Hz with a bandwidth between 0.16 Hz to 43 Hz [23]. As a calibration is needed, a series of Stroop Colour-Word Test is given as the stimuli to evoke different levels of workload and the EEG data are recorded at the same time. Together with the data obtained during performing of the ATC tasks, these two sets of data are used as the input to the EEG-based mental workload algorithm introduced in Section 2.

3.2.2. Questionnaires

Participants were also required to complete two sets of questionnaires after each task. The first one is NASA-TLX. Besides NASA-TLX questionnaire, we also include one more workload rating which is usually used in the proposed EEG-based system to label EEG data (ranges from 1 to 9 for low to high workload).

4. Results

In our previous study [24], we showed that results of the method used for labeling of the EEG data with workload levels is significantly correlated with NASA-TLX workload ratings results, and the EEG-based workload evaluation is reliable to be used in the recognition of workload level in the ATC tasks. In this paper, we continue the analysis using the continuous workload levels recognized from EEG.

For real-time continuous workload analysis, three levels of workload were used for a trade-off between higher accuracy and wider spread of the mean. The three levels of EEG-based recognized workload for each of the one-hour session was split into the interval of 5 minutes with an average taken for each 5 minutes interval.

We first apply mixed ANOVA test to the collected data, where "time" is the within-subjects factor, display mode and CRA conditions are the between-subjects factor. The results are given in Table 1. The factor of time ($p= 0.025$) and the interaction between time and display mode ($p= 0.04$) were significant (both p values < 0.05 as shown in Table 1). It can be concluded that ignoring the CRA setting and display mode, the workload values measured through the 12 time intervals are significantly different. Likewise, ignoring the CRA setting, the effect of display modes on workload is significantly different for different time intervals. The focus will be placed on the analysis of the effects of automation aids on workload at different time intervals. To analyse the effects of automation aids on workload, the mean and standard error of the workload measured for different time intervals are shown in Figure 1, for display modes and time intervals are shown in Figure 2, and for CRA settings and time intervals are plotted in Figure 3.

Table 1. Mixed ANOVA test for within-subjects effect analysis.

Factors	**F**	**p**
Time	3.493	0.025*
Display Mode * Time	2.438	0.04*
CRA * Time	1.914	0.083**

*Significant at α =0.05; **Significant at α =0.1

In the red box of Figure 1, it can be observed that the mean workload is relatively high at the start of the session and from the 20th minute to the 45th minute. The significance value obtained from the mixed ANOVA test for time factor is 0.025, which means that ignoring the CRA setting and display mode, the workloads measured through the 12 time intervals are significantly different. The workload was highest in the middle of the simulation task. This finding could be due to higher number of aircraft after ramp-up period. Afterward, however, the workload was getting lower along the time as ATCOs became more familiar with the aircraft and airspace structure.

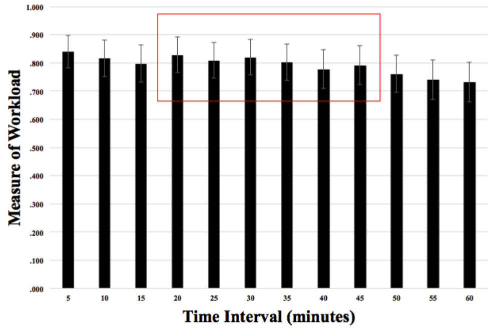

Figure 1. Measure of workload ignoring the CRA setting and display mode in 12 time intervals.

In the red box of Figure 2, it can be observed that the mean workload measured is the highest under the trajectory prediction and similar for the display and non-display mode. It can be associated with the information overload in the trajectory prediction where ATCOs also need time and effort in interpreting the complete information regarding aircraft status.

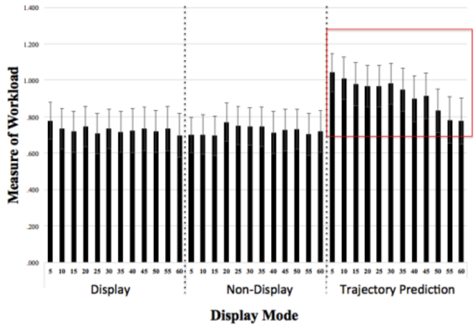

Figure 2. Measure of workload based on display mode Non-Display (ND), Display (D), and Trajectory Prediction (TP) in 12 time intervals.

In Figure 3, the mean workload obtained from the interaction between CRA setting and time intervals similar to the trend observed in Figure 1 except that the mean workload at the start of the unreliable CRA setting is relatively low, as shown in red box of Figure 3. In addition, no physiological workload difference was observed across different CRA conditions, indicating that the CRA brings neither benefit nor cost on ATCOs' physiological workload. In [25], lower objective workload as indicated by shorter ready response latency in SPAM was found under the CRA conditions as compared to manual condition. Here, the results of our physiological-based workload measurement contradict to this finding in [25]. However, the subjective workload was not statistically different across the CRA conditions in [25], which is consistent with our EEG-based results. We can infer that the physiological workload was in line with the subjective workload but not with the objective workload. This fact deserves further investigation on different workload measurement methods.

EEG-based workload evaluation method includes the continuous measurement of a subject's workload, which allows a more detailed analysis of experiment results obtained. From the current ATC study, it could be found that the complexity of trajectory prediction has resulted in a higher workload compared to other display aids in all time points, and the reliability of CRA has minimal effect on workload. Therefore, the trajectory prediction aid needs to be redesigned to meet the aim of reducing workload experienced by air traffic controllers.

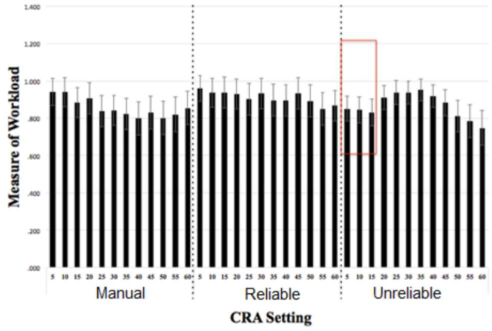

Figure 3. Measure of workload based on the CRA settings Manual, Reliable and Unreliable in 12 time intervals.

5. Conclusion

In this paper, we presented the EEG-based mental workload recognition for human factors study in ATC systems. An experiment was carried out with three display groups and each group completed the ATC tasks in three CRA conditions. We analyzed the continuous workload changes in a 5 minutes interval. It was discovered that the workload was highest in the middle of the simulation tasks. In addition, workload was the highest with the trajectory prediction. This could be due to information overload that was encountered by ATCOs. This opens a possibility to further investigate the provision of the trajectory prediction display given a longer training period to familiarize ATCOs with the new display. Lastly, no physiological workload difference was found across the CRA conditions.

Therefore, by utilizing the proposed EEG-based system, true understanding of ATCOs' working pattern along the time can be obtained. Based on the analyses of the objective recognized brain states together with the subjective feedback from ATCOs, we are able to reliably evaluate current ATC systems and refine new concepts of future ATC system. However, the further analysis between EEG and objective workload measurement methods deserves continuing investigations.

Acknowledgments

This research was supported by Civil Aviation Authority of Singapore (CAAS) and Air Traffic Management Research Institute (ATMRI) Project ATMRI: 2014-R5-CHEN and by the National Research Foundation, Prime Minister's Office, Singapore under its International Research Centres in Singapore Funding Initiative.

We would like to acknowledge the final year project students of School of MAE of Nanyang Technological University and personally Sim Jun An for their contribution in this work.

References

[1] Civil Aviation Authority of Singapore, "LEADING THE HEART OF AVIATION," Singapore: Civil Aviation Authority of Singapore, 2014, p. 102.

[2] K. Allendoerfer, F. Friedman-Berg, and S. Pai, "Human Factors Analysis of Safety Alerts in Air Traffic Control," Federal Aviation Administration, National Technical Information Service (NTIS) 2007.

[3] F. Trapsilawati, X. Qu, C. D. Wickens, and C.-H. Chen, "Human factors assessment of conflict resolution aid reliability and time pressure in future air traffic control," *Ergonomics,* vol. 58, pp. 897-908, 2015.

[4] U. Ahlstrom, "Experimental Evaluation of the AIRWOLF Weather Advisory Tool for En Route Air Traffic Controllers," *Aviation Psychology and Applied Human Factors,* 2015.

[5] T. Edwards, S. Sharples, J. R. Wilson, and B. Kirwan, "Factor interaction influences on human performance in air traffic control: The need for a multifactorial model," *Work-Journal of Prevention Assessment and Rehabilitation,* vol. 41, p. 159, 2012.

[6] M. Z. Weiland, D. M. Roberts, M. S. Fine, and M. S. Caywood, "Real Time Research Methods Monitoring Air Traffic Controller Workload During Simulation Studies Using Electroencephalography (EEG)," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2013, pp. 1615-1619.

[7] A. Majumdar and W. Ochieng, "Factors affecting air traffic controller workload: Multivariate analysis based on simulation modeling of controller workload," *Transportation Research Record: Journal of the Transportation Research Board,* pp. 58-69, 2002.

[8] S. Robertson, *Contemporary Ergonomics 1996*: CRC Press, 2003.

[9] J. Annett, "Subjective rating scales: science or art?," *Ergonomics,* vol. 45, pp. 966-987, 2002.

[10] W. Karwowski, *International encyclopedia of ergonomics and human factors* vol. 3: Crc Press, 2001.

[11] S. Miller, "Literature Review: Workload Measures," *The University of Iowa,* 2001.

[12] S. G. Hart, "NASA-task load index (NASA-TLX); 20 years later," in *Proceedings of the human factors and ergonomics society annual meeting*, 2006, pp. 904-908.

[13] S. G. Hill, H. P. Iavecchia, J. C. Byers, A. C. Bittner, A. L. Zaklade, and R. E. Christ, "Comparison of four subjective workload rating scales," *Human Factors: The Journal of the Human Factors and Ergonomics Society,* vol. 34, pp. 429-439, 1992.

[14] D. Damos, *Multiple task performance*: CRC Press, 1991.

[15] C. D. Wickens, J. G. Hollands, S. Banburry, and R. Parasuraman, *Engineering Psychology and Human Performance*, fourth ed. New Jersey: Pearson Education Inc., 2013.

[16] S. Loft, D. B. Morrell, and S. Huf, "Using the situation present assessment method to measure situation awareness in simulated submarine track management," *International Journal of Human Factors and Ergonomics,* vol. 2, pp. 33-48, 2013.

[17] P. Barrett and P. Sowden, "Psychophysiological methods," *Research methods in psychology,* pp. 160-173, 1995.

[18] J. Frey, C. Mühl, F. Lotte, and M. Hachet, "Review of the use of electroencephalography as an evaluation method for human-computer interaction," *arXiv preprint arXiv:1311.2222,* 2013.

[19] X. Hou, Y. Liu, W. L. Lim, Z. Lan, O. Sourina, W. Mueller-Wittig, *et al.*, "CogniMeter: EEG-Based Brain States Monitoring," in *Transactions on Computational Science XXVIII: Special Issue on Cyberworlds and Cybersecurity*, ed: Springer Berlin Heidelberg, 2016, pp. 108-126.

[20] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D: Nonlinear Phenomena,* vol. 31, pp. 277-283, 1988.

[21] Z. Lan, O. Sourina, L. Wang, and Y. Liu, "Real-time EEG-based emotion monitoring using stable features," *The Visual Computer,* pp. 1-12, 2015.

[22] W. L. Lim, O. Sourina, L. Wang, and Y. Liu, "EEG-based Mental Workload Recognition Related to Multitasking," presented at the 10th International Conference on Information, Communications and Signal Processing (ICICS), 2015.

[23] Emotiv. *http://www.emotiv.com*.

[24] X. Hou, F. Trapsilawati, Y. Liu, O. Sourina, C.-H. Chen, W. Mueller-Wittig, *et al.*, "EEG-based Human Factors Evaluation of Conflict Resolution Aid and Tactile User Interface in Future Air Traffic Control Systems," in *Advances in Human Aspects of Transportation*, ed: Springer, 2017, pp. 885-897.

[25] F. Trapsilawati, C. D. Wickens, X. Qu, and C.-H. Chen, "Benefits of imperfect conflict resolution advisory aids for future air traffic control," *Human factors,* vol. 58, pp. 1007-1019, 2016.