A Theoretical Model of Mental Workload in Pilots Based on Multiple Experimental Measurements

Zongmin Wei, Damin Zhuang, Xiaoru Wanyan, Huan Zhang, and Chen Liu

Department of Human-machine Environment Engineering, School of Aeronautic Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing, China
`weizongmin8411@163.com`

Abstract. The present study attempted to establish an effective discrimination and prediction model that can be applied to evaluate mental workload changes in human-machine interaction processes on aircraft flight deck. By adopting a combined measure based on primary task measurement, subjective measurement and physiological measurement, this study developed both experimental measurement and theoretical modeling of mental workload under flight simulation task conditions. The experimental results showed that, as the mental workload increased, the peak amplitude of Mismatch negativity (MMN) was significantly increased, SDNN (the standard deviation of R-R intervals) was significantly decreased, the number of eye blink was decreased significantly. Finally, a comprehensive mental workload discrimination and prediction model for the aircraft flight deck display interface was constructed by the Bayesian Fisher discrimination and classification method. The model's accuracy was checked by original validation method. When comparing the prediction and discrimination results of this comprehensive model with that of single indices, the former showed much higher accuracy.

Keywords: Mental workload, Human-machine interaction, MMN, SDNN, Eye blink.

1 Introduction

Mental Workload has always been deemed as an important factor that influences pilots' performances, because under excessive mental workload, the pilots may exhibit delayed response to some of the incoming information; while under inadequate mental workload, the pilots may lack vigilant awareness or even miss some abnormal flight information. Aircraft flight deck display interface, as an important human-computer interaction of providing flight information to pilots, its design should provide appropriate mental workload, otherwise may seriously affect flight safety.

Accordingly, there are multiple methods for the measurement and evaluation of mental workload on the aircraft flight deck display interface, among which NASA_TLX scale subjective evaluation method is mostly widely used. NASA_TLX scale has the advantage of comprehensive evaluation, and it takes six factors into consideration, i.e. mental demand, physical demand, temporal demand, effort,

D. Harris (Ed.): EPCE 2014, LNAI 8532, pp. 104–113, 2014.

performance and frustration level [1]. However, this method could only be conducted after the flight tests, which brings difficulties to the measurement and evaluation of aircraft cockpit human-machine interaction system design in earlier time, because once problems are detected, the system needs to be redesigned and the flight test and NASA_TLX scale subjective evaluation needs to be carried out once again, which puts great pressure on human and material resources. In addition, the evaluation results of NASA_TLX scale may be confusing due to the subjective differences among individuals [2].

Physiological evaluation method is another important measurement and evaluation method of mental workload on the aircraft flight deck display interface, which could provide real-time and objective physiological change of operators to measure mental workload. There are generally three categories of physiological evaluation methods: electroencephalogram (EEG), electrooculogram (EOG), and electrocardiogram (ECG) measurements. Former studies about flight simulation tests demonstrated that MMN index of EEG measurement [3], SDNN index of ECG measurement [4], as well as eye blink number of EOG measurement [5] could effectively reflect the mental workload level of flight, respectively.

However, current studies show that any single index or single method for the measurement and evaluation of mental workload has its own advantages and limitations [4]. Each index or method may provide useful information reflecting mental workload, but none could comprehensively reflect the mental workload under different task conditions. The multi-dimensional nature of mental workload makes each index sensitive to only one or several dimensions of mental workload, instead of all dimensions. That is to say, it is impossible for a single mental workload index attained from a single evaluation method to be appropriate for all different task conditions. Therefore, joint measurement and evaluation of a variety of indices of mental workload is seen as a more feasible approach and one of the future trends of mental workload evaluation.

The present study attempts to evaluate flight mental workload by synthetically combining above three categories of physiological measurements based on designed flight test tasks. It provides real-time and objective evaluation of the mental workload which the aircraft flight deck display interface puts on the pilots under different flight task conditions, and discriminates the mental workload level based the Bayesian Fisher discrimination analysis method, thereby guides the aircraft cockpit human-computer interaction systems mental workload task design. It is expected that the present study can be applied to the early design stage of aircraft flight deck display interface, and the designers can optimize the aircraft flight deck display interface by adjusting the mental workload task design, based on the discrimination and predication of mental workload.

2 Methods

2.1 Subjects

14 male flying cadets (range of age: 22-28 years old; mean of age: 24.6 years old) from Beijing University of Aeronautics and Astronautics participated in the present study. All subjects were right-handed with normal or corrected-to-normal hearing and

vision. Each subject was well trained to be good at simulated flight operations. After a complete description of the study, informed written consent was signed by each subject before the experiment.

2.2 Experimental Task

The experiment task was to monitor indicators based on a flight simulator. Subjects were required to accomplish the whole dynamic process of flight simulation, including take-off, climb, cruise, approach and landing. During the flight simulation process, they were asked to monitor the status of flight indicators presented on the Head up displays of simulation model, and recover the information state when abnormal information was detected, by pressing certain keys as quickly and accurately as possible. The simulation model was designed with reference to the typical information layout of Head up displays, and could display several kinds of flight indicators, including pitching angle, air speed, altitude, heading angle, rolling angle, and fuel status, as shown in table 1. Subjects' mental workloads were manipulated by adjusting the quantity of flight indicators and information refresh frequency: high mental workload was set as 6 flight indicators, 2s duration of abnormal information, and random inter-stimulus interval between abnormal information; low mental workload was set as 3 flight indicators, 1.5s duration of abnormal information, and random inter-stimulus interval between abnormal information.

Table 1. The scope setup for abnormal flight indicators

No.	Flight indicators	Abnormal scope
1	Pitching angle	Exceed 10'
2	Air speed	Exceed 400nautical mile/h
3	Altitude	Exceed 10000 feet
4	Heading angel	Exceed 50'
5	Rolling angel	Exceed 20'
6	Fuel status	Abnormal

2.3 Experimental Procedure

Within-subject factorial design was implemented in the experiment. The mental workload was divided into three levels, i.e. high, low and baseline. All the 14 subjects participated in the flight simulation experiments under the three levels of the mental workloads, respectively. The order of the high and low experiment tasks within the sessions was counterbalanced across participants to minimize the learning effect. In order to record the EEG, EOG and ECG data, all the participants were asked to wear EEG electrode cap, EOG electrodes and ECG electrodes throughout the experiment. After each session, each subject was instructed to take a 15-minnuts rest, meanwhile completed subjective evaluation of NASA_TLX (Task Load Index).

2.4 Experiment Data Recording and Analysis

Three types of indices were attained, including the subjects' performance evaluations (accuracy of the primary task and reaction time), subjective evaluations (the score of NASA_TLX), and three different physiological evaluations (EEG, ECG and EOG). By analyzing the changes of these different indices under different mental workload conditions, this study explored the sensitivity and diagnosability of these different indices to pilot mental workloads and provided a foundation for the design of display interface mental workload task.

Performance data recording and analysis. By computer programming, the system automatically recorded the two indicators of performance evaluation, including operation accurate rate and reaction time (the time interval between the detection of the abnormal information and correct responding).

Physiological data recording and analysis.FX-7402 12-channel automatic analysis of ECG machine was adopted to synchronously record the ECG signals. The ECG data recorded included the heart rates of subjects measured every 5 min, time series during R-R period, ECG within this period and the electrode placement arranged as per lead II. The heart rate value range was 20~300bpm, the heart rate detection accuracy was ±2bpm, the sampling frequency was 0.05-150Hz, and the waveform recording speed was 25mm/s. Relevant studies showed that, standard deviation of normal R-R intervals (denoted "SDNN"), one of the time-domain related indexes of HRV, could effectively reflect the sensitivity levels of mental workload[4]. Therefore, the present study analyzed the index of SDNN.

EOG signals were recorded from electrodes placed above and below the left eye. Relevant studies showed that eye blink number was closely related to mental workload level. Therefore, eye blinks data were analyzed in present study.

EEG signals were recorded from FZ electrode site using Neuroscan Nuamps Amplifier. Electrode placed at the forehead was grounding. Electrode impedances were maintained below 5KΩ. The recording band-pass was 0.05~100Hz and the sample rate was 1000 Hz. After correcting the eye movements, the epoch was set as 1300 ms, including a 200ms pre-stimulus baseline. Any epoch containing residual artifact voltages exceeding ±150μv was rejected. Relevant studies showed that the peak amplitudes of MMN of EEG measurement could effectively reflect the sensitivity of mental workload [3]. Therefore, the present study adopted the peak amplitudes of MMN of EEG。

Subjective data recording and analysis. NASA_TLX score was used for subjective analysis. For the convenience for the subjects to accurately and effectively complete subjective evaluation, in the present study, the NASA_TLX scale was presented in numerical value, i.e. score from 0 to 100, with 0 representing no effort and 100 representing maximum effort. First, a score (from 0 to 100) was obtained on each dimension according to the subjects' subjective feelings on the flight related mental workload. Then, a paired comparison task was performed for all pairs of the six dimensions, which required the subjects to choose which dimension had a greater relevance to the overall mental workload. After that, each of the six dimensions was given a specific weight according to the number of times each dimension was chosen in pared comparison and the six dimensions were sorted. The final mental workload score was got by multiplying each individual dimension scale score by its respective weight and dividing the total score of all dimensions by 15 (the total number of paired comparisons). Repetitive measure analysis of variance (ANOVA) was employed for the analysis of the above data by using SPSS 17.0 statistical package.

3 Results

3.1 Flight Task Performance Measurement

At two different mental workload levels (high and low levels), the accuracy rate and reaction time of subjects responding to abnormal flight indicators were shown in Table 2. Two (high and low) × two (accurate rate and reaction time) repeated measure analysis of variance (ANOVA) showed a significant ($P<0.001$) main effect of mental workload. As the mental workload increased, the performance level decreased significantly ($P<0.001$), which was specifically demonstrated by the successive decrease ($P<0.001$) of the accuracy rate of participates as well as the successive increase ($P<0.001$) of the reaction time at high and low mental workload levels, respectively.

Table 2. Flight task performances under the high and low mental workloads

Mental workload	High	Low
Accuracy rate /%	74.14±5.67	97.88±1.75
Reaction time /ms	862.47±52.67	809.18±68.52

3.2 Subjective Measurement

Results of NASA_TLX-based subjective evaluation were shown in Table 3. Result of the single-factor repeated measure ANOVA suggested a remarkable ($P<0.001$) main effect of mental workload. With the increase of mental workload, the subjective evaluation scores of NASA_TLX gradually increased ($P<0.001$).

Table 3. Subjective measurement under the high and low mental workloads

Mental workload	High	Low
NASA_TLX	65.39±5.27	57.10±4.78

3.3 Physiological Measurement

For the peak amplitudes of MMN at Fz, the main effect of mental workload was significant ($p=0.008$). As the mental workload increased, the peak amplitudes of MMN decreased significantly. The result of a further paired comparison suggested that, the peak amplitudes of MMN at high mental workload was significantly higher than that at low mental workload (P=0.035).

For the SDNN index, the main effect of mental workload was significant (p<0.001). As the mental workload increased, the value of SDNN decreased significantly. The result of a further paired comparison suggested that, the value of SDNN at high mental workload was significantly lower than that at low mental workload (P=0.013).

For the eye blinks index, the main effect of mental workload was significant (p=0.002). As the mental workload increased, the number of eye blinks decreased significantly. The result of a further paired comparison suggested that, the number of eye blinks at high mental workload was significantly decreased than that at low mental workload (P=0.003).

4 Modeling

4.1 Modeling Method

Based on the analysis results of experimental measurements, a comprehensive mental workload discrimination and prediction model for the aircraft flight deck display interface was constructed by the Bayesian Fisher discrimination and classification method. To ensure the comprehensiveness of the discrimination, the general discrimination analysis method (all-factor analysis method) was employed in the present study, i.e., the discrimination model included the peak amplitude of MMN, the value of SDNN, and eye blink number. The discrimination equations of the model included two discrimination functions, respectively representing two different mental workload levels. Substitute various index values obtained under the aircraft cockpit display interface successively into the two discrimination functions to calculate three scores, and the largest score represents the corresponding mental workload level.

4.2 Validity Check of the Model

The original validation method was used to check the predication and discrimination accuracy of the constructed Bayesian Fisher discrimination function. It substituted the 26 groups of subject sample data measured back into the constructed discrimination function to evaluate the accuracy level of predication and discrimination, and showed the check results in Table 4.

Table 4. Results of discrimination and predication accuracy rate

Method	**Predicted Mental Workload Accuracy Rate (%)**			
	Mental Workload	Low	High	Total
Original	Low	69.23	30.77	100.0
	High	15.38	84.62	100.0

4.3 Establishment of the Model and Instructions

It could be known from the comparative results of Table 4 that, when employing the general discrimination analysis method, the average discrimination and prediction accuracies of original check method was 76.92%. Specifically, the discrimination and prediction accuracies for low workload and high workload were 69.23% and 84.62%, respectively. The discriminate functions are as follows:

$$y_1=0.406x_1 - 0.075\ x_2 + 0.037\ x_3 -12.666 \tag{1}$$

$$y_2=0.307\ x_1- 0.222\ x_2 + 0.024\ x_3 - 8.149 \tag{2}$$

In the equations, y_1, y_2 represent the discriminate function value of the low and high levels of mental workloads, respectively. And x_1 value represents the SDNN value, x_2 value represents the peak amplitude of MMN, x_3 value represents eye blink numbers. According to the values of x_1, x_2, and x_3, the values of y_1, y_2 were calculated and compared. If the y_1 value is bigger, subjects are considered at a low level of mental workload. If the y_2 value is bigger, subjects are considered at a high level of mental workload.

5 Discussion

The major findings of the present study can be summarized as follows: as the mental workload increased, the three categories of indices represented different changes. For the performance evaluation indices, the detection accuracy of flight operation abnormal information was significantly decreased, and the response time was extended remarkably. For the subjective evaluation indices, the score of NASA_TLX was significantly increased. For the EEG evaluation indices, the peak amplitude of MMN was significantly increased; for the ECG evaluation indices, SDNN was sensitive to the mental workload change, which was significantly decreased as the mental workload increased; for the EOG evaluation indices, the number of eye blink was decreased significantly.

5.1 Sensitivity of Performance and Subjective of Indices to Mental Workload Change

The behavioral results indicated that the task performances of the subjects had been clearly distinguished between the high and low mental workloads conditions. Under the high mental workload condition, the accuracy rate of detecting abnormal information declined and the reaction time delayed. The outcomes implied that changing mental workloads during flight simulation condition affected subjects' operation performance significantly, and the results were consistent with the prior studies in other fields [6,7].

In the present study, the score of NASA_TLX increased significantly as the task got more difficult, which was consistent with former study results about flight tasks [5]. The present result also demonstrated that the setting of mental workload levels in this experiment for different flight task difficulties showed significant disparity from the participants' subjective perspective, which is accorded with the expectations.

5.2 Sensitivity of Integrated Evaluation Model to Mental Workload Change

The subjects' physiological indicators under different mental workload were chosen to construct the discrimination and prediction models of mental workload. Single physiological indicator evaluation model, two physiological indicators integrated evaluation model, and three physiological indicators integrated evaluation model, were detected by the Bayesian Fisher model to test their discrimination accuracy, respectively. The results are shown in Table 5.

Using original check method, by contrast, the integrated evaluation model based on three physiological indicators had the highest prediction accuracy (76.92%), followed by two physiological indicators integrated evaluation model by MS (MMN and SDNN) (73.08%) and SE (SDNN and Eye Blink) (69.23%), then was the two physiological indicators integrated evaluation model by EM (Eye Blink and MMN) (57.69%) and the single indicator evaluation model by Eye Blink index (57.69%), and finally the single indicator evaluation model by SDNN index (53.85%) and MMN index (53.85%).The overall comparison results of discrimination accuracy among different models showed that the discrimination accuracy of the model based on three physiological indicators is higher than the model based on two physiological indicators, which is higher than the model based on single physiological indicator. It demonstrated that multi-dimension physiological integrated evaluation model was generally more effective than single physiological indicator to discriminate the mental workload. In addition, among the single physiological indicator evaluation models, the Eye Blink indicator has the highest prediction accuracy for mental workload.

Table 5. Evaluation results of models based on single physiological indicator and multiple physiological indicators

Method	**Evaluation indicators**	**Predicted Mental Workload Accuracy Rate (%)**		
		Low	High	Total
Original	SDNN	46.15	61.54	53.85
	MMN	61.54	46.15	53.85
	Eye Blink	46.15	69.23	57.69
	SE (SDNN, Eye Blink)	61.54	76.92	69.23
	MS (SDNN, MMN)	76.92	69.23	73.08
	EM (Eye Blink, MMN)	61.54	53.85	57.69
	SME (SDNN, MMN, Eye Blink)	69.23	84.62	76.92

5.3 Comparison of Integrated Evaluation Model and NASA_TLX Score

The results of discrimination accuracy comparison between evaluation based on three physiological indicators integrated evaluation model and evaluation based on NASA_TLX score are shown in Table 6.

Using original check method, by contrast, under low mental workload, the evaluation based on three physiological indicators evaluation model had lower prediction accuracy than the evaluation based on NASA_TLX score; while under high mental workload, the evaluation based on three physiological indicators evaluation model had higher prediction accuracy than the evaluation based on NASA_TLX score; as for the average predication accuracy, the prediction accuracy of evaluation based on three physiological indicators evaluation model was the same as the evaluation based on NASA_TLX score. It demonstrated that it was feasible to substitute the three physiological indicators evaluation model for NASA_TLX score evaluation.

Table 6. Evaluation results of models based on three physiological indicators and NASA_TLX score

Method	**Evaluation indicators**	**Predicted Accuracy Rate (%)**		
		Low	High	Total
Original	SME (SDNN, MMN, Eye Blink)	69.23	84.62	76.92
	NASA_TLX	76.92	76.92	76.92

The overall comparison results of discrimination accuracy among different models showed that the discrimination accuracy of the model based on three physiological indicators is higher than the model based on two physiological indicators, which is higher than the model based on single physiological indicator. It demonstrated that multi-dimension physiological integrated evaluation model was generally more effective than single physiological indicator to discriminate the mental workload. In addition, among the single physiological indicator evaluation models, the Eye Blink indicator has the highest prediction accuracy for mental workload.

Therefore, the present study provides a method to extract the objective evaluation indicators sensitive to mental workload by experimental measurement, and then construct objective and real-time integrated discrimination and prediction model for mental workload change during flight process. It can help to determine and predict the mental workload task design for the aircraft cockpit human-machine interaction system design.

6 Conclusion

In conclusion, the discrimination and prediction results of the model proposed in this paper revealed a satisfactory consistency with the experimental measured results, and the model can accurately reflect the variation characteristics of the mental workload of the aircraft flight deck display interface, and provide a sound foundation for the ergonomic evaluation and optimization design of the aircraft flight deck display interface in the future.

Acknowledgments. This research is supported by National Basic Research Program of China (NBRPC, Program Grant No. 2010CB734104). The opinions and conclusions expressed in this article are those of the authors and do not necessarily reflect the views of the NBRPC.

References

1. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. North-Holland, Amsterdam (1988)
2. DiDomenico, A., Nussbaum, M.A.: Effects of different physical workload parameters on mental workload and performance. Int. J. Ind. Ergon. 41(3), 255–260 (2011)
3. Wei, Z., Wanyan, X., Zhuang, D.: Measurement and evaluation of mental workload for aircraft cockpit display interface. J. Beijing Univ. of Aero. and Astro. 40(1), 86–91 (2014)
4. Wei, Z., Zhuang, D., Wanyan, X., Liu, C., Zhang, H.: A model for discrimination and prediction of mental workload of aircraft cockpit display interface. Chi. J. Aero. (accepted)
5. Ryu, K., Myung, R.: Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. Int. J. Ind. Ergon. 35(11), 991–1009 (2005)
6. Cohen, J., Polich, J.: On the number of trials needed for P300. Int. J. Psychophysiol. 25(3), 249–255 (1997)
7. Ullsperger, P., Freude, G., Erdmann, U.: Auditory probe sensitivity to mental workload changes–an event-related potential study. Int. J. Psychophysiol. 40(3), 201–209 (2001)