Experimental Study of Task Load Measurement for Basic Flight Operation Task

Lijing Wang [a], Xueli He [a], Yuhui Wang [a] and Yingchun Chen [b]

[a] Fundamental Science on Ergonomics and Environment Control Laboratory
School of Aeronautic Science and Engineering
Beihang University
Beijing, 100191, China

[b] Shanghai Aircraft Design and Research Institute
Shanghai, 201210, China

ABSTRACT

The aim of this research is to do experimental measurement research of the task load by subjective and objective measurement, and find the most effective measurement tool for the basic flight task load, for providing experimental research foundation for task load quantitative research. Three sets of experiments were designed in this research. The first two sets of experiments were based on human information processing stages and different processing modalities, designed as following: Visual (Perception) - Cognition – Manual (Responding), and Auditory (Perception) - Cognition –Manual (Responding), with consideration of different processing codes; the third set of experiment was designed based on flight operation task (obtaining instrument data information), including two experiments. Physiological parameters of participants were recorded with apparatus, and questionnaires were used to record subjective evaluation results. The experimental data were processed with correlation and significance analysis. The correlations between the pupil diameter, subjective evaluation and their average value are high ($p<0.05$). The significance between each set of data of the pupil diameter are high ($p<0.05$). The task load of basis flight operation task units can be measurement based on the pupil diameter and subjective assessment, the task load of perception tasks, cognition tasks, responding tasks divided from flight operation task can be measured by subjective evaluation.

Keywords: Task load; Physiological Index; Subjective Evaluation; Basic Flight Task

INTRODUCTION

Task analysis is widely used in workload predict assessment in aircraft design. Bierbaum (1987) used task analysis to establish workload assessment model, and applied this model in the design of UH-60 aircraft. Bierbaum (1990) described the using instructions of task analysis method used in workload assessment in detail, and applied it to compare the crew workload of MH-47E aircraft to CH-47D aircraft (Hamilton, 1991), to optimize the cockpit design. The first step of task analysis method is the decomposition of a task scenario into a sequence of task elements, and then a variety of different methods can be used to measure the task load of each task elements.

In workload assessment model established by Linton et al. (1989), classification of task load is based on cognitive channel, and task elements values of each channel are obtained by subjective estimate method. In CRAWL (Computerized Rapid Analysis of workload) (Thompson et al., 1986), operating tasks are divided into task elements, expert advice method (subjective measurement) and traditional analysis method are combined to assess task load of task elements. Aldrich, et al. (1989) classified task elements based on five behavioral channels (cognition, vision, hearing, muscle movement and psychomotor), the task load of task elements is obtained by the subjective measurement with a 7-point scale. In the task load model (Task Loading Model TLM) established by Staveland et al. (1991), a more detailed decomposition of already classified cognitive channel (visual, auditory, cognitive, action) was conducted based on operation action demands, then the task elements were categorized according to the various channels, it also used a subjective evaluation to define task unit load.

Subjective and objective measurements (EEG measurement, eye movement measurement and ECG measurement) both are usually used to measure workload (O 'Donnell, 1986). The average value of different participants measured in experiment values were taken as the analysis basis. For example, the flight mission mental workload study calculated the measurement data average values of ten crews, including subjective workload ratings, eye blink changes, eye durations etc., to analyze the workload of the F4 phantom aircraft pilots and weapon systems officers (Wilson, 1993). At the same time, in his one other research on validity and reliability of these physiological indicators, these evaluation values (Wilson, 2002) are being analyzed. In researches of Lee and Liu (2003) on flight workload also, the average value of heart rate, RR interval indicator was obtained before further analysis.

Therefore the aim of this research is to doing experimental measurement research of the task load, analyze the rationality of using physiological indictors average value from different participants as the analysis basis of task load, analyze the correlations between data obtained through subjective and objective measuring method, and the application of these two methods aims at verification the valid measurement method used in the task load assessment of basis operation action. In this study, first two sets of experiments are designed based on the multi-resource model information processing stage (perception, cognition and responding) proposed by Wickens (2002), then basis flight task of obtaining instrument information is designed as the third set of experiments.

EXPERIMENTS

Experiments were designed based on human information processing stages (e.g., perception, central processing, and response execution), where they require different processing modalities (e.g., visual/auditory channels) and where they rely on the different processing codes (e.g., Verbal/spatial). The first and the second set were the basic experiments based on perception, cognition and responding, and the third set of experiments was designed based on flight operation task.

The first set included three different basis tasks: Visual Tracking, Visual-Cognition and Visual-Cognition-Responding. The second set of experiments included two different basis tasks: Auditory-Cognition, Auditory-Cognition-Responding. The third set included two different basis flight tasks.

Experimental Contents

The first set of experiments: Visual Tracking: The cursor "+" was showed on the computer screen randomly. The switching time of two "+" interval was set based on the optimum perform time obtained from the prepare-experiment. In this experiment, the total duration time is 60s. Subjects were required to visual tracking “+" cursor on screen; Visual-Cognition: One letter was shown on the computer screen randomly one by one. The switching time of two letters interval was set based on the optimum perform time obtained from the prepare-experiment. The total duration of the each trial is 60 s. Subjects were required to see the letters and try to recognize them; Visual-Cognition-Responding: Letter was shown on the computer screen randomly. The end of the last operation is the beginning of the next letter appearing. The total duration of the each trial is set to 60s. The subject were required to see the letters and cognitive and respond on the keyboard with the right keys.

The second set of experiments: Auditory-Cognition: Subjects were required to cognitive after hearing the letters via sound. The experimental requirements were based on hear letters to identify them; Auditory-Cognition-Responding: Subjects were required to cognitive after hearing the letters via sound. The experimental requirements were based on hearing letters to identify them. Subjects were required to see the letters and respond on the keyboard with the right keys.

The third set of experiments: Visual-Cognition (obtaining static instrumental data): The instrument graph with different static data (Fig.1) was shown on the computer screen randomly one by one. The total duration of the each trial is 60 s. Subjects were required to see the letters and try to recognize them; Visual-Cognition (obtaining dynamic instrumental data): The dynamic instrument with changing data was shown on the computer, the data in yellow oval (Fig.1) changes after a short time. The total duration of the each trial is 60s. Subjects were required to see the letters and try to recognize them.

Fig.1 Instrument with static data

Participants

Ten students (undergraduate or graduate) participated in this experiment, male, age 19-25, visual acuity 1.0-1.5, and signed a consent form, were paid for their participation.

Apparatus

Stimuli were displayed in computer screen. The whole experiment was presented using pictures on the computer screen. The content of the experiment was edited and shown in the software (Experiment Buide, EB) of Eyelink II. Participants sat approximately 70 cm from the bottom of the stimulus. Eye movement data were collected using data analysis software of Eyelink II. The pupillometer consisted of a video camera and infrared light source that were pointed at a participant's eye, and a device that tracked the location and size of the pupil using these tools. Pupil size was recorded with 250 Hz or 500 HZ. HRV index were measured by Biofeedback 2000 x system.

Measurement methods

Physiological measurements

Measurement of eye movement: Pupil diameter was real-time measured and recorded by Eyelink II eye movement measurement system.

HRV measurement: LF, HF, LF/HF, SDNN, RMSSD, Pulse, HRV index were real-time measured and recorded by Biofeedback 2000 x system.

Subjective rating evaluation

After the end of the physiological experiments, all of the subjects were asked to evaluate the task load during the experiments. The meaning description of evaluation scores is shown in Table1.

The evaluation questionnaires include two parts; the first part aims to evaluate the experiment tasks, including all the three set, seven experiments. The second part aims to evaluate the basis tasks divided from the seven experiments: Perception task: visual (cross), visual (letter), visual (static number), visual (dynamic number), auditory (letter); Cognition task: cognition (position), cognition (letter), cognition (static number), cognition (dynamic number); Responding task: responding (press letter).

Table 1 the description of scores

score	description
0	No load (easy to complete the task perfectly)
1	Load is small (no burden to complete the task)
2	Mild load (little effort to complete the task)
3	Median load(more effort to complete the task, task is difficult)
4	Heavy load(a lot of effort to complete the task perfectly)
5	Very heavy load(great effort to complete the task)

Analysis method

Data processing

Pearson Correlation was used to test the correlation between each set of experimental data and their average value, and the correlation between each average value.

Wilks'Lambda was used to test the significance between each set of data of the pupil diameter.

All analyses were completed using the SPSS 18.0 (IBM Corporation, Chicago, USA).

Sensitivity, diagnosticity and intrusiveness

Sensitivity is the capability of a technique to discriminate significant variations in the workload imposed by a task or group of tasks.

Diagnosticity is capability of a technique to discriminate the amount of workload imposed on different operator capacities or resources (e.g., perceptual versus central processing versus motor resource).

Intrusiveness is the tendency for a technique to cause degradations in ongoing primary task performance.

DATA PROCESSING

Results of data process

The correlations between average value of all the measurement indictors from all participants of the same task and the pupil diameter of each participant were analyzed by Person Correlation. The analysis results are shown in Table 2.

The correlations between the pupil diameter average value from all participants of the same task and the pupil diameter of each participant are all high ($p<0.05$). At the same time, the correlations between the subjective evaluation average value from all participants of the same task and the subjective evaluation value of each participant are all high ($p<0.05$). The correlations between other sets of data and their average value are low ($p>0.05$).

Table 2 Analysis of correlations between each set of data and their average value

		subject1	subject2	subject3	subject4	subject5	subject6	subject7	subject8	Subject9	subject10
LF	Pearson Correlation	.663	--	.271	-.210	.324	.048	-.061	--	.862**	--
	Sig.	.073	--	.516	.618	.433	.910	.885	--	.006	--
HF	Pearson Correlation	.368	--	.418	.313	.434	.174	-.085	--	.757*	--
	Sig.	.370	--	.302	.450	.283	.680	.842	--	.030	--
LF/HF	Pearson Correlation	.595	--	.266	.732*	.235	.382	.768*	--	.588	--
	Sig.	.120	--	.525	.039	.575	.351	.026	--	.125	--
SDNN	Pearson Correlation	.775*	--	.416	.741*	-.160	.399	.053	--	.170	--

		subject1	subject2	subject3	subject4	subject5	subject6	subject7	subject8	Subject9	subject10
	Sig.	.024	--	.305	.035	.706	.328	.901	--	.688	--
RMSSD	Pearson Correlation	.510	--	.864**	.819*	.731*	.697	.798*	--	.233	--
	Sig.	.197	--	.006	.013	.039	.055	.018	--	.579	--
PNN50	Pearson Correlation	.705	--	.894**	.468	.417	.471	.013	--	.319	--
	Sig.	.051	--	.003	.242	.304	.238	.976	--	.441	--
Pulse	Pearson Correlation	.054	--	.430	.716*	.383	-.344	-.199	--	.634	--
	Sig.	.898	--	.287	.046	.349	.404	.637	--	.092	--
pupil diameter	Pearson Correlation	--	--	-.840**	.984**	.945**	.988**	.892**	.980**	.742*	.938**
	Sig.	--	--	.009	.000	.000	.000	.003	.000	.035	.001
HRV- -index	Pearson Correlation	.872**	.871**	.761*	.333	.771*	.626	.530	--	-.048	--
	Sig.	.005	.005	.028	.420	.025	.097	.176	--	.910	--
subjective evaluation	Pearson Correlation	.789*	.941**	.590	.737	.646	.708	.482	.789*	.789*	.767*
	Sig.	.035	.002	.163	.059	.117	.075	.273	.035	.035	.044

**. Correlation is significant at the 0.01 level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed); Blanks with transverse line in the table is due to absence of data

Analysis of correlations between each set of average value and subjective evaluation

Subjective evaluation was used to measure workload mostly, which can be used as the test basis of other measurement indictor. The correlations between pupil diameter and other measurement indictors were presented by Pearson correlation, shown in Table 3. The correlation between pupil diameter and subjective evaluation is high ($p<0.1$). The correlation between other sets of data and subjective evaluation is low ($p>0.1$).

Table 3 Correlations analysis between each set of average value and subjective evaluation

		Pupil diameter	LF	HF	LF/HF	SDNN	RMSSD	PNN50	Pulse
subjective evaluation	Pearson Correlation	.669	-.042	.253	-.488	-.401	.048	.162	-.105
	Sig.	.070	.921	.546	.220	.325	.910	.701	.804
	N	8	8	8	8	8	8	8	8

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Analysis of significance between each set of data of the pupil diameter

The significance between each set of data of the pupil diameter are high ($p<0.05$), shown in Table4.Difference is significant between sets of data in pupil diameter experiment; pupil data variance is insignificant from person to person.

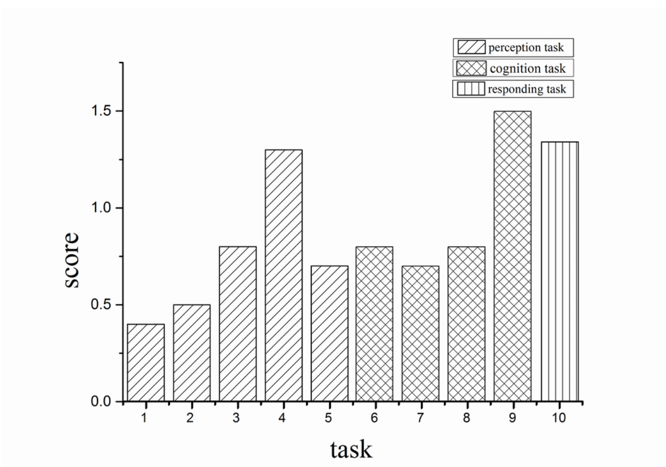
Table 4 Significance analysis between each set of data of the pupil diameter

Multivariate Testsb						
Effect		Value	F	Hypothesis df	Error df	Sig.
test	Wilks' Lambda	.012	23.865a	7.000	2.000	.041

a. Exact statistic; b. Design: Intercept Within Subjects Design: test

Subjective rating evaluation of decomposition tasks

The tasks divided from the three sets experiments are measured by subjective evaluation, the evaluation results are shown in Fig.4. Task 1, task2, task3, task4, and task5 belong to perception task; task 6, task 7, task 8, and task 9 belong to cognition task; task 10 belongs to responding task. From the results, the task load of dynamic data perception task, dynamic data cognition task, and responding task are higher than other tasks.



task1: visual (cross); task2: visual (number); task3: visual (static number); task4: visual (dynamic number); task5: Auditory (letter); task6: cognition (position); task7: cognition (number); task8: cognition (fixed number); task9: cognition (changing number); task10: responding (press letter);

Fig.4 The subjective evaluation results of ten subtasks

Sensitivity, diagnosticity and intrusiveness

In subjective rating evaluation of decomposition tasks, the mean data of task load is 0.88, standard deviation is 0.36. There are significant differences between different tasks with the subjective evaluation. There are sensitivity and diagnosticity with the pupil diameter measurement and subjective evaluation. And there is no intrusiveness with those task load measurement method with the subjective rating from participants, and the experimental results from with pupillometer and non-equipment.

CONCLUSION

In this study, three sets of experiments were designed based on the multi-resource model information processing stage (Perception, cognition, responding) and simple flight operation procedure. Pearson Correlation was used to test the correlation between each set of experimental data and their average value, and the correlation between each average value, for finding the most efficient method to measure basis flight operation task units and perception task, cognition task, responding task divided from flight operation task.

The correlations between the pupil diameter average value from all participants of the same task and the pupil diameter of each participant are all high. At the same time, the correlations between the subjective evaluation average value from all participants of the same task and the subjective evaluation value of each participant are all high. So the average value of pupil diameter and subjective evaluation can be used to descript the change trend of different experiments.

There is a strong correlation between pupil diameter and the subjective evaluation. The significance between each set of data of the pupil diameter is high. Difference is significant between sets of data in pupil diameter experiment; pupil data variance is insignificant from person to person. The pupil diameter and subjective evaluation can also be used as the data basis used to the quantitative analysis of task load.

By conducting the analysis of the correlation between the measurement data of the pupil diameter and the subjective evaluation value, it can be found that the subjective evaluation method can be used for the basis operation load measurement which cannot be measured by designing the experiment content.

There are sensitivity, diagnosticity and non-intrusiveness in the task load measurement by the pupil diameter and subjective assessment.

DISCUSSION

The ECG indicators can reaction mental load changes perfectly in many researches of mental load, but in this experiment, it didn’t work. The main reason is that the experimental time was shortened to reduce the affect caused by fatigue accumulated over time. The time that the ECG indicators reach steady with the tasks change is long, So ECG indicators couldn’t reflect the change of physiological indicators with the task change in this research.

The paper was aimed to find a method to measure basis task load from many measurement methods. And the ultimate purpose of the research is to measure the load of the basis task and its subtasks (perception, cognition, operation) by variety measurement methods. In order to measure the task load, not the workload, we need to eliminate human factor in experiment, fatigue, experience, mood, and so on. So we control the experimental time, training time, experimental environment, and the subjective evaluation instruction.

ACKNOWLEDGEMENTS

This research was supported by the strong support of 973 Project of National Technological Department (2010CB734103).

REFERENCES

Aldrich, T. B., Szabo, S. M., & Bierbaum, C. R. (1989). The development and application of models to predict operator workload during system design.*Applications of human performance models to system design*, 65-80.

Bierbaum, C. R., Szabo, S. M., & Aldrich, T. B. (1989). *Task Analysis of the UH-60 Mission and Decision Rules for Developing a UH-60 Workload Prediction Model. Volume 1. Summary Report* (No. ASI690-302-87). ANACAPA SCIENCES INC FORT RUCKER AL.

Bierbaum, C. R., Fulford, L. A., & Hamilton, D. B. (1990). *Task Analysis/Workload (TAWL) User's Guide. Version 3.0* (No. ASI690-323-89). ANACAPA SCIENCES INC FORT RUCKER AL.

Hamilton, D. B., & Bierbaum, C. R. (1992). *Operator Workload Predictions for the Revised AH-64A Workload Prediction Model. Volume 2. Appendixes A Through H* (No. ASI690-354-92-II). ANACAPA SCIENCES INC FORT RUCKER AL.

Lee, Y. H., & Liu, B. S. (2003). Inflight workload assessment: Comparison of subjective and physiological measurements. *Aviation, space, and environmental medicine*, *74*(10), 1078-1084.

Linton, P. M., Plamondon, B. D., Dick, A. O., Bittner Jr, A. C., & Christ, R. E. (1989). Operator workload for military system acquisition. *Applications of human performance models to system design*, 21-46.

O'donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. *Measurement Technique*, *42*, 5.

Staveland, L. (1991, October). MIDAS TLM: man-machine integrated design and analysis system task loading model. In *Systems, Man, and Cybernetics, 1991.'Decision Aiding for Complex Systems, Conference Proceedings., 1991 IEEE International Conference on* (pp. 1219-1223). IEEE.

Thompson, M. W. and R. P. Bateman, (1986). *A Computer-Based Workload Prediction Model*, SAE Aero Tech.

Wickens, C. D. (2002). Multiple resources and performance prediction.*Theoretical issues in ergonomics science*, *3*(2), 159-177.

Wilson, G. F. (1993). Air-to-ground training missions: A psychophysiological workload analysis. *Ergonomics*, *36*(9), 1071-1087.

Wilson, G. F. (2002). An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology*, *12*(1), 3-18.