Dean&Francis

ISSN 2959-6157

Research on the Evaluation Method of Pilot Workload in High Load Scenarios during Aircraft Climb Phase

Hongyu Yi

School of mechanical and Power Engineering, Zhengzhou university, Zhengzhou, Henan Province, China, 450001
1743088108@qq.com

Abstract:

To address the challenge of accurate evaluation of pilot workload during aircraft climb phase, an innovative evaluation method combining subjective and objective factors was proposed in this paper. Three high load scenarios were defined by systematic analysis of the climb phase mission. In terms of research methods, the AWDS subjective scale is optimized, and the degree problem and closed loop logic verification are used to improve data reliability. Meanwhile , eye movements and heart rates were collected with non-contact instruments, with precise calibration achieved by establishing individual physiological baselines. Experimental validation demonstrated a significant correlation between subjective workload scores and objective indicators, such as pupil diameter change rate and blinking frequency change rate, indicating high consistency between subjective and objective data.. The final comprehensive evaluation model provides an effective solution for real-time monitoring of pilot workload and cockpit human-computer interaction optimization.

Keywords: aircraft climb phase, workload, subjective scale, physiological data, Verification of consistency

1. Introduction

Aviation safety confronts multiple severe challenges across various levels, among which human factors have long been a critical issue. Historical data shows that approximately 70% - 80% of aviation accidents are directly or indirectly related to human factors[1]. Pilot workload is an important factor, making its evaluation an essential component of aviation safety research. Traditional methods of workload assessment mainly include subjective assessment and physiological measurement. The most commonly used subjective evaluation method is the NASA Mission Load Index scale, which evaluates workload from six dimensions: mental needs, physical needs, time needs, performance levels, effort levels, and frustration levels[2]. Participants rate each dimension's score on the appropriate scale based on their actual experience. The overall workload feel score can be calculated through weighted summation of these

ISSN 2959-6157

dimension scores, enabling the workload level of the task[3]. Physiological measurement objectively assesses pilots' physiological arousal and psychological workload by monitoring changes in physiological signals during missions. Key measured parameters include EEG (α, β, θ waves), heart rate, heart rate variability, pupil diameter, and blinking frequency[4].This paper focus on developing an effective and reliable workload assessment method for high load scenarios suitable for aircraft climb phase, and verifying the consistency between subjective and objective assessments.

This research adopts a multi-modal, layered fusion evaluation system. Its core ideas are multi-mode data synchronous acquisition, layered processing, fusion analysis and cross-validation. The system architecture comprises three core layers: data acquisition, fusion analysis, and application verification. This study establishes a comprehensive subjective-objective integrated evaluation framework, combining the optimized subjective scale with non-contact physiological measurement to form a more robust and complementary evaluation system.

2. High load scenario analysis

The aircraft climb phase starts with the landing gear retraction after takeoff and ends when the aircraft reaches cruise altitude. High load characteristics in this process are not uniformly distributed, but concentrated at specific key nodes or time windows.

This study defines three typical, repetitively induced high load scenarios:

First, departure Initial Climb and Standard Instrument Departure Procedures Execution Phase

(1) Time window: From takeoff to about 1500 feet above ground level, lasting 3 - 5 minutes.

(2) High workload sources: Time pressure, multi task parallel, and high working memory load.

Second, crossing Transition Altitude/Transition Altitude Layer

(1) Time window: When approaching the locally specified transitional altitude layer, usually in the middle of the climb.

(2) High workload sources: Critical cognitive transitions, serious consequences of procedural errors, and task interruption-recovery processes.

Third, ATC Directive Intensive Period and Complex Airspace Environment

(1) Time window: Mid to late stages of the climb phase, especially when overflying busy terminal areas or en route rendezvous points.

(2) High workload sources: Dramatic increases in communications loads, dynamic decision and plan updates, constant distraction.

3. Optimization and implementation of subjective evaluation methods

3.1 Analysis of limitations of traditional subjective scale

The same individual may provide different subjective evaluations under varying training levels, environments, times, or psychological activities. A reasonable and scientific design of certain definitions and criteria in the subjective evaluation table can mitigate the differences, but cannot fully eliminate them[5]. Additionally, traditional evaluation processes lack a data reliability self-check mechanism. If contradictory statements arise, the scale provides no basis for determining accuracy.

3.2 Design of the subjective evaluation scale after optimization

This study designed a new quantitative tool for assessing mission workload, the "Aircraft Workload Dynamic Scale (AWDS)". It is more intuitive, muanced, and intelligent, aligning closely with pilot perception. While retaining the core ideas of NASA-TLX, the dimensions were restructured and contextualized to better reflect pilot's actual experience. For response modes, numerical scoring is replaced with visual analog sliders. Each dimension is paired with a slider, with icons and brief descriptions at both ends to indicate extreme values. Additionally, the evaluation process removes redundant weight assignment tasks. The system derives implicit weights based on pilot's slide selection mode. The six dimensions are shown in Table 1.

Table 1. six dimensions of AWDS

Dimension	Problem description	Visualization icon
D1: Mission complexity	What is the amount of information I need to think and process simultaneously at the mission stage just now?	A gradient icon from „single flow diagram“ to „complex network diagram“

D2: A sense of time urgency	How well do I feel the time available for operations and decisions is adequate?	A gradient icon from a „slow flowing hourglass“ to a „sharp twinkling countdown“
D3: Operational smoothness	Is my interaction with the aircraft systems smooth and intuitive?	A gradient icon from „Smoothed Line“ to „Clayton‘s Sawtooth Line“
D4: Distraction	To what extent do I need to switch my attention frequently between different sources of information and tasks?	A gradient icon from „focused spotlight“ to „scattered multiple points of light“
D5: Situational clarity	How clear is my grasp of current flight status and future trends?	A gradient icon from „Clear Sky“ to „Misty Diffuse“
D6: Psychological comfort	What level of tension or discomfort do I feel during this phase?	An icon that transitions from a „relaxed face“ to a „tense sweating face“

3.3 Closed loop subjective evaluation work flow

Flow description: Pilot completes preliminary evaluation → system detects inconsistencies or extreme selection → triggers secondary confirmation problem automatically → guides pilot to reconsider decisions, improving data reliability. Figure 1 illustrates the workflow of aligning subjective scale data and non-contact physiological data (e.g., eye movement, heart rate) via timestamp markers, then normalizing the data (e.g., calculating physiological change rates relative to individual baselines) for subsequent consistency analysis.

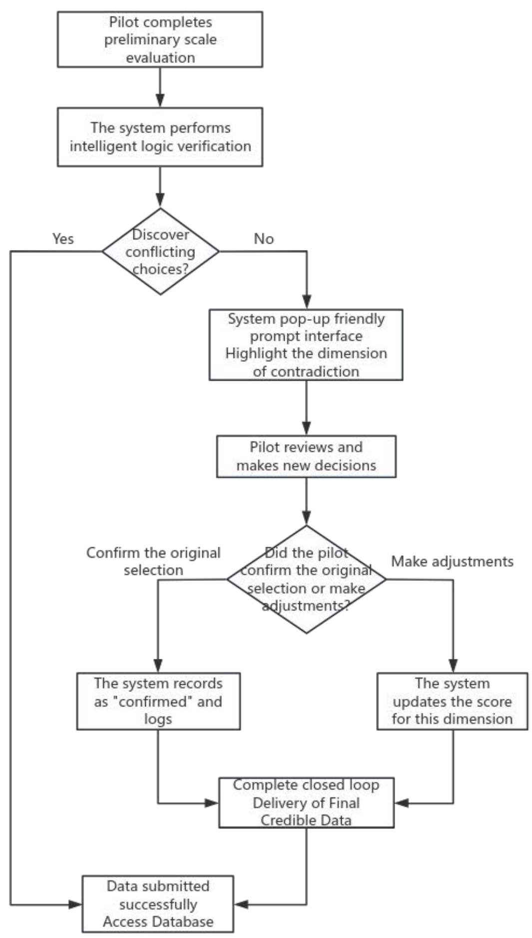

Figure 1. Data Synchronization and Standardization Process for Multimodal Pilot Workload Assessment

ISSN 2959-6157

Step 1: Preliminary evaluation and data submission

After simulated flight, pilots complete the AWDS scale on a tablet, adjust sliders for six dimensions to reflect their perceived workload levels, then submit the data.

Step 2: System Intelligent Logic Checkout

Submitted data is not immediately stored in the database. Instead, it is first sent to an intelligent logic check module, which contains a pre-established library of contradictory rules based on human factors engineering principles and flight experience. Examples of contradictory rules include:

(1) Rule 1 (time to operation contradiction): Triggered if time urgency ≥ 80 and operational smoothness ≥ 70.

(2) Rule 2 (Psychological to mission contradiction): Triggered if psychological comfort<20 and mission complexity<30.

(3) Rule 3 (Situational to attention paradox): Situational score>80, attention score>75, triggers.

Step 3: Judgment and Interactive Validation

The system identifies conflicts based on a rule base. If no conflicts are detected, the data is submitted directly to the database; If a conflict is found, a friendly, non-critical prompt window pops up. The slider associated with the contradictory response is highlighted, prompting pilots to review their selection.

Step Four: Pilot's Second Decision and Closed Loop Formation

Faced with a prompt from the system, pilots have two options:

1. Determine the original selection: If pilots believe their initial response accurately reflects the situation. They can select "confirm submission". The system labels the data as "confirmed manually" and saves it to the database. This data itself is valuable, as it may indicate unique workload patterns.

2. Make adjustments: If pilots recognize errors in their initial submission or seek greater accuracy, they can modify the highlighted sliders directly on the prompt screen and then tap "Resubmit".

The above design not only makes the AWDS scale more modern and user-friendly but also enable it to transcend traditional scales. It provides subjective data for subsequent pilot workload studies that is more abundant, diverse, and ecologically valid.

4. Non-contact Physiological Data Acquisition and Processing

4.1 Physiological index selection and theoretical basis

The selection of indicators follows three principles: sensitivity to cognitive and emotional workload, reliable access through non-contact measurement, and clear neurophysiological explanations. The parameters of pupil diameter, blinking frequency and heart rate were selected in this study.

4.2 Non-contact data acquisition scheme

1. Equipment selection and configuration

(1) Eye Movement Data: Select Tobii Pro Fusion or an equivalent accuracy desk mounted eye tracker.

(2) Electrocardiogram data: Select the Logitech Brio4K HD webcam.

2. Mitigation of foreign object sensation and strategies for ecological validity assurance

(1) Early adaptation:

All participants complete at least 2 simulator acclimatization sessions of no less than 1 hour each before the formal experiment. The identical equipment configurations in training is used to ensure pilots adapted to the presence of equipment in the formal experimentation.

(2) Stealth of equipment:

All equipment adopt a black matte finish and are cleverly embedded into the simulator environment, avoiding reflective and obtrusive looks, enabling it to blend into the cockpit background.

5. Multimode data fusion and consistency analysis

5.1 Construction of Individualized Physiological Baseline Library

Pilots are asked to read neutral, non-aviation-related magazines or books in a quiet preparation room with soft lighting for 30 minutes. Throughout this period, pilots maintained a stress-free state with low cognitive workload. Physiological data were recorded during the 30-minutes baseline data acquisition and using the same non-contact instrumentation and parameter settings as the formal experiment. For baseline value calculations, data from the first 5 minutes (adaptation period) and the last 5 minutes (possible fatigue period) were excluded. The average value of the stable middle 20 minutes data was defined as the individual baseline value for each pilot.

The calculation formula is:

Baseline of individual physiological parameters=mean (physiological_parameterbaselines).

A personalized baseline database was established for each pilot, serving as the reference standard for subsequent data calibration and analysis.

5.2 Data synchronization and standardization

1. Multimode data synchronization

An event marker synchronization method is used. Predetermined code in the flight simulator software, at the precise moment of the beginning of the three high load scenarios, simultaneously sending a unique timestamp marker accurate to milliseconds to all data streams. In subsequent data processing, all data streams are aligned on a timeline using these common event markers as reference points.

2. Data standardization and feature extraction

(1) Time window delineation: A fixed window of analysis time is defined for each high load scenario (30 seconds before the start of the scenario to 30 seconds after the end, for a total of 2 minutes), based on event markers.

(2) Physiological Characteristics Calculations:

Individual baseline values were retrieved, and raw physiological data were converted into relative change rates compared to the individual baseline

$$\text{Physiologicalindexchangerate}=\frac{\text{Missionwindowaverage-PersonalBaselineValue}}{\text{PersonalBaselineValue}}\text{¡Á}100\% \quad (1)$$

(3)Subjective data matching: After the mission, pilots completed the AWDS scale by recalling and scoring specific mission phases. Each pilot's composite load score on the scale corresponding to that high load scenario was paired one-to-one with the physiologic indicator change rate calculated over the synchronized time window. This creates a pair of data (subjective score, objective physiological rate of change).

5.3 Analysis of subjective and objective data consistency

This study used the t-P confidence analysis to verify the consistency between subjective and objective data.

1. Analysis logic

(1) Core issue: Is the correlation between the subjective scores we observe and the rate of change in physiological indicators real or caused by random sampling errors.

(2) Original hypothesis (H0): There is no linear correlation between the subjective load score and the rate of change of physiological indicators (ρ=0).

(3) Alternate hypothesis (H1): There is a linear correlation between subjective load score and the rate of change of physiological indicators (ρ≠0).

2. Analysis steps

(1) Compute the Pearson correlation coefficient (r).

Substituting all paired data (subjective scores, rate of change of physiological indicators) into the Pearson formula calculates the correlation coefficient r:

$$r=\frac{\sum\left(x-\bar{x}\right)\left(y-\bar{y}\right)}{\sqrt{\sum\left(x-\bar{x}\right)^2\sum\left(y-\bar{y}\right)^2}} \quad (2)$$

The range of values for r is [- 1,1]. R>0 indicates positive correlation, r<0 indicates negative correlation. Its absolute value | r | The closer to 1, the stronger the linear correlation.

(2) Perform a t test to calculate the p value.

Using the formula to express the calculated t statistic is:

$$t=r^*\sqrt{\frac{n-2}{1-r^2}} \quad (3)$$

(where n is the sample size of paired data, i.e.: number of pilots × number of high load scenarios)

From the t value and the degree of freedom dF=n-2, the t distribution table can be looked up or its corresponding P value automatically generated by the software directly.

(3) Results Determination (Confidence Analysis):

Set a significance level α=0.05.

If the calculated p<0.05, the null hypothesis (H_0) was rejected, and the alternative hypothesis (H_1) was accepted at the 95% confidence level. This means that the observed correlations are statistically significant, with both subjective and objective data consistent.

If p≥0.05, there was insufficient evidence to reject H_0, and no significant correlation could be concluded.

5.4 Preliminary construction of comprehensive evaluation model

The purpose of constructing a comprehensive evaluation

ISSN 2959-6157

model is to determine whether pilots' subjective workload perception can be more accurately predicted using only objective physiological indicators[6]. This is essential for the future development of real-time monitoring systems.

1. Model selection and construction

(1) Model type: A Multiple linear regression model is adopted.

(2) Variable definitions:

Dependent variable (Y): Pilot's subjective composite load score.

Independent variable (X): Standardized key physiological indicators such as pupil diameter change rate (X_1), blinking frequency change rate (X_2), and heart rate (X_3).

(3) Model equation:

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3 \tag{4}$$

2. Model training and evaluation

Regression coefficients (β0, β1, β2, β3) are obtained using the least squares method with training set data, yielding the optimal fitting model.

Given the data sets Y, X1, X2, X3, independent variables are first standardized to eliminate the influence of dimensional differences between variables:

$$Z=\frac{X-\mu}{\sigma} \tag{5}$$

Get the normalized data set for Y, Z1, Z2, Z3.

The matrix form of a multiple linear regression:

$$Y = X\beta + \varepsilon. \tag{6}$$

Of which:

$$Y = [Y1, Y2, Y3, Y4, \ldots]^{T} \tag{7}$$

$$X=\begin{bmatrix} \\ 1,Z_{A1},Z_{A2},Z_{A3} \\ 1,Z_{B1},Z_{B2},Z_{B3} \\ 1,Z_{C1},Z_{C2},Z_{C3} \\ 1,Z_{D1},Z_{D2},Z_{D3} \\ ? \end{bmatrix} \tag{8}$$

(A, B, C, D refer to the number of different experimental subjects)

Solution using least squares:

$$\beta\text{¡®}=\left(X^{T}X\right)^{01}X^{T}Y \tag{9}$$

After matrix operations, the values of β ? 0, β ? 1, β ? 2, β ? 3 can be obtained. Then there are:

$$Y=\beta\text{¡®}_0+\beta\text{¡®}_1Z_1+\beta\text{¡®}_2Z_2+\beta\text{¡®}_3Z_3 \tag{10}$$

Substituting $Z_1=\frac{X_1 0\mu}{\sigma},Z_2=\frac{X_2 0\mu}{\sigma},Z_3=\frac{X_3 0\mu}{\sigma}$ into this equation gives:

$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3 \tag{11}$$

(3) Model Evaluation:

Determination coefficient (R2): find R2 on the test set, the fraction of the variance of the subjective load fraction that the model can explain. The formula for calculating R2 is as follows:

$$SST=\sum\left(Y_i 0 \bar{Y}\right)^2 \tag{12}$$

$$SSR=\sum\left(??0 \bar{Y}\right)^2 \tag{13}$$

$$R^2=\frac{SSR}{SST} \tag{14}$$

The closer R2 is to 1, the higher the prediction accuracy of the model.

6. Conclusion

This study revolves the core issue of "pilot workload assessment under high load scenarios during aircraft climb phase", and the main work can be summarized in four aspects: the construction of a comprehensive assessment system framework with subjective and objective fusion, the innovation of traditional subjective assessment methods, the implementation of a set of contactless physiological data acquisition and processing scheme, the completion of rigorous experimental verification and consistency analysis.

Despite the expected results of this study, there are still limitations in sample size and diversity, inherent boundaries of the simulated environment, and external disturbances of physiological signals that require refinement in future work.

This study not only provides a novel, effective and user-friendly approach to pilot workload assessment, but also bridges the gap between subjective perception and objective physiology, as well as between laboratory research and real-world applications. Future research will continue to explore deeper, more practical, and broader directions based on this foundation.

References

[1] Yu Sixian (2019) General Aviation Safety Risk Analysis Research (Master's Thesis, Nanjing University of Aeronautics and Astronautics). masterhttps://doi.org/10.27239/d.cnki.

gnhhu.2019.001151..
[2] Li Li (2024). A Study on the Evaluation of the Workload of Flight Training for the ACPC (Master's Thesis, China Civil Aviation Flight Academy). masterhttps://doi.org/10.27722/d.cnki.gzgmh.2024.000325
[3] Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C., & Christ, R. E.. (1992). Comparison of four subjective workload rating scales. Human Factors The Journal of the Human Factors and Ergonomics Society, 34(4), 429-439.
[4] (2016).Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. Frontiers in Human Neuroscience,10,539.
[5] Li Linjun & Li faces south. (2019). Flight load assessment techniques and methods Helicopter technology,(02),37-41+46.
[6] Li Li (2024). A Study on the Evaluation of the Workload of Flight Training for the ACPC (Master's Thesis, China Civil Aviation Flight Academy). masterhttps://doi.org/10.27722/d.cnki.gzgmh.2024.000325