Pilot's mental workload nonlinear correlation among objective and subjective measurements

A Alaimo¹, A Esposito¹,* and C Orlando¹

¹Università degli Studi di Enna Kore, Enna, 94100, Italy

E-mail: Antonio.esposito@unikore.it

Abstract. In the aviation framework, the workload evaluation is influenced by human factors effects both for pilots and ground operators. Nevertheless, the measurement of workload is of great importance to prevent human errors due to fatigue. In the present work, correlation analysis has been carried out to investigate the relationship between subjective and objective workload measurements. An experimental campaign was conducted using a full flight business aircraft simulator. Particular attention has been paid to the NASA-TLX subscales contribute. In terms of objective measurements, three different indexes gathered from heart rate variability have been considered. As the main result, the frequency-domain index reveals a large correlation effect with the workload subscale for the maneuvers taken into account.

1. Introduction

The mental workload can be assumed as the amount of brain activity per unit of time, namely, the resources brain's rate occupation or the ability to process the information during the done work activity [1]. Three factors are believed to have a strong influence on the workload: the time, the intensity of the action (i.e., the difficulty or complexity level that requires resources in terms of attention) and the psychophysiological functional state [2]. Therefore, the mental workload is a multidimensional concept, i.e., a framework that cannot be studied directly but can be deduced from different and quantifiable variables through physiological, cognitive, psychological, behavioral, and other disciplines [3, 4]. Workload assessment is recurrent in aviation, both for pilots and ground staff, such as air traffic controllers. Today's search for workload evaluation methods is based on three main criteria: i) process of evaluating work performance; ii) subjective evaluation; iii) physiological assessment method.

The performance measurements alone cannot be a reference as an adequate metric for attributing the workload because the pilot's performance can be good even though a great level of workload has been requested. On the other hand, the pilot's performance cannot be adequate when his/her workload level is too low. Similar considerations can also be drawn in other aviation tasks such as Air Traffic Control (ATC) or Remotely Pilot Aircraft Systems (RPAS) and drone remote piloting activities. For such reasons, the objective and subjective aspects can support the indications derived from the performance indexes to evaluate the workload levels.

1.1. Subjective evaluation method

Generally speaking, subjective measurements include perceived stress levels recognition. The subjective evaluation method is focused on the participant's perception and results in correlation with
personal motivation. Subjective metrics are generally easy to use, as they are of the post-activity type; substantially, they do not interfere with work and have become a popular and most effective method for assessing mental workload [1]. This category includes both a one-dimensional and multidimensional scale. The first scale focuses on an overall measure of the workload, while the second concerns single components. One of the most popular aviation field measures is the NASA TLX, a multidimensional rating scale that allows operators to indicate their mental workload. A careful process based on experimentation has led to identifying six dimensions of the workload [5]. A weighted procedure gives information about each subscale to determine its contribution and nature on the perceived overall workload.

1.2. Physiological evaluation method

One of the primary references for mental workload evaluation based on physiological parameters can be traced back to the Spyker [6] research. It combines electrocardiogram, respiration, skin impedance, and electroencephalogram to analyze the workload of pilots. At present, the most used physiological indicators are based on brain activity, eye tracking, and electrocardiograms [4]. The analysis of physiological parameters measured by contactless devices has also been recently being evaluated [7].

Aricò et al. [8] have shown that theta, alpha, beta, and other EEG bands' energy are sensitive to mental workload changes. Activity-related eye indicators are widely used in processing human information such as perception, memory, reasoning, and thinking. Several ocular signal processing methods can be used to process primary eye markers [9] (pupil diameter, blink, blink rate) to explore the relationship between ocular markers and brain resource occupation induced by flight tasks. Zhang et al. [10] prove that the pupil diameter and eye-opening have significant differences under different mental loads during the flight simulation and the blink rate decreases.

Heart rate and heart rate variability (HRV) are indexes of the electrocardiogram (ECG). Closely related to the autonomic nervous system, HRV can reflect people's psychological stress and is sensitive to changes in the needs of the current activity and their efforts; for this reason, it is commonly used as workload tracking indicators. In 1971, Spyker [6] reported that the ECG characteristics are suitable for identifying the workload of pilots. In particular, it was found that the ECG was highly affected by tasks demands. Starting from the ECG heartbeat measure, different indexes can be derived in terms of heart rate variability in the time domain, in the frequency domain, and as nonlinear index associated with the acquired data's statistical characteristics.

2. Experimental set-up and measurement

A group of 23 pilots was taken into account for the study. Before the test, each pilot filled the privacy form with the authorization to wear the chest belt heart sensor and declaration about no use of heart drugs. Concerning the maneuvers that characterize the simulator flight test, a certified instructor informed the pilots of the whole procedure during a pre-flight briefing. In the present work, the analysis was focused on two segments: takeoff with the initial climb phase, namely "TO", and the final approach with the landing phase "LA". These segments were recently investigated in terms of flight performance by the Authors [3]. In the present work, attention has been paid to correlation analysis between the measured subjective and objective indexes.

For evaluating subjective workload perception, the NASA TLX was used with the evaluation of subscale weighting. In terms of objective measurement, to assess the workload level, it was decided to consider an index for every research domain. It was decided to use the standard deviation of all normal-to-normal beats interval SDNN [11] as a parameter in the time domain. When considering an ECG signal that lasts for no more than five-minute, the heart rate variability can be characterized by three frequency bands: very low frequency (≤ 0·04 Hz), low frequency (0·04 – 0·15 Hz), and high frequency (0·15 – 0·4 Hz). It has been found in [12] that, compared to other frequency bands, the middle-level frequency band of the heart rate is more sensitive to mental load. In this work, importance has been given to the low to high-frequency ratio LF/HF [11]. Finally, among the so-called nonlinear indexes, it was decided to evaluate the SD1 [11] parameter, which identifies the ellipse's
minor axis generated by the Poincaré plot. More in detail, when the workload increases, the SDNN, and the SD1 reduce, while the LF/HF increases.

3. Results
The primary purpose of the present work is to identify a correlation between the NASA-TLX and the HRV parameters with particular attention to the subscales. Before proceeding with the correlation analysis between the paired indexes, it is opportune to introduce the sample's characteristics through a boxplot. In Figure 1, the central mark indicates the median, and the outliers are plotted using the '+' symbol. Z-Score [13] standardization of data was used to put data on the same scale. The acronyms for the six subscales are Mental Demand (MD), Physical Demand (PD), Temporal Demand (TP), Performance (OP), Effort (EF), and Frustration (FR), respectively. Along with the six subscales, Figure 1 shows the overall workload "OW" computed by a weighted procedure [3] and the HRV measurements; Moreover, with a circled number, the corresponding weight is shown at the top of the six subscales obtained as the mean of all pilots sample. Comparing the TO and LA phases, it is possible to highlight for the TP, EF, and FR subscale that LA phase edges of the box (i.e., the 25th and 75th percentiles) are narrower than the TO phase. MD box for the TO phase presents a higher median value with respect to the LA phase, which is zero; meantime, the TP median in the TO phase is lower than the LA phase. This result suggests a general perception of the TO phase of higher mental demand associated with a slow exercise pace. Quite the opposite, the physical demand is higher for LA than TO since the final approach has a higher demand than the climb phase. The weighted procedure gives an overall workload with comparable median value and narrower whiskers for TO. The mean value of the sample, for OW and HRV measurements, are listed in Table 1. OW's mean value for the LA phase reveals a slightly higher workload level; this result agrees with HRV indexes' analysis. A higher LF/HF ratio is associated with a higher workload level; simultaneously, an increase in workload is related to a decrease of SD1 and SDNN. Therefore, the sample's subjective and objective mean values are in agreement in terms of overall workload estimate.

![Boxplot TO and LA phase for NASA-TLX subscale and overall workload.](image)

**Figure 1. Boxplot TO and LA phase for NASA-TLX subscale and overall workload.**

|       | OW  | LF/HF | SD1  | SDNN |
|-------|-----|-------|------|------|
| TO    | 56.31 | 4.91  | 8.02 | 20.27 |
| LA    | 59.85 | 5.51  | 6.86 | 19.22 |
With the aim of evaluating the subscale relationship with the HRV indexes, correlation analysis has been carried out and presented. Recent studies in simulator flight activities [3] have proven a non-monotonic relationship between objective and subjective measurements. Nevertheless, a monotonic relationship can be investigated using Spearman's rank correlation analysis after a transformation of raw data. Considering a set (X, Y) of variables, in this case, an objective and a subjective measurement, the strength of the correlation can be estimated and a nonlinear transformation of raw data can improve the correlation coefficient [14]. The search for mathematical functions that change the variables' measurement scales to maximize the correlation coefficient is the focus of this work.

No assumptions about the distribution of the variables were initially made because the procedure transforms the distribution of original measurements. A group of mathematical functions expressed by sinusoidal terms like equation (1) is tested for all the possible pairs between objective and subjective measurements for TO and LA phases. The three considered coefficient's search domains were set in the ranges 0 ≤ α ≤ 2, 0 ≤ b ≤ 3, and 0 ≤ c ≤ 2. In equation (1) I is the HRV index, the star is used to label the transformed data, i = 1, 2, . . . , 23 relates the variable to the pilot index while α = {TO, LA}.

\[ l_{I,\alpha} = a \cdot \sin(b \cdot l_{I,\alpha}) + \frac{c}{2} \sin(2l_{I,\alpha}) \]  

(1)

The results obtained for all the pairs are presented in Table 2 in terms of Spearman's rank correlation coefficient, putting into evidence the values of "r" greater than 0.5. Moreover, the star denotes a significance p-value lower than 0.05 and a double star p < 0.005. Only the positive "r" coefficient was considered as a consistent solution of transformation data. Identifying a robust function able to return low significance value as much as possible for the TO and LA phase is essential. For this reason, the results in Table 2 were obtained for the TO and the LA with the same set of coefficients for each considered pair. For completeness, Table 3 shows coefficients used in equation (1), highlighting the pairs for which the coefficient c is equal to zero. It is convenient to mention that a value of 0.3, in Table 2, implies a medium effect, and a value close to 0.5 implies a large correlation. By the analysis of results listed in Table 2, it emerges a high relationship between LF/HF ratio and NASA-TLX subscale, exception made for PD. Nevertheless, the lower p-value was obtained for the SD1 index in pairs with MD and EF; this result does not emerge from the overload workload analysis alone, for which a low/medium effect is reported.

| Phase | Index | MD   | PD   | TP   | OP   | EF   | FR   | OW   |
|-------|-------|------|------|------|------|------|------|------|
| TO    | LF/HF | 0.39 | 0.25 | 0.44 | 0.48 | 0.41 | 0.51 | 0.51 |
|       | SD1   | 0.56 | 0.30 | 0.26 | 0.34 | 0.57 | 0.43 | 0.24 |
|       | SDNN  | 0.33 | 0.24 | 0.34 | 0.36 | 0.40 | 0.29 | 0.33 |
| LA    | LF/HF | 0.32 | 0.24 | 0.37 | 0.52 | 0.34 | 0.53 | 0.50 |
|       | SD1   | 0.26 | 0.23 | 0.25 | 0.35 | 0.51 | 0.21 | 0.24 |
|       | SDNN  | 0.42 | 0.48 | 0.43 | 0.39 | 0.37 | 0.29 | 0.35 |

Table 3. Coefficients of equation (1) for each (HRV, NASA-TLX) pair.

| MD   | PD   | TP   | OP   | EF   | FR   | OW       |
|------|------|------|------|------|------|----------|
| TO/LA| LF/HF| (2.1,10.1) | (1.03,1.13,0.1) | (0.94,0.66,0.3) | (1.07,0.90,1) | (1.15,1.23,0.5) | (1.0,1.12,0.0) | (0.26,1.09,0.1) |
|      | SD1  | (1.4,40.0) | (0.99,2.95,0.1) | (1.68,0.96,0.1) | (0.70,1.08,1) | (1.27,1.33,1) | (1.0,1.34,0.0) | (1.50,1.25,0.1) |
|      | SDNN | (1.2,24.0) | (1.24,2.23,0.1) | (2.00,2.20,0.1) | (1.0,0.42,0) | (1.65,0.40,1.0) | (0.89,1.85,0.1) | (1.25,2.23,0.1) |

4. Conclusions
The correlation analysis conducted has shown that some of the subscales investigated by NASA-TLX have a more significant relationship than the values obtained from the pairs (HRV,OW). Indeed,
analyzing the (HRV, OW) pairs, the frequency domain index appears in a large relationship effect; nevertheless, subscales analysis presents better results for the SD1 index, particularly with mental demand and effort subscales in the TO phase.

The evaluated indexes' relationships cannot be considered unquestionable because they are obtained for two flight segments with a similar workload level. Future works could prove the statistical consistency of the mathematical functions found and corresponding coefficients values. Future research should be conducted on different maneuvers with different workload levels to validate the model presented in this work.

Acknowledgement

The study was funded by the Italian Ministry of Education, Universities and Research (MIUR) grants under the RPASInAIR project (P.O.N. Grant ARS01_00820 CUP J66C18000460005).

References

[1] Fan X, Zhao C, Hu H and Jiang Y 2019 Proc. Of the AHFE 2019 Int. Conf. on Applied Human Factors and Ergonomics Washington D.C., USA (Cham, Switzerland: Springer) pp 165-172
[2] Hendy KC, Liao J and Milgram P 1997 Combining time and intensity effects in assessing operator information-processing load Human Factors 39 (1) 30-47
[3] Alaimo A, Esposito A, Orlando C and Simoncini A. 2020 Aircraft pilots workload analysis: heart rate variability objective measures and nasa-task load index subjective evaluation Aerospace 7 (9) 137
[4] Liu J, Gardi A, Ramasamy S, Lim Y and Sabatini R 2016. Cognitive pilot-aircraft interface for single-pilot operations Knowledge-based systems 112 37-53
[5] Hart SG and Staveland LE 1988 Advances in psychology vol°52 (North-Holland Elsevier) pp 139-183
[6] Spyker DA 1971 Development of techniques for measuring pilot workload (Washington, D.C.: National Aeronautics and Space Administration)
[7] Alaimo A, Esposito A and Orlando C 2018 4th Int. Forum on Research and Technology for Society and Industry (Palermo, Italy ) (Piscataway: IEEE) pp 1-4
[8] Aricò P, Borghini G, Di Flumeri G, Colosimo A, Bonelli S, Golfetti A, Pozzi S, Imbert JP, Granger G, Benhacene R and Babiloni, F 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
[9] Duchowski AT, Krejtz K, Gehrer NA, Bafna T and Bækgaard P 2020 Proc. of the 2020 CHI Conf. on Human Factors in Computing Systems (Honolulu, HI, USA) (New York: Association for Computing Machinery) pp 1-12
[10] Zhang H, Wanyan X, Zhuang D and Wu X 2014 6th Int. Conf. on Intelligent Human-Machine Systems and Cybernetics (Hangzhou, China) vol°1 (Piscataway: IEEE) pp 110-113
[11] Electrophysiology TF of the ES 1996 Heart Rate Variability Circulation 93 (5) 1043-65
[12] Dellaux S, Delafarge A, Deharo JC and Chaumet G 2019 Mental workload alters heart rate variability, lowering non-linear dynamics Frontiers in physiology 10 565
[13] Salkind N. 2007 Encyclopedia of measurement and statistics (London: Sage Publications) vol°3 pp 1055-1058
[14] Tabachnick BG et al. 2007 Using multivariate statistics (Boston: Pearson) vol°5 pp 481-498