The Enhancement of Training of Military Pilots Using Psychophysiological Methods

G. Kloudova, M. Stehlik

Abstract—Optimal human performance is a key goal in the professional setting of military pilots, which is a highly challenging atmosphere. The aviation environment requires substantial cognitive effort and is rich in potential stressors. Therefore, it is important to analyze variables such as mental workload to ensure safe conditions. Pilot mental workload could be measured using several tools, but most of them are very subjective. This paper details research conducted with military pilots using psychophysiological methods such as electroencephalography (EEG) and heart rate (HR) monitoring. The data were measured in a simulator as well as under real flight conditions. All of the pilots were exposed to highly demanding flight tasks and showed big individual response differences. On that basis, the individual pattern for each pilot was created counting different EEG features and heart rate variations. Later on, it was possible to distinguish the most difficult flight tasks for each pilot that should be more extensively trained. For training purposes, an application was developed for the instructors to decide which of the specific tasks to focus on during follow-up training. This complex system can help instructors detect the mentally demanding parts of the flight and enhance the training of military pilots to achieve optimal performance.

Keywords—Cognitive effort, human performance, military pilots, psychophysiological methods.

I. INTRODUCTION

MILITARY pilots are experiencing a high cognitive load on an everyday basis due to the environment requiring them to adapt to a variety of very different conditions. It is a demanding workspace requiring the pilot to hold a lot of information in their working memory while still seeking other information to perform important decision-making processes and to pay attention to secondary tasks, such as answering a radio call. During a real flight, there are also stressors like noise, vibrations, light, and pressure changes that can lead to the depletion of mental resources. It is no wonder that the human factor plays a crucial role in aviation accidents, as 80-85% are caused by human error [19]. Automation technology has helped compensate for human limitations in terms of information processing, but the final decision is still in the hands of the operator who has to interface with the system to achieve optimal human-machine performance. It is not surprising that this kind of demand produces challenging tasks and mental workload for pilots.

The enhancement of a pilot’s performance primarily depends upon thorough training which is able to cover all of the probable circumstances involved. The effective training of technical skills is a must, but within the context of the work of professional pilots, there are even bigger challenges. In this stage of training, the need is to focus on the mental health of the pilots and their abilities to perform at peak performance level. For this purpose, the quantification of mental workload is crucial.

Mental workload is a complex construct and can be measured by a variety of metrics. The methods to measure the mental workload of the pilot are many, but it has usually been based on subjective questionnaires [8]. This type of evaluation is too subjective, and there is often a problem with distorted answers due to the pilots’ fears of being excluded from flying duty. Combining such data with psychophysiological data provides more objective results, and recent technological progress offers a nonintrusive way of collecting data in real time during all kinds of tasks of both simulated and real flight and do not interfere with the primary task of piloting. The accuracy of detecting their mental workload can be up to 90% if we include physiological data in the testing [18].

There are several psychophysiological methods used in the context of assessing human factor behavior. Primarily, they are focused on the measurement of the activity of the autonomous nervous system. The most commonly used and easy-to-apply method is HR monitoring [7], [12], [20], [21]. Increased HR can indicate increments of task difficulty, so the cardiovascular response is usually used to evaluate the mental workload in aviation under different flight conditions, both real and simulated [3], [10]. However, different variables, including muscle activity and anxiety, have to be taken into account in an evaluation during piloting.

The activity of the central nervous system can be seen by the cortical brain waves over the scalp that are measured by EEG. It’s one of the most reliable contemporary methods, although it’s mostly used under laboratory conditions given the usual environmental constraints. There are only a few in-flight studies [7], [16], [22] which have mainly captured data without muscle activity. The analysis of EEG waveforms is very complex, and most of the studies show that activity in the beta frequency band (β, 13–30 Hz) should indicate arousal, anxiety, alertness, and signs of acute stress [4], [23]. Alpha (α, 8-13 Hz) bands occur during relaxation and low attention levels and when anxiety is suppressed [11], [15]. Alpha activity decreases during complex and cognitively demanding tasks. Theta waves (θ, 4–8 Hz) are indicators of deep relaxation states and drowsiness, and they mainly occur during sleep states. Delta band activity is a tricky wave range,
because it is usually influenced by muscular activity. If
artifacts are carefully eliminated, it might indicate alertness or
a reaction to ground tasks [9], but it may also indicate fatigue
or a hypnagogic state [14]. Finally, in recent studies, research
has also focused on gamma band activities (γ, 30–70 Hz) that
are probably related to vigilance level, memory, situational
awareness, and other important cognitive tasks [17] that are
mainly present during flight.

Other methods used in the context of aviation are
electrooculogram (EOG) – it is a modern eye-tracking system
which returns information about eye movements, pupil
dilatation, and eye blinking. Other good indicators of mental
workload are electrodermal activity (EDA) and biochemistry.

Based on our previous research [13], we focused later
research and data analysis only on HR monitoring and EEG,
which proved to be the most reliable and nonintrusive methods
to assess the mental workload of military pilots under both
real and simulated conditions.

II. SUBJECTS

The subjects in this study were five military pilots flying the
Mi-2 and Mi-171 helicopters in their first stages of training.
All of them were men aged 25-31 years who voluntarily
participated in this research. The subjects underwent a regular
full physical examination. They were all in good health, with
no cardiac disease, vision defects, or brain damage. They were
flight status ready, non-smokers and right-hand dominant.

None of them had ever suffered any chronic
disease. Their family situation and social network was stable.
They were all assigned a standalone identifier, so the
anonymity of the pilots was ensured.

III. METHODS

Electroencephalogram: To measure electroencephalogram
signals, we used bipolar EEG with a Multicap electrode cap
from GVB geliMed. This cap normally contains 19 electrodes
according to the international 10/20 system. Based on previous
findings, we extracted the least relevant electrodes to reduce
manipulation time and to increase the efficacy of
measurement. Electrodes from this cap are replaceable, flatter,
and more suitable for measurement under a helmet. The final
number of electrodes used was eight, including the two
reference electrodes M1 and M2 (see Table I). Occipital
electrodes (O1, O2) proved to be effective in the
differentiation of mental workload. Then this was continued in
a toposcopic line using the central electrodes (C3, C4) and
frontal electrodes (F3, F4), which cover the most important
parts of the cerebral cortex. Placement of the electrodes in all
frontal parts (Fp1, Fp2) was not possible due to the higher
concentration of muscle artifacts and the restrictions of
wearing a pilot’s helmet. All of the recordings were made by
the Somte PSG ambulatory digital device, along with Somte
software which provides comprehensive analyses of the
physiological data.

Heart rate monitoring: The heart rate was recorded by the
Polar V800 advanced sports GPS watch. Based on the simple
data needed in this research, we moved from the classic but
bulky ECG device to this easy-to-use, compact watch. Heart
rate was measured by two electrodes placed on the sternum of
the subjects. Besides heart rate, it also captured heart rate
variability and provided for wireless online data collection.

Table I

<table>
<thead>
<tr>
<th>EEG Channel</th>
<th>Electrode</th>
<th>Reference electrode</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C3</td>
<td>C3 – M2</td>
</tr>
<tr>
<td>2</td>
<td>C4</td>
<td>C4 – M1</td>
</tr>
<tr>
<td>3</td>
<td>F3</td>
<td>F3 – M2</td>
</tr>
<tr>
<td>4</td>
<td>F4</td>
<td>F4 – M1</td>
</tr>
<tr>
<td>5</td>
<td>O1</td>
<td>O1 – M2</td>
</tr>
<tr>
<td>6</td>
<td>O2</td>
<td>O2 – M1</td>
</tr>
</tbody>
</table>

Questionnaires: To compare subjective ratings and
objective psychophysiological data, we created a post-flight
questionnaire. It contains questions about the mental workload
perceived during training flight and about task difficulty. To
cover most of the variables, the questionnaire also included
questions about the current mental and physical health of the
pilots, their social support, and relationships at work.

Another questionnaire was created for instructors to
evaluate each of the flight tasks according to difficulty and the
performance of the pilot. Difficulty was rated on a scale from
1-10, where 10 was the most difficult task such as take-off or
landing and the performance of the pilot was rated A (without
mistakes), B (some mistakes but still satisfactory) and C
(unsatisfactory).

To eliminate personality variables in the testing, we
included a personality questionnaire NEO-PI-R [6], which
contains the main scales of personality traits, such as
neuroticism, extraversion, agreeableness, conscientiousness,
and openness to experience. Another set of 30 subscales
provided more detail in completing the remainder of the
personality profile. This method provided information on
general personality characteristics and their specific
personality traits.

IV. STUDY PROTOCOL

Data were collected at LOM PRAHA’s Flight Training
Center (CLV) in Pardubice, Czech Republic. This training
facility offers complex services for both fixed wing and rotary
wing aircraft pilots. In this study, we only measured helicopter
pilots in their first stages of training using Mi-2 and Mi-17
helicopters during all airborne training and also during one
simulator training session.

Participants came to the briefing room in the morning and
started the day with a daily squadron meeting about the flight
plan and weather conditions. Later, they went to a separate
room where the researchers applied the above-mentioned
devices, such as the heart rate monitor and the EEG. Pilots
were instructed before the testing to be well rested, in good
health, and free of any prohibited substances. First, the
electrodes of the heart rate monitor that are attached to a strap
were placed on the pilot’s sternum. Later, the Polar watch was put on the subject’s wrist, and the researchers waited for the devices to be connected. Subjects were then informed about the function of the watches and the signs that flash if it turns off or if there is some other problem. They were asked to report such an issue to testing personnel to avoid data destruction. To secure optimal testing conditions for the heart rate measurement, we captured the data 1.5-2 hours after eating and without any medication. It was not possible to meet all the requirements for the standard procedure according to Bayevsky [2], since the research was conducted in a real flight environment. After connecting the devices, we began to measure the resting heart rate while the subject was sitting quietly without any distractions for five minutes.

Next, the electrodes of the EEG were placed on the scalp of the measured pilot using eight electrodes incorporated in the EEG cap. The connection was potentiated by a special conductive gel which is applied in the electrodes with a blunt needle. After checking the right impedance, it was necessary to make a calibration of the EEG. The subject sat for two minutes with eyes open and two minutes with eyes closed to ensure there were changes during the closed-eyes phase and to have the resting phase data available for later comparison.

Once all the used devices were correctly set up, the pilot was picked up by his instructor and taken by car to the helicopter. During the flight, there was a psychologist or his assistant present to ensure the devices were properly functioning and to record data for the whole flight, including the flying conditions and the reactions of the pilot. There were different flight tasks depending on the stage of training. These ranged from simple take-offs to more difficult ones, such as autorotation and group flying. On average, the training flight lasted about one hour and navigation flights up to two or two and a half hour. In total, all of the pilots flew at least 10 measured training flights. After landing, both the pilot tested and the researcher went back to the examination room where all of the devices were removed. Then the pilot sat down to a computer and filled out the post-flight questionnaire. Meanwhile, the researchers downloaded and saved all of the data. After finishing the questionnaire, which took about 10-15 minutes, the pilot could leave the room and continue with his duties. The final part of the research was done by the instructor who filled out an evaluation of the pilot after the training based on his notes and recordings of the flight.

V. RESULTS

The aim of this study was to find out if psychophysiological methods are useful for improving the training of military helicopter pilots. For that purpose, pilots were measured during actual flight to observe their level of mental workload during specific flight tasks. Based on previous research, it was decided to create an individual model for each pilot to better target the training and to ensure greater effectiveness.

First, the data from the heart rate monitor were analyzed. It was necessary to convert the data into the same format at a sampling frequency of 5Hz. In some of the recordings, there were technical problems, so there were finally 25 recordings made from five subjects. Heart rate and EEG data were organized according to the number of the flight (#001-…), date (YY-MM-DD), and pilot (A-E). Later, the data from the heart rate monitor were processed and displayed in a graph where it was possible to compare them with flight conditions (see Fig. 1). To see if there were significant heart rate differences between the resting phase and during flight, the means of these two situations were compared. The average heart rate of the resting phase was compared to the heart rate during flight using a t-test that showed significant differences for p<0.01. Also, data from the EEG were put into the same format and timeline. First, it was necessary to remove artifacts, mostly eye movements and muscle activity. This elimination was made off-line before data averaging using an automated method and later proofed manually. The automatic method was based on an algorithm searching for outliers in the recordings. This was made by the Grubb test [1] at a level of significance where p<0.05. These segments can be expected to have a high probability of artifacts, are not used in further analysis, and do not have any effect on the final classification model. Artifact rejection is a key part in the EEG analysis to clear all the data that will later be compared. To better read the signal, the EEG data were shown in a spectrogram where the blue color represents the minimum and the red color the maximum in the measured power band. On the vertical axis, frequency is shown. On the horizontal axis, the color scale is represented by the module of a complex spectral function. The basis for calculating the complex spectrum of each segment is the discrete Fourier transformation (DFT). In the artifact parts where too much muscle activity occurs, either the electrode fell off or there were other technical problems which are marked in the spectrogram as vertical lines. The artifact parts correlate with increasing heart rate, so it is obvious that there are mainly muscle type artifacts. During flight, the artifacts mostly occurred at frequencies of 4Hz, 7Hz, 12.5Hz, 25Hz, 38Hz, 42Hz, and 50Hz. So, later on, we focused on parameters excluding these frequencies. For further analysis two frequency bands were used, one in the beta waveband (16-22Hz) and another one in the gamma waveband (28-35Hz). A spectrogram with a spectral analysis for the EEG signal that was measured between two electrodes, e.g. C3-M2 (see Fig. 2), is shown below. At a higher resolution, it is possible to see the critical parts simply by visual analysis. But, for the purpose of higher quantification, a statistical analysis was made to prove there was a significant difference between the resting phase and real flight.
Fig. 1 Heart rate of Pilot A during specific flight situations
The statistical analysis was made using all of the parameters of EEG and HR. Each parameter was later compared between the resting phase and periods of demanding flight situations from all of the recordings. First, the Shapiro-Wilk normality test was used and later the t-test to compare the two phases. The results of this test are presented in Fig. 3, which shows that the most significant features were: C4-M1 LOG BETA, C4-M1 LOG GAMMA, O1-M2 LOG GAMMA, and the HR MEAN. This means that for future analysis these are the best parameters for the detection of mental workload. Still, the accuracy of classification was not very high, so it was decided to design a more individualized model based either on types of flight situations or individuals (pilots). The model based on flight situations was too large, since there were 17 situations and for each of them there were 2-3 features. For example, it was possible to characterize take-off by C3-M2_STD, HR_MEAN and O1-M2_GAMMA, but this would have been too complicated to use in practice. The individual model for
every pilot was based on a self-learning model, which was shown to be more effective and deterministic. For each pilot, a model was created which enabled to differentiate the resting phase from the demanding in-flight phase.

<table>
<thead>
<tr>
<th>Resting phase</th>
<th>Demanding flight situations</th>
<th>Significance level (p)</th>
<th>p &lt; 0.05</th>
<th>p &lt; 0.01</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1-ME-STD</td>
<td>0.21 ± 0.61</td>
<td>0.17 ± 0.22</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>C1-ME-LOG BETA</td>
<td>0.4 ± 0.37</td>
<td>0.28 ± 0.18</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>C1-ME-LOG GAMMA</td>
<td>0.3 ± 0.15</td>
<td>0.28 ± 0.15</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>C4-ME-LOG BETA</td>
<td>0.52 ± 0.61</td>
<td>0.36 ± 0.26</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>C1-ME-LOG GAMMA</td>
<td>0.49 ± 0.14</td>
<td>0.29 ± 0.16</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>PS-STD</td>
<td>0.31 ± 0.10</td>
<td>0.18 ± 0.02</td>
<td>0.014</td>
<td>0.014</td>
</tr>
<tr>
<td>PS-LOG BETA</td>
<td>0.43 ± 0.13</td>
<td>0.32 ± 0.13</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td>PS-LOG GAMMA</td>
<td>0.41 ± 0.16</td>
<td>0.37 ± 0.18</td>
<td>0.056</td>
<td>0.056</td>
</tr>
<tr>
<td>PM-STD</td>
<td>0.27 ± 0.11</td>
<td>0.21 ± 0.04</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>PM-LOG BETA</td>
<td>0.44 ± 0.11</td>
<td>0.31 ± 0.10</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>PM-LOG GAMMA</td>
<td>0.5 ± 0.11</td>
<td>0.35 ± 0.18</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>DI-ME-STD</td>
<td>0.22 ± 0.10</td>
<td>0.10 ± 0.02</td>
<td>0.024</td>
<td>0.024</td>
</tr>
<tr>
<td>DI-ME-LOG BETA</td>
<td>0.36 ± 0.01</td>
<td>0.34 ± 0.23</td>
<td>0.009</td>
<td>0.009</td>
</tr>
<tr>
<td>DI-ME-LOG GAMMA</td>
<td>0.39 ± 0.04</td>
<td>0.44 ± 0.36</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td>DI-PS-STD</td>
<td>0.28 ± 0.14</td>
<td>0.27 ± 0.01</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>DI-PS-LOG BETA</td>
<td>0.45 ± 0.13</td>
<td>0.35 ± 0.16</td>
<td>0.077</td>
<td>0.077</td>
</tr>
<tr>
<td>DI-PS-LOG GAMMA</td>
<td>0.38 ± 0.10</td>
<td>0.42 ± 0.12</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td>DI-MEAN</td>
<td>0.31 ± 0.12</td>
<td>4.7 ± 1.7</td>
<td>0.002</td>
<td>0.002</td>
</tr>
</tbody>
</table>

Fig. 3 Statistical analysis of the counted features

This model is based on a decision tree [5] which was created by a CART (classification and regression tree) algorithm. The decision trees used for this model were binary, so in each non-leaf node the data were divided into two sets. The final decision trees were calculated from all available features, with the resting phase given as class 0 and all difficult flight situations where an increased mental workload was expected as class 1. All the data from each subject were divided into 10 parts and a final decision tree was compiled using a 10 fold-cross-validation. This final model may be used for the classification and detection of mental workload in new recordings. This system of modeling was used in past recordings, so it was possible to make a classification model of the level of mental workload for each pilot. Every individual had his own classification features based on EEG and heart rate monitoring recordings. It is a self-learning model, so with every new recording, the classification becomes more and more accurate. The minimum number of recordings needed for this model to be effective was estimated to be five recordings. However, even with just one or two recordings, it is already possible to classify the most mentally demanding parts of the flight. For future recordings, this easy-to-use application was created for the instructors to evaluate the flight regarding the presence of mental workload. However, it is necessary to train the personnel how to apply the devices so that the recordings can be evaluated. First, the recorded data has to be downloaded from the devices. Next, the data has to be properly formatted, and then it is uploaded to the application. The application automatically rejects artifacts and shows a graph of the most mentally demanding parts of the flight. Based on these results, the instructor, or even pilot himself, can see which part of the flight was the most difficult and would be appropriate to replicate and focus on during future training flight conditions.

Finally, the post-flight questionnaire was used to evaluate the development of the pilot during training. The physical and mental state of the pilots was stable through all the phases of training with slight deviations that did not influence measurement of the psychological data. Also, their social support and work relationships were evaluated as good and did not change during the course of training. It was determined that no extreme situations occurred during the training that could have influenced the reliability of data measurements. The pilots also evaluated the difficulty of the flight and their level of mental workload during specific flight tasks. The tendency was very similar in all of the individuals. Initially, training was perceived to have a high level of difficulty. However, the perceived mental workload fell as the training continued. In the middle stage of the training, tension slightly increased again. During the final phases, tension fell to a moderate level. The evaluation made by the instructors was mainly used for the pilots to see their progress in the training. For use in our research, it was possible to compare data from psychophysiological methods, such as EEG and heart rate monitoring, to the performance of the pilot. It was already proven in previous research [13] that a highly mentally demanding task is handled worse than a task without a perceived high mental workload. This was also confirmed in this research. This finding supports the theory for the need of monitoring mental workload during flight training to enhance pilot performance. The results from the personality inventory NEO-PI-R did not show any signs of destabilization of the mental health of the pilots nor acute tendencies for a radical change in the future. The group of pilots tested was largely homogenous and indicates a good pilot selection process regarding their mental stability.

VI. CONCLUSION

Military aviation is one of the key components of modern defense forces therefore enhancing and optimizing its performance is critical to military success. The technical aspects of military aviation are definitely important, but the human factor still plays a crucial role in innovation. As stated above, human error causes 80-85% of aviation accidents. To avoid such tragedies, a military pilot has to be ready to perform in extreme conditions while fulfilling all kinds of specific flight tasks and stay focused on solving any unexpected problems. The training and requirements for a military pilot do not differ much whether the pilot flies a military helicopter or a fighter aircraft. They all need to go through specific flight tasks during training to become elite pilots. As technological progress in the automation of aircraft increases, technical training needs to keep pace by focusing on the pilot’s mental resilience. This includes the capability of handling stress, reducing fatigue, staying focused, and being precise and quick in decision-making processes.

During flight, it is difficult to determine if the pilot is
handling the flight situation calmly or if he’s going through a
lot of stress. If there is an objective tool that can choose key
parts of the flight that should be trained on more thoroughly, it
can help the pilot to gain confidence and to automate
completion of the learning process. This tool is based on
physiological data such as EEG and heart rate monitoring. The
psychological questionnaires supported the hypothesis about
decreasing tension during training and the need for objective
measures of mental workload.

First, it was necessary to process data from the recordings
and to convert them into same format. Later, the problem with
artifacts from EEGs regarding muscle activity and other
interference had to be resolved. For this purpose, an automated
rejection process of artifacts was created, which was later
incorporated into the final application. For each pilot, an
individual model was created based on chosen features
regarding mean heart rate and frequency band on the EEG
electrodes. This model is self-learning. So, with every new
recording, it gets better and better in detecting the mental
workload of the pilot during flight. This calculated model is
already incorporated in a hands-on application which trained
instructors could use on an everyday basis. One complication
is in the measurement itself which, without proper training, it
would not be possible to make. For example, the EEG has its
own specific application methods that are important in the
avoidance of artifacts, the displacement of the electrode, and
other signal interference. So far, it is not possible to make
this measurement without the presence of one of our professionals.
However, in the future, there is hope to train instructors or the
pilots themselves to incorporate this new technique in their
training as a routine procedure. To ensure that the data
obtained are not influenced by other variables, such as
physical or mental state, social support, working environment,
or some unusual personality characteristics, a series of
questionnaires were included in the testing. All pilots proved
to be mentally healthy without any significant deviations
throughout the testing.

The aim of this study was to find new, innovative ways to
improve the training of military pilots. Since the focus is now
turning from the machine towards human-machine interface, it
was only appropriate to try to use the psychophysiological
measurements of the pilot and his level of mental workload.
For this purpose, we created an objective model that could
help instructors and their pilots to enhance their performance
and to easily handle their working tasks. Its advantage lies in
the individually selected features that help to specify the
indicators of mental workload for each pilot. Small obstacles
can occur during the application of the devices which still
requires a well-trained professional. Now, future effort calls
for an empirical investigation of the efficacy of this presented
model and of finding possible ways of increasing the
resilience and mental health of military pilots to lead to
optimal human performance.

REFERENCES
Outliers in Statistical Data. Wiley.