Optimizing Performance Based Training: Monitoring the Flow of Cognitive Load based on Psychophysiological Measurements in a Fighter Cockpit Simulator

Maykel van Miltenburg
1059 CM Amsterdam
THE NETHERLANDS

Maykel.van.Miltenburg@nlr.nl

Jur Crijnen
1059 CM Amsterdam
THE NETHERLANDS

Jur.Crijnen@nlr.nl

ABSTRACT

Current training of fighter pilots is, almost without exception, designed for a fixed number of hours and on a specific time schedule. Performance Based Training is a training concept aimed at optimizing training, preferably in a personalized way. It is about preventing training/performance gaps beforehand, instead of solving them afterwards. Effective personalized learning assumes an optimal level of difficulty in the learning task provided. Therefore, an optimum load model is developed capable of classifying a pilot's cognitive load in real-time based on various cognitive load metrics. The study was set up to test electroencephalography (more specifically, the individual upper alpha band power and theta band power) as one of the cognitive load metrics of this model in a fighter cockpit environment. A total of four participants took part, all of whom were former F-16 pilots. Each of the participants performed three sessions with multiple runs. The cognitive load is expected to be higher during the first run (Retention Test) compared to the last run (Performance Test) within each session. While performance and subjective workload are respectively higher and lower during a Performance Test compared to a Retention Test, the cognitive load metric showed mixed results between both tests which could be attributed to high inter- and intra-individual differences.

***Keywords**: Performance Based Training, cognitive load, electroencephalography, fighter cockpit simulator, retention interval, tactical intercept*

1.0 INTRODUCTION

Current training of fighter pilots is, almost without exception, designed for a fixed number of hours in a specific time schedule for qualification training programs as well as annual training programs. Generally, deviations from the programs relate to organisational demands, less to individual pilot demands. Performance Based Training is a training concept aimed at optimizing training, preferably in a personalized way, such that relevant training events are offered at the correct time and with the correct resources. It is about preventing training/performance gaps, instead of solving them afterwards and can be used to aim for the highest personal standards, instead of ensuring the minimal standard. Performance Based Training requires advanced techniques for measuring and recording performance and behavior of both pilot and system. It also requires advanced analysis techniques. Both types of techniques are hardly used in practice for a variety of reasons. They require expertise that is lacking in the organisation, they take time to use, and they may influence the execution of the task (they are "intrusive"). We anticipate these limitations will disappear as technology advances in the next decade.

In the search for optimal learning conditions, researchers from education sciences have developed the concept of personalized learning. Similar to the successful development of personalized medicine, personalized learning seeks to identify genetic, neural and behavioural predictors of individual differences in learning and aims to use predictors to help create optimal teaching paradigms [1]. Effective personalized learning assumes, at least, an optimal level of difficulty in the learning task provided. The optimal task difficulty relates to a balance of performance and cognitive load ('Cognitive Load Theory') [2]. An adaptive level of task difficulty is required to maintain the balance while learning is progressing. Suboptimal task difficulty leads to ineffective training, e.g., presenting experts with a training suitable for novices for example has been found to adversely affect their learning progress [3]. The balance can be achieved faster by enhancing motivation [4]. These findings are consistent with Vygotsky's Zone of Proximal Development for children [5] and the Flow concept of Csíkszentmihályi [6] as applied in game design [7]: the 'flow channel' between anxiety and boredom (see Figure 1). This latter concept stresses the importance of a bandwidth in which the task difficulty fluctuates between challenging and easy activities, while avoiding states of extreme frustration or boredom. Such progress may be stimulating not only for gaming, but for any type of activity, including learning as shown by Kiili [8].

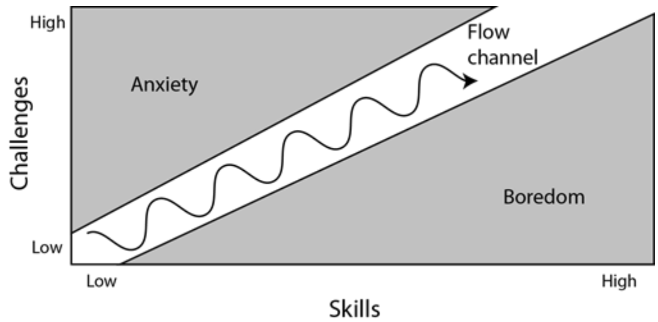

Figure 1: Flow Channel based on Schell (2014).

The optimal level of task difficulty and its related flow channel greatly differs between individuals, and therefore would be a promising candidate to incorporate in a personalized learning environment. However, the optimal level of task difficulty for a particular individual is also mostly unknown. It depends on a number of individual and environmental variables, and the subjective appraisal thereof, both at the cognitive (e.g., task performance, cognitive load) and the emotional (e.g., stress, anxiety, fatigue) levels, which are often inaccessible to consciousness. Currently, instructional designers, instructors, as well as game designers make assumptions about the optimal task difficulty for their audiences and try to adapt as needed. The relevance of using either performance measures as well as mental states is acknowledged but these measures often are inaccurate and subjective. Also, despite growing availability of a wide range of learning analytics techniques, these are hardly employed to deliver personalized training and they are mostly developed for post-hoc adaptation: providing feedback after an assignment or selecting the next learning task [9], not adapting the learning task during the task performance. However, selecting the optimal next learning task should be considered most important. Adapting the learning task in real-time should be no more than finetuning since too big adaptations are undesirable and will also not be feasible in practice.

1.1 Measures from the brain

Psychophysiological measures have been used in an attempt to provide objective measures for mental states such as cognitive load. They can be obtained fairly easily and noninvasively from the surface of the human body are characterized by sufficient individual variation [10] to warrant their use in personalized learning. Much is known about the physiological basis of most psychophysiological measures, as well as their relation to psychological states and processes, but an overall theoretical framework within the context of (personalized) learning is lacking.

As learning and task difficulty are influenced by cognition as well as emotion, it is relevant to incorporate measures at both levels in personalized learning. In the present context, measures from the brain (i.e. electroencephalography; EEG) and the autonomic nervous system (i.e. electrocardiography; ECG) are most important, because they can be consistently related to behaviour. Measures from the brain mostly tap on cognitive processes, whereas measures from the autonomic nervous system are mostly related to emotional states, although this distinction is by no means very sharp. In the EEG, we will focus on the bands most commonly associated with cognitive load, viz. theta and upper alpha. Reduced power in the (upper) alpha band and/or increased power in theta band covary with cognitive load in a variety of tasks [11, 12].

1.2 Individual differences

There are individual differences in EEG in either rest or task conditions. These are contributed, to a large extent, to age and genetic factors. Some researchers advice to identify individual peak frequencies for bandwidths [10], which varies over time within a participant (SD of 1 Hz), but certainly varies between participants (SD is 2.8 Hz). Not only does the magnitude of alpha power vary between individuals, the location of the alpha peak power varies as well, and therefore also the definition of the alpha sub-bands. In a healthy population of adults, the average *individual alpha frequency* (IAF) is usually thought to be around 10 Hz with a standard-deviation of 1 Hz [13]. The definition of the sub-bands 8-10 and 10-12 Hz is based in this finding. However, a standard-deviation of 1 entails that 95% of the IAF across individuals lie within a confidence interval of 8 and 12 Hz even in a relatively homogeneous population of young adults. This suggests that the alpha sub-bands should also be personalized so that an individual with an IAF of, say 8.5 Hz works with a low alpha band of 6.5-8.5 Hz and a high alpha band of 8.5-10.5 Hz. Individualized (i.e. IAF-based) alpha sub-bands have successfully been used before [14] but to our knowledge never in an (applied) training setting.

1.3 Optimum Load Model

Effective personalized training assumes, at least, an optimal level of difficulty in the task provided [2, 7]. Prior to the study described here, a first version of our optimum load model which is capable of classifying a pilot’s cognitive load in real-time while performing the Multi Attribute Task Battery (MATB-II) [15] based on electroencephalography (EEG) and heart rate (variability) signals was developed. The MATB-II consisted of five different levels with an increasing complexity. The resulting optimum load model consisted of three different layers: 1) simple rules and metrics including a multimodal approach in which different response systems (i.e. EEG and ECG) are analysed, 2) more complex criteria by using different classifiers (like k-Nearest-Neighbours, Support Vector Machine or Random Forest) based on EEG data and 3) deep learning models using Convolutional Neural Network called EEGNet (also based on EEG data).

The goal of this study is to test these three different layers of the optimum load model within a more realistic setting. The current paper describes a small scale experiment focussing on only the first layer (rules & metrics) of the optimum load model given constraints common for simulator training. Measurements from the brain (EEG) and autonomic nervous system (ECG) are collected on aggregated and individual level. However, only the EEG measurements will be analysed within current paper.

1.4 Hypotheses

The hypotheses are that the individual upper alpha band power will decrease with an increasing cognitive load [16] and that the theta band power will increase with an increasing cognitive load [17, 18]. In addition, the cognitive load is expected to be higher during the first run of a session compared to the last run of a session.

2.0 APPROACH

Participants were provided a total of three F-16 simulator sessions with varying retention times in between. included. Each session consisted multiple runs. The first run after a retention interval is called Retention Test (RT). The last run of the session is called Performance Test (PT). The first session consisted of multiple

Tactical Intercept (TI) runs with an increasing complexity. The second and third sessions differed from the first session in that they started and ended with a run of equal complexity. The pilots are scored for the PT and RT on performance criteria, which are discussed in Paragraph 2.4.

2.1 Participants

A total of four participants with a former position as F-16 pilot participated. They varied in age (37 – 50 years), flight experience, and diversity in fighter pilot retirement ranging from five months to twelve years. Two out of four pilots still perform flights on a regular basis as airline pilot, however this is not expected to impact the performance as the skill researched within this paper has only limited overlap with civil flight operation skills.

2.2 Experiment sessions

The goal of the TI sessions is to train the participants, independent of their fighter pilot skills prior to the experiment, to be "current" on the single ship TI task. Within this task the fighter pilot is commanded to eliminate hostile aircraft while being assisted by fighter control. Although the environment is known to the participants, the single ship TI is a non-standard mission type as regular TI's are performed in a multi ship formation. Since the relation between skill improved and cognitive load has our interest it is favorable to have no interreference from task dependent experience. Additionally, even the most skilled pilots will be able to improve themselves in executing this mission type due to its novelty. The single ship TI task therefore allows to take measurements. In each session the pilot is trained by means of multiple runs with different complexity. The level of complexity is determined by the amount, the formation, and the speed/altitude of the hostile aircraft. Within each session the complexity is increased and under supervision of a flight instructor the runs are debriefed to determine points of improvement. Before the start of first run, the pilots first familiarised with the setup of controls and cockpit layout. The experiment design is visualized in Figure 2.

Throughout the runs, baseline and activity measurements are taken for the EEG. In each run, the baseline captures approximately two minutes of data ranging from ten seconds after the pilot is 'released' in the air until the fighter control announces the presence of hostile aircraft ("*New Picture*"). Activity measurements range from 40 seconds prior to the release of a missile from the participant to 20 seconds after. This timeframe is expected to indicate the highest amount of cognitive load because prior to weapon release the pilot has to achieve a radar lock on the enemy and afterwards has to maintain this lock for several seconds before performing a defensive maneuver to lower vulnerability towards the opponents' shots. If activity timeframes overlap because multiple shots are released within the 'prior to' or 'after' missile release window, the measurements are stacked together in a larger timeframe.

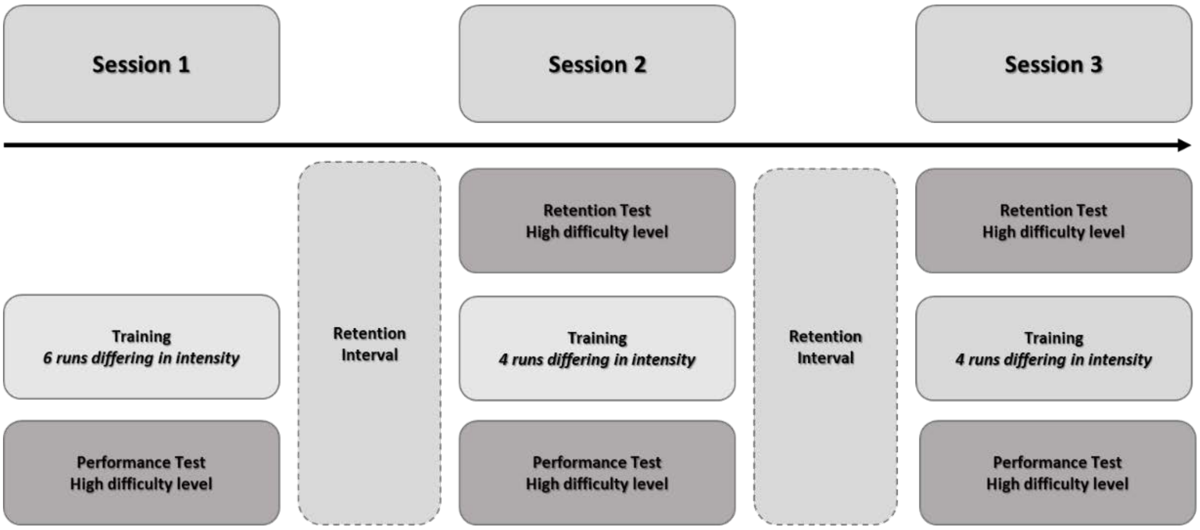

Figure 2. Experiment Design

2.2.1 Retention interval

In between the TI sessions a varying interval time is assigned to the pilots. Retention intervals varied between 33 to 85 days. The retention interval in between session 1 and 2 is on average 48 days (SD = 18) and for session 2 to 3 62 days (SD = 15). Especially scheduling restrictions limited tighter and less dispersed intervals between the sessions.

2.3 Procedure

During the TI sessions the participants were first briefed on the tasks they were trained on. In between all test and training runs the scenario was debriefed to emphasise on the point of improvement for the upcoming runs. After each run the participant was asked to fill in the NASA TLX form.

2.4 Measures

During each run, EEG data is collected and simulator data is logged. After each run, subjective workload is assessed with the NASA-TLX. Retention time is logged based upon the total amount of days between the different sessions for each individual participant.

After all sessions, the three PTs and two RTs for each pilot are graded according to their *Flight Geometry*, *Weapon Management* and *Rules of Engagement/Communication* skills. These criteria are fundamental to a successful completion of a (single ship) TI and are defined in cooperation with the grader, a former F-16 pilot and weapon instructor. *Flight Geometry* yields the positioning of the aircraft with respect to the opponent. This is crucial to get the opportunity to attack the hostile aircraft and involves tactic skills. Accomplishment of radar lock, shooting and hitting the enemy by the participant is considered *Weapon Management*. Finally, *Rules of Engagement/Communication* comprises adherence to the combat rules, and indicates the level of competence in communicating with fighter control and acting on cues. The three criteria are graded by expert judgement according to a five-point Absolute Category rating scale which ranges from bad performance to excellent Performance. The expert is provided with a Common Operational Picture (COP) view and radio communication recording of the sessions. The displayed information involves the participants' and enemies' aircraft positioning, speed, and weapon release together with hit result. The PTs and RTs are provided in a random order and the expert is not aware about the run being a PT or RT while grading. The training runs in between the RT and PT are not graded as these are of a less demanding levels which therefore could interfere with the data.

For electrophysiological data acquisition, the EEG machine TSMi SAGA32 was used to record brain activity. The electrode selected for alpha rhythm was Pz and the one for theta Fz [19–21]. To research the hypotheses, alpha is investigated further: Besides using the total range of alpha, it is also subdivided into a lower (8-10Hz) and a higher alpha (10-12Hz). Furthermore, the individual peaks of the alpha bands are determined to locate the Individual Alpha Frequency (IAF). Additionally, this IAF is divided into a lower and higher IAF according to the peak. The range of lower IAF is defined as IAF-2 to IAF, whereas the range of higher IAF is defined as IAF to IAF+2. This resulted in six measures of alpha: total alpha, lower-alpha, higher-alpha, total IAF, lower IAF, and higher IAF. For each run, only the relative band power higher IAF (i.e. IAF to IAF+2Hz) and relative band power theta (i.e. 4-8Hz) are analysed.

Since there are many individual differences in EEG in either rest or task conditions, band power will be reported by using the difference (i.e. delta) of the baseline and TI (i.e. TI minus baseline). We assume intra-individual differences in EEG, so before each TI run there will be new a baseline.

After each run, the NASA-TLX workload rating scale [22] was performed to take measurements of the participant's subjective feelings of workload. The overall task load index is measured by averaging the results on these subscales (without any weighting factors). Research [23] have found in two studies that the NASA-TLX demonstrated validity and reliability.

3.0 RESULTS & DISCUSSION

Analysis showed promising results as was expected according the optimum load model. The hypotheses were that individual upper alpha band power decreased and theta band power increased with increasing cognitive load. Besides, a decreased cognitive load within RTs compared to PTs was expected as well. First, we will present and discuss the aggregated (i.e. all participants) and, second, the participant-dependent results. Sample size is too low to perform statistical analysis so it should be taken with caution and only as a trend analysis. More sessions are planned in the future.

3.1 Aggregated Results

Before discussing the results on the cognitive load based on EEG, we will start with the aggregated results on NASA-TLX and Performance Assessment.

3.1.1 Subjective workload

The NASA-TLX score in Figure 3 indicates the subjective workload perceived by the participants on each first and last run. On average, it can be seen that for the PTs (i.e. after training) the workload is graded lower than for the RTs (i.e. first run of the session after a retention interval). This is according to our expectations as the pilots are tasked to perform a relatively difficult exercise after the retention interval only with a short familiarization to prepare.

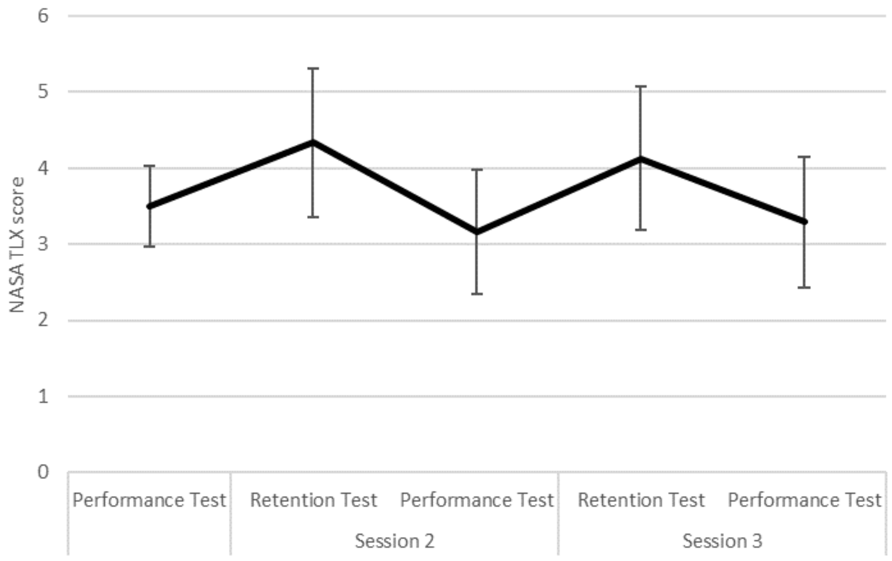

Figure 3. Result for each Test on the average (aggregated) NASA-TLX score. Error bars display the Standard Error.

3.1.2 Performance

The pilots' performance of both the PTs and RTs are indicated in Figure 4. The bars represent the criterium dependent aggregated scores and the average for the three criteria is shown as a line. As expected, generally the performance drops after a retention interval, which is indicated by the two 'dents' in the line. Training runs appear to restore ones' skills before they are tested during the PT. Noticeably is also the lower average performance increase during Session 3 compared to Session 2, as well as the overall peak performance (PT) in Session 3 remaining lower than Session 2.

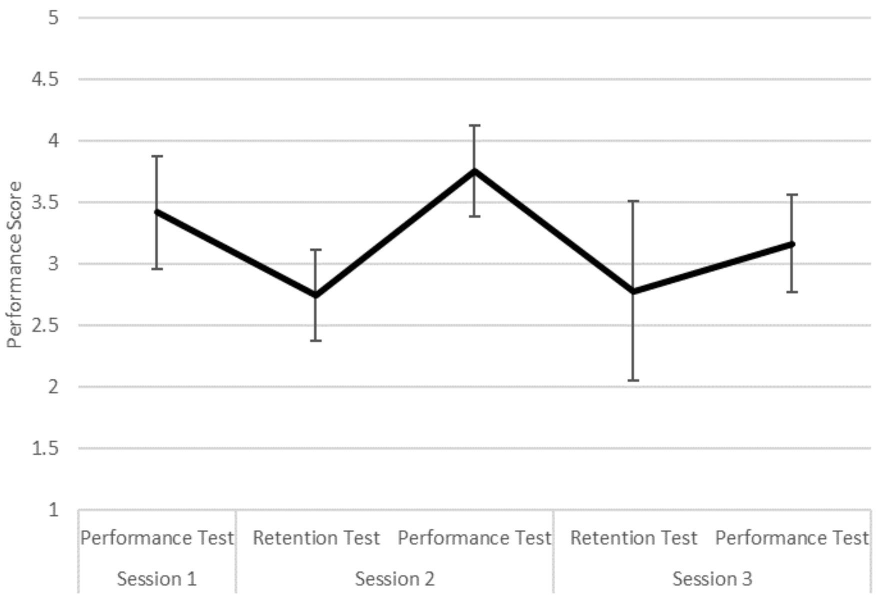

Figure 4. Result for each Test on the average (aggregated) Performance score. Error bars display the Standard Error.

3.1.3 Cognitive Load Metric: EEG

First, an analysis was conducted to compare the total alpha activity of the resting states for each participant. Only for each run of Participant 4, there was no visual difference in upper alpha activity during eyes closed compared to eyes open. For now, Participant 4 is considered Brain-Computer Interface (BCI) illicit [24] and is removed from further EEG analysis and results.

In Figure 5, the aggregated average performance and delta (i.e. Tactical Intercept minus Baseline) in relative band powers for individual upper alpha and theta are plotted during the PTs and RTs for each session. A positive delta means the power in TI was higher compared to the baseline within the same run. A negative delta means that the power in TI was lower compared to the baseline within the same run. For all runs, a positive delta in individual upper alpha band power and a negative delta in theta band power were expected as this would indicate increased cognitive load during the TI. Apart from both band powers in Session 1, the results confirm these hypotheses.

In addition, when comparing PT and RT more closely, a lower cognitive load for PTs than for RTs is expected, as training aids in both increasing performance and lowering the cognitive load to execute their run. Apart from the theta band power in PTs of Session 2 and 3 (compared to RT within the *same* session) and the upper alpha band power in PT of Session 3 (compared to RT within the *same* session), this hypothesis can be confirmed. The delta in both band powers in PT of Session 2 is *smaller* when compared to its band power in RT within the same (and even, but only for upper alpha, the next) session(s) which indicates that cognitive load is *decreasing*. This could be induced by a constantly improving performance in combination with a lowering cognitive load in each consecutive run.

However, this doesn’t appear to apply for all PTs and RTs. For the cases where the hypothesis cannot be confirmed, PT resulted in an increased cognitive load compared to the RT *within* the same session. As observed performance is lower during the RT the decreased workload (according to NASA-TLX) might reveal the reduced motivation and effort of the pilots to push the envelope of the TI and eliminate the hostile aircraft when this appears to be out of limits. The PTs’ increased delta in both band powers could illustrate the endeavour to push extra for scenario fulfilment. Interestingly, the results on NASA-TLX suggest the opposite. Therefore, a mismatch between actual (objective) cognitive load and perceived (subjective) cognitive load is identified. We need to look at the individual results to check this phenomenon.

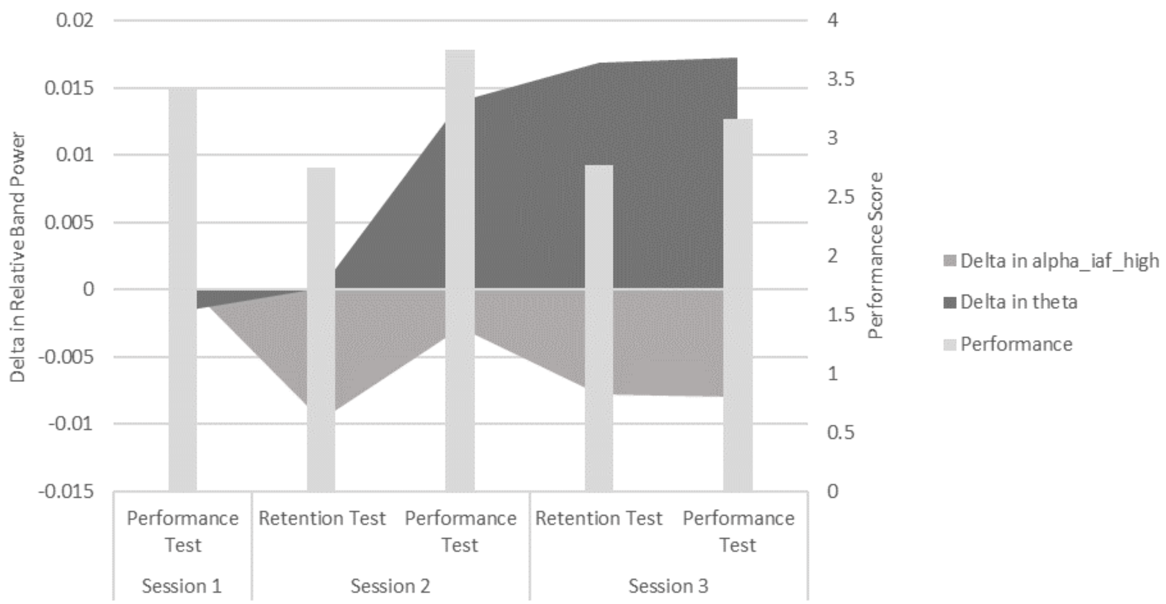

Figure 5. Average delta (TI minus Baseline) in relative band power of theta and individual upper alpha (i.e. alpha_iaf_high) compared to performance (Aggregated).

3.2 Participant dependent results

Results from one participant are discussed to zoom in on the suitability of individual performance – cognitive load assessment.

3.2.1 Performance and workload

In Figure 6 the performance assessment and subjective workload (NASA-TLX) of one participant is visualised. Within this diagram a sawtooth-like pattern is identified for the combined Performance Score (average of Flight Geometry, Communication / Rules of Engagement, and Weapon Management scores). The sawtooth illustrates a performance drop after a retention interval, and restoring of these skills during training to an equivalent or a higher level than before. The perceived workload by the participant is decreasing within the sessions, indicated by a lower NASA TLX score for the PTs (with exception from the first PT) compared to the RTs. Therefore, it appears that increased performance is negatively related to perceived workload. This meets the expectations, training improves performance and lowers the (perceived) workload when an equivalent task is executed.

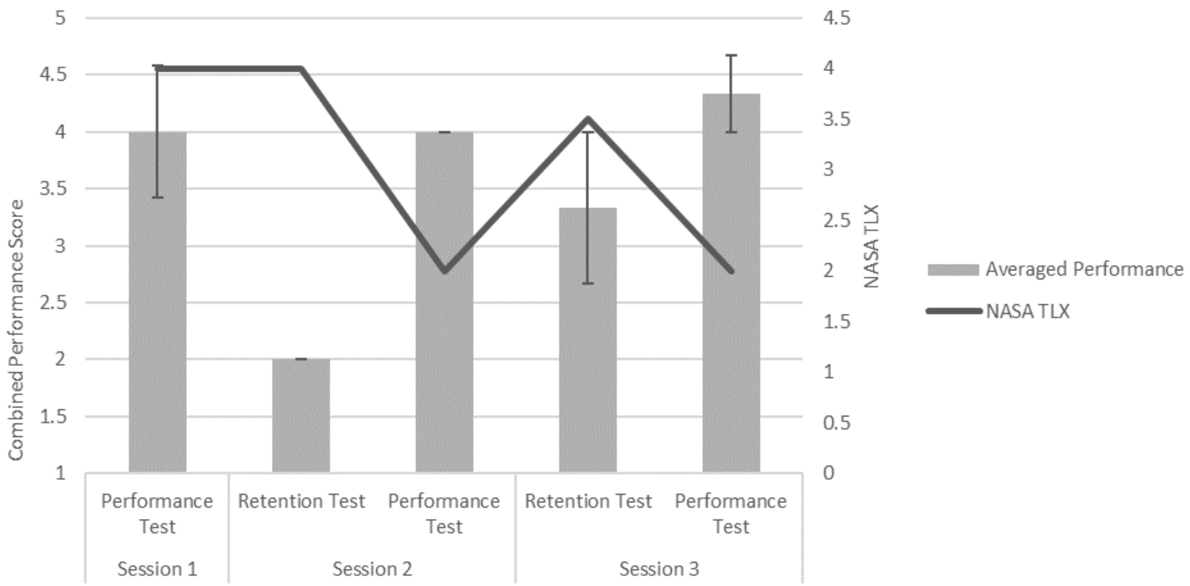

Figure 6. Individual performance (participant 1) with standard error compared to subjective workload (NASA TLX).

3.2.2 Cognitive Load Metric: EEG

In Figure 7, the performance and delta (i.e. Tactical Intercept minus Baseline) in relative band powers for individual (IAF-based) upper alpha and theta are plotted during the PTs and RTs for each session of one participant. When delta is positive that means the power in TI was higher compared to the baseline within the same run. When delta is negative that means that the power in TI was lower compared to the baseline within the same run. The results of Participant 1 are similar to the aggregated results.

For all runs, a positive delta in individual upper alpha band power and a negative delta in theta band power were expected as this would indicate increased cognitive load during the TI. Apart from theta band power in PT Session 3, the results confirm these hypotheses.

In addition, when comparing PT and RT more closely, a lower cognitive load for PTs than for RTs is expected, as training aids in both increasing performance and lowering the cognitive load to execute their run. Apart from the theta band power in PTs Session 1 and 2, this hypothesis can be confirmed. The delta in both band powers in PT of Session 2 is *smaller* when compared to its band power in RT within the next (and, but only for upper alpha, even the same and previous) session(s) which indicate(s) that cognitive load in PT is *decreasing* compared to RT. This could be induced by a constantly improving performance in combination with a lowering cognitive load in each consecutive run.

However, this doesn’t appear to apply for all PTs and RTs. For the cases where the hypothesis cannot be confirmed, PT resulted in an increased cognitive load compared to the RT *within* the same session. As observed performance is lower during the RT the decreased workload (according to NASA-TLX) might reveal the reduced motivation and effort of the pilots to push the envelope of the TI and eliminate the hostile aircraft when this appears to be out of limits. The PTs’ increased delta in both band powers could illustrate the endeavour to push extra for scenario fulfilment. Interestingly, the results on NASA-TLX suggest the opposite. Therefore, a mismatch between actual (objective) cognitive load and perceived (subjective) cognitive load is identified. We need to collect more data of more sessions and analyse different types of data (such as ECG) with the same (and new) participants to further research this phenomenon.

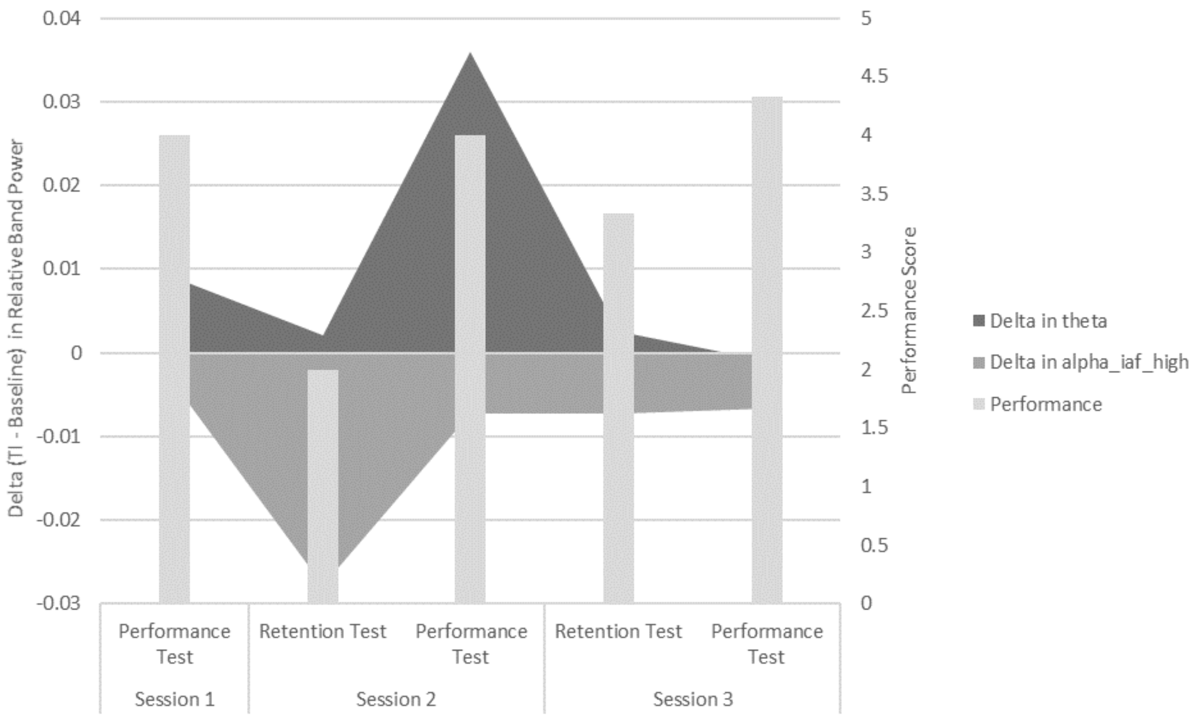

Figure 7. Average delta (TI minus Baseline) in relative band power of theta and individual upper alpha (i.e. alpha_iaf_high) compared to performance (Participant 1).

3.3 Lessons Learned

The EEG data are difficult to interpret due to intra- and inter-individual differences. In general, an increased cognitive load is identified compared to its baseline. Delta of the band power during later sessions increased compared to the first sessions which illustrate the endeavour to push extra for scenario fulfilment.

Overall (i.e. aggregated results), a small difference is observed in EEG between the RTs and PTs. Unlike expected, the cognitive load sometimes even suggest a more demanding PT compared to the RT. Therefore, a mismatch between actual (objective) cognitive load and perceived (subjective) cognitive load is identified. Pilot perform worse in RTs and find it more demanding (NASA TLX), however, their objective cognitive load appears to be often higher in the PT. Because of the nature of the scenario, this possibly suggest that the pilots are abandoning the run when they are aware that the hostile target is not within reach anymore.

Looking at the training sessions few improvements should be implemented to make the PT and RT more equal for all fighter pilots. It is tempting to make the starting conditions more predefined, such as releasing the pilot always at the same location, altitude, and speed. This was deliberately not done to prevent the pilot to become too acquainted with the scenario. A more diverse set up scenarios could avoid this, however, using numerous scenarios again can lead to differences in comparing. The balance of establishing equal conditions while preventing participants from becoming to familiarised will always remain a challenge to scenario design for retention training.

4.0 FUTURE WORK

Future work will focus more on using the retention performances and load measures to determine personalized retention schemes. Implementing personal schemes may raise scheduling difficulties in training organisations, which in turn require more advanced scheduling optimization methods [25] that need to be developed.

Furthermore, we plan to develop optimum load model which automatically and in real-time adapts the task complexity during a simulator session. More specific, complexity can be adapted by adjusting computer generated forces. For example, when the computer generated forces appear to be too challenging or too comforting, then the complexity of the computer generated forces will be adjusted accordingly in real-time. As a result of this adaptive automation, these computer generated forces will behave and act within the flow channel (see Figure 1) of the individual. Specific activities are needed to deal with intra- and inter-individual differences of known and new participants.

Also, the recorded heart rate (variability) data is to be analysed for the performed sessions, and will be used to substantiate the cognitive load data. This might reveal findings which explain the contradictory EEG and NASA TLX observations and give more insight in the intra- and inter-individual differences.

With advanced automated schedules and adaptive training, further support is required for both instructor and pilot. We plan to develop a dashboard for the instructor and pilot to enhance insight in training progress and plans. This dashboard should present the pilot's current and historical cognitive load based on (psycho)physiological data in an intuitive way so both (i.e. instructor and pilot) can use this information for their benefit. For example, the pilot (or trainee) could use this information to understand where he/she had a higher cognitive load during which events of a specific training.

On the short term, more sessions are planned which should deal with the low sample size and expose whether PBT is ready to become an unambiguously and interpretable training philosophy for broadly applicable situations.

5.0 CONCLUSIONS

Overall, a relation between performance, subjective workload, and (objective) cognitive load is identified. The TI sessions appeared to be a suitable training setting to identify these relations and individual differences.

Acknowledgements

This research has been partially funded by The Netherlands MoD, research program OiT (L1806, *Opleiden en Individueel Trainen*). The presented results are part of a series of publications. We thank the involved former F-16 pilots, fighter controller, and former F-16 weapon instructor who observed the experiment sessions for performance grading.

REFERENCES

[1] Wong PCM, Vuong LC, Liu K (2017) Personalized learning: From neurogenetics of behaviors to designing optimal language training. Neuropsychologia 98:192–200. https://doi.org/10.1016/j.neuropsychologia.2016.10.002

[2] Sweller J, van Merrienboer JJG, Paas FGWC (1998) Cognitive Architecture and Instructional

Design. Educational Psychology Review 10:251–296. https://doi.org/10.1023/A:1022193728205

[3] Sweller J, Ayres PL, Kalyuga S, Chandler P (2003) The expertise reversal effect. Educational Psychology Review 38:23–31

[4] Paas F, Tuovinen JE, van Merriënboer JJG, Aubteen Darabi A (2005) A motivational perspective on the relation between mental effort and performance: Optimizing learner involvement in instruction. ETR&D 53:25–34. https://doi.org/10.1007/BF02504795

[5] Hedegaard M (2005) The zone of proximal development as basis for instruction. In: An Introduction to Vygotsky, 2nd ed. Routledge

[6] Csikszentmihalyi M (2000) FLOW: The Psychology of Optimal Experience. 6

[7] Schell J (2014) The Art of Game Design, 0 ed. A K Peters/CRC Press

[8] Kiili K, de Freitas S, Arnab S, Lainema T (2012) The Design Principles for Flow Experience in Educational Games. Procedia Computer Science 15:78–91. https://doi.org/10.1016/j.procs.2012.10.060

[9] Colchester K, Hagras H, Alghazzawi D (2016) A Survey of Artificial Intelligence Techniques Employed for Adaptive Educational Systems within E-Learning Platforms. Journal of Artificial Intelligence and Soft Computing Research 7:47–64

[10] Haegens S, Cousijn H, Wallis G, et al (2014) Inter- and intra-individual variability in alpha peak frequency. NeuroImage 92:46–55. https://doi.org/10.1016/j.neuroimage.2014.01.049

[11] Käthner I, Wriessnegger SC, Müller-Putz GR, et al (2014) Effects of mental workload and fatigue on the P300, alpha and theta band power during operation of an ERP (P300) brain–computer interface. Biological Psychology 102:118–129. https://doi.org/10.1016/j.biopsycho.2014.07.014

[12] Puma S, Matton N, Paubel P-V, et al (2018) Using theta and alpha band power to assess cognitive workload in multitasking environments. International Journal of Psychophysiology 123:111–120. https://doi.org/10.1016/j.ijpsycho.2017.10.004

[13] Shaw JC (2003) The brain’s alpha rhythms and the mind. BV Elsevier Science

[14] Klimesch W, Sauseng P, Hanslmayr S (2007) EEG alpha oscillations: the inhibition–timing hypothesis. Brain research reviews 53:63–88

[15] Santiago-Espada Y, Myer RR, Latorella KA, Comstock JR The Multi-Attribute Task Battery II (MATB-II) Software for Human Performance and Workload Research: A User’s Guide. 269

[16] Klimesch W (2012) Alpha-band oscillations, attention, and controlled access to stored information. Trends in Cognitive Sciences 16:606–617. https://doi.org/10.1016/j.tics.2012.10.007

[17] Brookings JB, Wilson GF, Swain CR (1996) Psychophysiological responses to changes in workload during simulated air traffic control. Biological Psychology 42:361–377. https://doi.org/10.1016/0301-0511(95)05167-8

[18] Fairclough SH, Venables L (2006) Prediction of subjective states from psychophysiology: A multivariate approach. Biological Psychology 71:100–110. https://doi.org/10.1016/j.biopsycho.2005.03.007

[19] Gevins A, Smith ME (2003) Neurophysiological measures of cognitive workload during human-

computer interaction. Theoretical Issues in Ergonomics Science 4:113–131. https://doi.org/10.1080/14639220210159717

[20] Gevins A, Smith ME, Leong H, et al (1998) Monitoring Working Memory Load during Computer-Based Tasks with EEG Pattern Recognition Methods. Hum Factors 40:79–91. https://doi.org/10.1518/001872098779480578

[21] Fournier LR, Wilson GF, Swain CR (1999) Electrophysiological, behavioral, and subjective indexes of workload when performing multiple tasks: manipulations of task difficulty and training. International Journal of Psychophysiology 31:129–145. https://doi.org/10.1016/S0167-8760(98)00049-X

[22] Hart SG (2006) Nasa-Task Load Index (NASA-TLX); 20 Years Later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50:904–908. https://doi.org/10.1177/154193120605000909

[23] Corwin WH, Biferno MH (1989) Assessment of Crew Workload Measurement Methods, Techniques and Procedures. DOUGLAS AIRCRAFT CO LONG BEACH CA 1:237

[24] Maskeliunas R, Damasevicius R, Martisius I, Vasiljevas M (2016) Consumer-grade EEG devices: are they usable for control tasks? PeerJ 4:e1746

[25] Fjaerbu EL, Svendsen GK, Van der Pal J Feasibility of Performance-Based Training Programs for Combat Aircraft Pilots. Paper presented at the NATO MSG virtual symposium Towards On-Demand Personalized Training and Decision Support. STO-MSG-177-RSY

AUTHOR BIOGRAPHIES

Maykel van Miltenburg earned his Master’s degree in Applied Cognitive Psychology (MSc) in 2014 and is a Human Performance Specialist within NLR's department of Safety and Human Performance. His expertise lies on measurement tools such as rating scales and a combination of (psycho)physiological measurements: EEG, Heart Rate Variability, speech volume/pitch and eye-tracking.

Jur Crijnen graduated in Cognitive Sciences & Artificial Intelligence (MSc) and Mechanical Engineering (BSc). His passion for aviation is reflected in the research he does at NLR, which focus on Modelling & Simulation topics and future training applications.