UNIVERSITY OF TWENTE.

MAX-PLANCK-INSTITUT
FÜR BIOLOGISCHE KYBERNETIK

Mental Workload in a Moving Helicopter Simulation:

An Analysis using Near-InfraRed Spectroscopy in a Moving Environment

by

Joost Zijlstra

Supervised by Dr. K.N. de Winkel from Max Planck Institute for Biological Cybernetics and Dr. R.H.J. van der Lubbe and Dr. S.M. van den Berg from University of Twente.

Research at Max Planck Institute for Biological Cybernetics.
Submitted for the title of Master of Science to the programme *Human Factors and Engineering in Psychology* at the *University of Twente*

Enschede, 22 July 2019

Abstract

Near-Infrared Spectrosopy or NIRS shows promise as an alternative to mainstream neuroimaging techniques. As this technique is based on light instead of electromagnetic signals and does not require a person to lie still in a machine, such as with mainstream methods like EEG or PET, it can prove especially useful for situations where a person is exposed to motion from the environment, such as in a helicopter. To assess how well NIRS can measure neural correlates of mental workload (MWL), an EEG-cap with NIRS optodes was placed over different regions of the frontal cortex. A moving-base simulator was used to simulate the environment of a moving helicopter. Participants completed a simulated helicopter flying task in either clear weather for a reference state for MWL, or with wind and fog, for increased MWL. These MWL manipulations were performed both with a moving simulator and with the simulator's movement turned off, to catch potential differences. Interestingly, there were indications that NIRS was able to detect differences in neural activation during the task. While it was not verified that these differences were MWL-related, detecting these differences could be used to encourage research into the possibility of using NIRS as an 'online' measurement technique. The experiment did not find the expected neural correlates, related to the manipulation of MWL. Possibly, because these occurred in an area where no optodes were placed, or the task did not elicit enough change in MWL to measure with NIRS.

Keywords: Near-Infrared Spectroscopy, NIRS, Mental Workload, CMS

Table of contents

Introduction

High mental demands can have detrimental effects on operator performance, effects that can result in human error. This can in turn result in large losses, especially when human lives or expensive technology are at stake (Durantin, Gagnon, Tremblay & Dehais, 2014). An accurate assessment of mental workload can help to predict performance decline, associated with work overload or under stimulation. In turn, this prevents operator error and allows for pertinent intervention (Ayaz et al., 2012; Hirshfield et al., 2009). To make an accurate assessment, it is necessary to assess operator performance not only in laboratory settings, but in more naturalistic settings as well, where motion and mobility are present (Mehta & Parasuraman, 2013).

Several methods exist to measure mental workload, each with specific advantages and disadvantages. However, when measuring mental workload in a natural environment, many of the mainstream methods like EEG, PET, fMRI or self-report measures, can run into several difficulties. They can be disruptive during a task or sensitive to electromagnetic interference. Additionally, they require a person to be relatively motionless or lying down. Research of the last several years suggests that Near-InfraRed Spectroscopy (NIRS) could play a role in filling the position of a mobile neuroimaging technique. NIRS uses near-infrared light to measure local changes in hemoglobin concentrations in the cortex, which are correlated with neural activation. These changes are known as the hemodynamic response and can influence mental workload (Ayaz et al, 2012; Strangman, Culver, Thompson, & Boas, 2002; Tai and Chau, 2009). Using this technique, Ayaz and colleagues (2012) were already able to assess mental workload in a relatively noninvasive way during a task. This is interesting, as the technique employs light for its measurements. This makes it in theory less susceptible to factors like electromagnetic noise, muscle activation or movement of the participant. Because of this, NIRS is a promising method for more natural environments, where these factors play an important role in measuring mental workload.

Mental workload

The term mental workload (MWL) has proven to be surprisingly difficult to define. The “resource model” gives an intuitive explanation, with workload being the general term used to express the demands that tasks impose on an operator’s limited information processing resources (Wickens, Hollands, Banbry & Parasuraman, 2015). MWL varies with task demands and the capacity of the operator to meet these demands. Furthermore, Wickens (1984) proposed that a human operator does not have one undifferentiated pool of resources, but instead relies on multiple to meet task demands. This theory was named the multiple resource theory and states that multiple tasks at the same moment can draw on different resources, if these resources are not closely related. This is especially relevant for work environments that have high demands and require multitasking, such as for vehicle operators.

Problems with workload can occur when task demands exceed an operator's cognitive capacity. Performance degradation, attentional lapses, cognitive tunneling and errors can occur due to high workload. This phenomenon is called overload (Loft et al., 2007). Conversely, underload may happen when there is too little stimulation, potentially causing similar performance decrements. For optimal performance, MWL should therefore not be minimized per se, but kept in between the bounds of under- and overload (Ayaz et al., 2012; Young, Brookhuis, Wickens & Hancock, 2015). To achieve this, reliable and robust methods of assessment for MWL are needed. Many methods already exist. Much of the current research of NIRS focuses on measuring activation in the prefrontal cortex (Ayaz et al., 2012). To understand why NIRS can contribute to existing methods, it is important to understand the advantages and disadvantages of the current methods to assess MWL.

Methods to assess MWL

Cain's 2017 meta-analysis gives an elaborate overview of methodologies to assess MWL. He categorizes them into three broad groups: self-report measures, performance measures and physiological measurements. Self-report measures usually take the form of questionnaires. Examples are the NASA Task Load Index (NASA TLX or TLX), Subjective Workload Assessment Technique (SWAT) and the Rating Scale Mental Effort (RSME) (Cain, 2007). Performance measures take many forms, as these are dependent on a specific task and its performance goals. Performance measures include tests of speed, accuracy, or error rates of an operator during a task (Cain, 2007). The physiological measures include measurements of heart rate, electromagnetic brainwave and of the hemodynamic response in the brain. Techniques include Electro-Cardiogram (ECG) for heart rate, Electro-Encephalogram (EEG) for electromagnetic brainwave measurements and functional Magnetic Resonance Imaging (fMRI) and Positron Emission Tomography (PET) for hemodynamic response in the brain (Ayaz et al., 2012).

Although each of these techniques has specific advantages, there are also specific disadvantages. Self-report measures are complex to design, difficult to validate and hard to generalize. Performance measures do not necessarily relate directly to MWL (Cain, 2007) and are difficult to generalize, as specific tasks in different applications often require different measurement types. A general problem with the self-report and performance measures is that they only provide relatively indirect measures of MWL. Moreover, if administered during the task, both techniques can interfere with the MWL being measured. This problem may be avoided by administering them afterwards, but they are then less likely to provide insight into the user's changing experience during the task itself (Ayaz et al., 2012; Hill & Bohil, 2016; Hirshfield et al., 2009).

Physiological measurements may offer a solution as they can provide continuous, unobtrusive monitoring and can be taken 'passively', without placing an extra, active mental load on a participant (Ayaz et al., 2012; Wickens, Hollands, Banbry & Parasuraman, 2015). There are several methods to assess MWL through measurement of neural correlates. EEG has become a dominant technique in this respect, thanks to its relatively low cost, high temporal resolution and ease of use. Several markers have been found to reflect MWL in midline central and parietal areas and right hemisphere frontal and temporal areas (Roy, Charbonnier, Campagne & Bonnet, 2016; Wickens, Hollands, Banbry & Parasuraman, 2015). Because of its higher spatial precision than EEG, fMRI has been used to more precisely pinpoint activity associated with MWL in the prefrontal, parietal and anterior cingulate cortex (Jansma, Ramsey, Coppola & Kahn, 2000), which is supported by PET research (Petersen, Van Mier, Fiez and Raichle, 1998). However, the physiological measures discussed above also have several limitations. EEG is sensitive to electromagnetic interference, eye-movement artefacts, and provides only a limited spatial resolution of underlying cognitive processes. FMRI requires restriction of movement in individuals, and exposes subjects to loud noise, and PET also has the additional problem that it requires the use of potentially harmful radioactive tracers (Izzetoglu, 2008). These drawbacks can be especially prohibitive in naturalistic settings.

Previous research

Results of several recent studies suggest Near-InfraRed Spectroscopy may be used to overcome certain limitations of other neuroimaging methodologies in the measurement of MWL (Ayaz et al., 2012, Ayaz et al., 2010; Herff et al., 2014; Izzetoglu, Bunce, Izzetoglu, Onaral & Pourrezaei, 2007; Keshmiri, Sumioka, Yamazaki & Ishiguro, 2017; De Winkel, Nesti, Ayaz & Bülthoff, 2017; Young, Brookhuis, Wickens & Hancock, 2015). These studies typically focused on placing demands on working memory to vary MWL, and measured activation in the prefrontal cortex (PFC), which is associated with regulation of the working memory system. To vary MWL, they applied N-back tasks. In a typical N-back task, a participant is presented a sequence of items in the form of letters, patterns or numbers. Participants are asked to indicate if the current stimulus is the same as the stimulus presented 'N' trials before. This N usually ranges from one to four stimuli. When N is higher than two, high MWL is reported, and the task requires continuous mental effort. Application of these tasks is popular, because they can be easily manipulated and have a well-characterized paradigm with good correlations between level of difficulty and cortical activation (Ayaz et al., 2012; Wickens, Hollands, Banbry & Parasuraman, 2015). These N-back studies found that an increase in MWL correlates robustly with an increase in oxygenation in the dorsolateral PFC (Ayaz et al., 2012; Herff et al., 2014). Furthermore, Fishburn, Norr, Medvedev and Vaidya (2014) found that this activation scales linearly with MWL on N-back tasks. This suggests that NIRS can provide an indicator to measure changes in MWL, using N-back tests.

As N-back alone was not enough to simulate real world environments, research has also endeavored to test NIRS in more ecologically valid settings. For example, Ayaz et al. (2012) used a pc-based air traffic simulation with experienced air traffic controllers. Large changes in difficulty resulted in statistically significant hemodynamic changes in the prefrontal cortex, particularly in the region of the left dorsolateral PFC and anterior medial PFC, close to respectively AF7 and AFz from the international 10-5 system. However, NIRS had problems with differentiating between smaller changes in difficulty in this experiment. Gateau et al. (2015) made an online NIRS system to discriminate between on-task and not-on-task, and between high and low working memory load. They tested their system on pilots in a non-moving flight simulator and made a system that was able to recognize different workload states using the hemodynamic response in the left and right dorsolateral PFC. Also, results from Unni et al.

(2015), who used a virtual reality driving simulator in combination with a speed sign N-back task, indicate that measuring the hemodynamic response in the lateral PFC could reliably quantify MWL levels in more naturalistic tasks. Also, more recently De Winkel, Nesti, Ayaz & Bülthoff (2017) found encouraging results for measuring hemodynamic responses with NIRS in a moving environment, suggesting that NIRS could be used in more ecologically viable settings. In short, the aforementioned research indicates that NIRS could be a valuable tool to measure MWL in more naturalistic settings. However, until now participants performed their tasks while sitting in a fixed, relatively movement-free position. This is acceptable for operators in a static environment but does not inform about the robustness of NIRS-measurements for operators in moving environments.

One place that could benefit from NIRS's potential is the work environment of vehicle operators. Helicopter pilots are an excellent example for this: not only do they have to move their head and eyes to read instruments and get a bearing of their surroundings, but they also operate a vehicle that can move freely and undergoes strong accelerations in many different directions. NIRS specifically, has a lot of potential here, as modalities such as EEG would encounter considerable electromagnetic interference in these areas. Mistakes due to under- or overload for operators in this area can become quite costly, not only resulting in the loss expensive equipment, but even in the loss of human life. Accurate online MWL assessment in the form of NIRS could prove a valuable tool. However, before adopting NIRS as a neuroimaging tool for human operators in moving environments, it is necessary to assess its robustness in such extreme conditions.

Near-InfraRed Spectroscopy (NIRS)

NIRS is a method that utilizes near-infrared light to measure cortical activity by measuring the concentration of hemoglobin in a certain area. Hemoglobin is used to transport oxygen in the blood. Hemoglobin with an oxygen molecule bound to it is called oxygenated hemoglobin (HbO2). When the oxygen is needed, it can be taken from the hemoglobin, resulting in deoxygenated hemoglobin (HbR, standing for 'reduced hemoglobin'). When neurons in the cortex are active, they use up oxygen from local hemoglobin, which results in a change of the concentration of HbO2 and HbR in a typical order. These typical changes to concentrations are known as the hemodynamic response and can be related to external stimuli. The hemodynamic response has practical use when trying to determine MWL, since measuring the hemodynamic response can provide insights into neural activity, related to MWL (Ayaz et al., 2012; Strangman, Culver, Thompson, & Boas, 2002; Tai and Chau, 2009). Hemodynamic responses typically appear in the following order: after the onset of neural activity in a certain area of the brain, there is a slight increase in the concentration of HbR in that part, due to the oxygen being used in the activation. Subsequently, a larger, delayed increase in HbO2 follows, peaking at around 10 seconds after activation, together with a corresponding decreased concentration of HbR (Tai and Chau, 2009). After the end of activation HbO2 levels decrease back to baseline and HbR levels increase back to baseline (Herff et al., 2014). The positive relation between HbO2 level and neural activation is most commonly used to measure neural activation. When cardiorespiratory changes are expected as part of a fatiguing task, the positive relationship between HbR level and neural activation might be more reliable (Mehta & Parasuraman, 2013). A lot of seminal research has included both in the analysis. To not miss responses in a relatively new experiment and also take into account possible physical fatigue from the experiment, both levels will also be included in the analysis here (Ayaz et al., 2012, Ayaz et al., 2010; Herff et al., 2014).

NIRS can measure these hemodynamic responses in target regions of the cortex. It uses the optical window of human tissue: photons in the spectrum of around 700 to 900 nm can penetrate human tissue relatively easily and are then either scattered by different layers of tissue within the head or are absorbed mainly by water and hemoglobin (Izzetoglu, 2008). Because the absorption rates of water, HbO2 and HbR differ substantially from each other in this window, it is possible to calculate the relative concentrations of these molecules after measurement (Izzetoglu, Bunce, Izzetoglu, Onaral & Pourrezaei, 2007; Tai & Chau, 2009). Consequently, the hemodynamic response over a certain region of interest can be measured by placing a grid of near-infrared light sources and light detectors over the part of the scalp covering this region.

NIRS offers several advantages for specific situations over other neuroimaging modalities, making it viable for more ecologically valid environments: it is not susceptible to electromagnetic interference or eye-movement artefacts and subjects can be allowed considerable freedom of movement. Furthermore, the equipment is relatively inexpensive, commercially available, is mobile, has no harmful materials, is non-invasive, and works soundlessly. NIRS provides spatially well-resolved information, and when compared to EEG has a better signal-to-noise ratio (Aqil, Hong, Jeong & Ge, 2012). Consequently, the advantages of

NIRS suggest it may be used to obtain reliable measurements of MWL in a wide variety of settings, provided that several drawbacks are taken into account. As NIRS measurements are based on blood-oxygen-level dependent imaging, measurements have a relatively low temporal resolution, requiring several seconds before neural activation can be measured (Mehta & Parasuraman, 2013). Furthermore, the near-infrared light has a penetration depth of around three centimeters, meaning that the measurements cannot go deeper into the head than cortex level. While these drawbacks can limit the usefulness of NIRS in settings that require high temporal resolution and measurements below the cortex level, the possible advantages suggest that NIRS can fill an important role for neuroimaging in settings with a lot of motion.

Present study

We hypothesized that different levels of MWL could be detected in the PFC in a moving environment with NIRS. Using a simulated helicopter in a moving-base motion simulator, we assessed this hypothesis. The level of MWL was manipulated by influencing task difficulty through different weather conditions during flight. As a helicopter can move in any direction, this provided a broad range of potential motions during testing.

There were two manipulations for every participant. The first manipulation was giving a participant two levels of task difficulty to manipulate MWL. This featured a clear weather condition, aimed at eliciting a lower level of MWL and a foggy, turbulent weather condition, aimed at eliciting a higher level of MWL. The second manipulation was a binary manipulation on motion of the environment, with motion of the simulator either being turned on or off during the task. The motion manipulation could give insight in if the current setup was successful in manipulating MWL and measuring it with NIRS without motion. Then, this then gave a comparison to the measurements when the motion of the simulator was turned on.

Participants had NIRS optodes placed on their forehead, measuring the hemodynamic response in the PFC. The expectation was that, irrespective of the motion manipulation, at a higher level of MWL NIRS would detect relatively higher concentrations of HbO2 and lower concentrations of HbR, relative to the lower level of MWL, in which NIRS was expected to detect relatively lower concentrations of HbO2 and higher concentrations of HbR. These changes were expected to be detected around the dorsolateral PFC, like in a lot of N-back based research with MWL. Most specifically in an area close to AF7 from the international 10-5 system as found by Ayaz et al. in 2012 (Gateau et al., 2015; De Winkel, Nesti, Ayaz & Bülthoff, 2017). The objective of the study was to manipulate mental workload for the participant and see if NIRS detected these manipulations in a moving environment with a more naturalistic setting.

Methods

Participants

A total of 28 participants were recruited for the experiment. Out of these, 21 of the participants had no previous experience with helicopter controls, as this was not necessary to complete the experiment. The other seven had some experience with the controls from prior, unrelated experiments. Of the participants, fifteen were male and thirteen were female. The ages of the participants ranged from 18 to 38 years old, with a mean of 25 years (SD = 3.9 years).

Due to safety regulations for the motion simulator that was used in the experiment, several exclusion criteria were put in place. Persons could only participate if they were shorter than 1.95m, weighed under 100kg, were not pregnant and were under 45 years of age. Persons with a (history of) vestibular illness, spinal problems, heart or circulatory disease, or with a pacemaker were excluded. Participants were offered a financial compensation for their time.

This experiment adhered to the declaration of Helsinki. The experimental protocol was approved by the ethical committee of the medical faculty of the Eberhard-Karls University of Tübingen, Germany, with the reference number 238/2017BO1. Participants were informed on their rights and the risks of the motion simulator before the experiment. Before starting the experiment, they were required to sign an informed consent form. At the start of the experiment participants were asked about their age, sex, handedness and if their vision was normal, corrected or uncorrected.

Design

The experiment consisted of a two by two within-subjects design. Two variables were manipulated: task difficulty and motion of a motion simulator (see *CyberMotion Simulator*). The first variable, task difficulty, was aimed at influencing MWL. This was manipulated through a helicopter flying task as described in *Helicopter simulation*. Task difficulty was binary, where clear weather would simulate less demanding flying conditions and turbulence and fog would simulate more demanding flying conditions. The second variable, motion, was aimed at providing NIRS measurements with and without motion, for comparison of the MWL neuroimaging measurements. Each of the two by two manipulations was given their own block, giving 2x2 = 4 blocks. Each of the four blocks consisted of five repetitions of the flying task, resulting in the participants performing the task a total of 4x5=20 times.

Task description

The participant was seated in a motion simulator with the NIRS optodes on the head. After this, the simulator started. The simulation consisted of a small airfield with floating markers demarcating the ideal path for the helicopter to take. The task consisted of a path with five subsequent maneuvers and was laid out in such a way that movements in heave, sway, surge, yaw and hover were all present at least once during the task. The experiment consisted of four blocks, each with one of the four experimental manipulations. A participant repeated the task a total of twenty times: five times per block, with around 30 second breaks in between each repetition and a longer ten-minute break after completing a block.

At the start of each task repetition, the helicopter was suspended in the air at a fixed starting position, facing a floating marker globe, with a marker line going downwards out from the bottom. Participants could begin the task by pressing a specific button on the helicopter controls. The helicopter would then begin hovering in the starting position. The participants had been asked to perform five maneuvers when the simulation started. After finishing one of the five maneuvers of the task the participant would press a button, signaling the end of the maneuver to Simulink and the NIRSport box. The five maneuvers of the task were:

1. Fly the helicopter downwards, facing the black line that stuck out the bottom of the first marker globe, until reaching the second marker globe.
2. From the second marker globe, participants were asked to fly sideways to the right, without any yaw rotation. To help them, they could see a row of evenly spaced, black vertical lines, jutting from the ground directly in front of them. These continued to the right. The tops of these lines marked the intended altitude. and direction during this maneuver. When this row of lines ended, the maneuver ended, the maneuver ended.

3. In front of the participant was a path, lined by black vertical lines on the left- and right-hand side. They were asked to fly forward, in between these lines until the path ended over a yellow and green square on the ground.
4. Over this yellow and green square, they had to stabilize the helicopter, line up their nose with a yellow and green target square and turn the nose 90 degrees to the right until they saw another yellow and green target square.
5. Finally, the participant would successfully complete the task by lining up their nose with another yellow and green target, and hover for 10 seconds over the yellow and green square on the ground.

After completing these maneuvers, the task repetition was over and the participants were moved back to the initial position, where they could then start again. A video of a helicopter task through the eyes of a participant can be found on this YouTube video.

Procedure

The total procedure took on average two and a half hours, of which the first hour consisted of instructions, practice and the preparation of the NIRS equipment on the head and then one and a half hour of the experiment. At the start of the first hour, participants were instructed on their rights, what the experiment entailed and risks associated with the motion simulator. They were then required to sign an informed consent form for their rights and an informed consent form on the risks. After this, they were given instructions on how to successfully complete the helicopter flying task. To stimulate optimal task engagement, participants were told it was important to perform the task to the best of their abilities. After safety instructions, they were seated in the CyberMotion Simulator and practiced the task without the movement of the simulator. Participants had a few practice runs, with the first half in the clear weather conditions and the second in the turbulent weather conditions. During this practice phase the experimenter provided additional instructions when necessary. When the participant could reliably finish the helicopter task, the experimenter verbally confirmed with the participant if the task was clear. After this, the optodes were placed on the participant's forehead and the experiment began.

The experiment consisted of four blocks, each block representing one of the four experimental conditions of the two by two design. Every block consisted of five tasks, to be completed by the participant. The order in which the blocks were presented to the participants was randomized and different per participant. After every task, the participant had a short break of around 30 seconds while the simulator was reset. During this short break, four scores were presented on-screen, representing their maximum deviation in meters from a perfect performance, in altitude, latitude, longitude and final heading. Participants could then use this score as feedback on their performance. After completing a block, participants were given a ten-minute break from the helicopter tasks. During this break, they could rest and fill out the TLX.

When a participant lost control of the helicopter and could not regain it, the task was stopped and had to be redone. This was done to make sure that possible task disengagement was not measured and to prevent possible motion sickness, stemming from uncontrolled movements in the simulation. When a participant had finished five tasks in each of the four conditions and had filled out the last TLX, they were taken out of the CMS, given a full explanation of the goal of the experiment and were debriefed. The debriefing would conclude the experiment.

Apparatus

Software

For the experiment, the participant had to complete a helicopter task in a simulated helicopter. This helicopter simulation was run from Simulink (The MathWorks, Inc., Natick, Massachusetts, United States). Unity version 4.2 was used for the visuals and provided a map to fly in, using a small airfield in the middle of mountain scenery as a setting. For the helicopter, a flight dynamics model for a Personal Aerial Vehicle was used, as described in the article by Perfect, Jump and White (2015). The idea behind this Personal Aerial Vehicle was to simulate a small helicopter of about 500kg, envisaged to require about the same skill level as is required for driving a car. It provided the participant with a small, stabilized helicopter that was easier to fly than a regular helicopter. This model was used to simulate a helicopter flight and enabled the experiment to be performed by participants with no previous flight experience.

Both turbulence and visibility could be manipulated in the simulation. Turbulence was created with the CETI method, as described in Perfect, Jump and White's article (2015). This method does not simulate full aerodynamic models for turbulence, but rather produces equivalent white noise in the control inputs. This white noise is then filtered according to existing models of helicopter turbulence, to produce the same effects as winds of around 40 knots (for filtering technique see Lusardi, 2004). Exponential squared fog mode was used to impair visibility for the higher task difficulty condition and could be switched on and off dynamically.

Simulator

The CyberMotion Simulator (CMS), shown in Figure 1, was used to simulate the helicopter flying task as described in the task description below. The CMS was developed at the Max Planck Institute for Biological Cybernetics, as a motion platform to research human control behavior. The setup for the CMS consisted of an enclosed cabin and an arm. These were positioned on a linear track of about 10m long. The arm provided motion in six degrees of freedom using seven axes and could also move along the track. The cabin was equipped with a curved projection screen in front of the participant. Two projectors on either side of the occupant provided a visual feed of the helicopter simulation on the projection screen (for more details see Nieuwenhuizen, 2013).

The participant was seated in a chair in the cabin and secured in place with a five-point safety harness. To control the simulated helicopter, the participant provided input using a cyclic stick for controlling tilt, a collective lever for controlling vertical movement and foot pedals for controlling yaw (developed by Wittenstein GmbH, Germany). Controls are also shown in Figure 3. Experimenter and participant could communicate with each other at any time through a headset. One video camera in the cabin was aimed at the participant, enabling constant monitoring of the participant. Another video camera provided an over-the-shoulder view of what the participant was seeing. Motion of the CMS could be switched off and on according to which experimental condition the participant was in. Motion provided the participant with inertial feedback of the helicopter's displacements in the simulation.

For protection of the relatively inexperienced participants, gain for motion of the cabin compared to motion of the simulated helicopter was set to 0.1 for longitudinal, lateral and vertical movement and yaw rotation. This meant the strength of the machine's motions reflected one tenth of the strength motions of the simulated helicopter. Gain for roll was put to 0.5 and pitch to 0.4. The participant was free to move their head during the experiment.

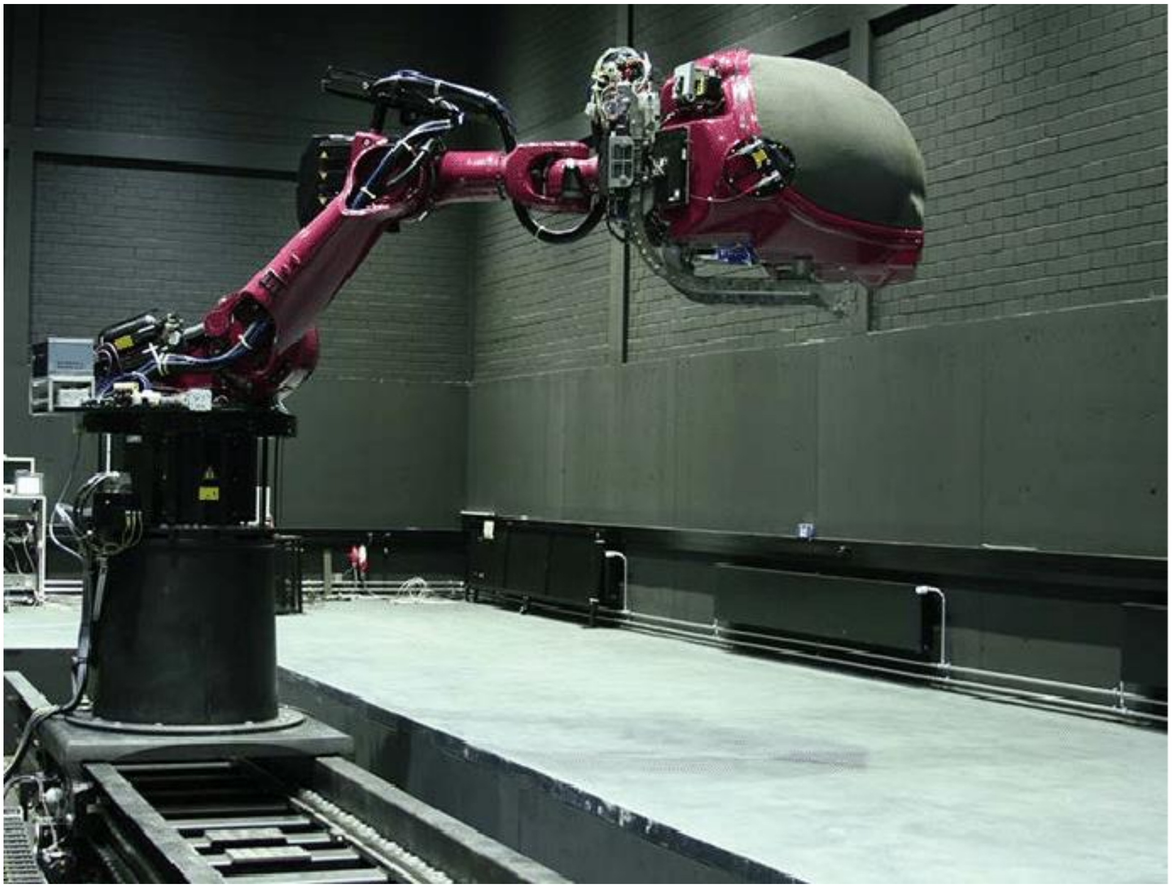


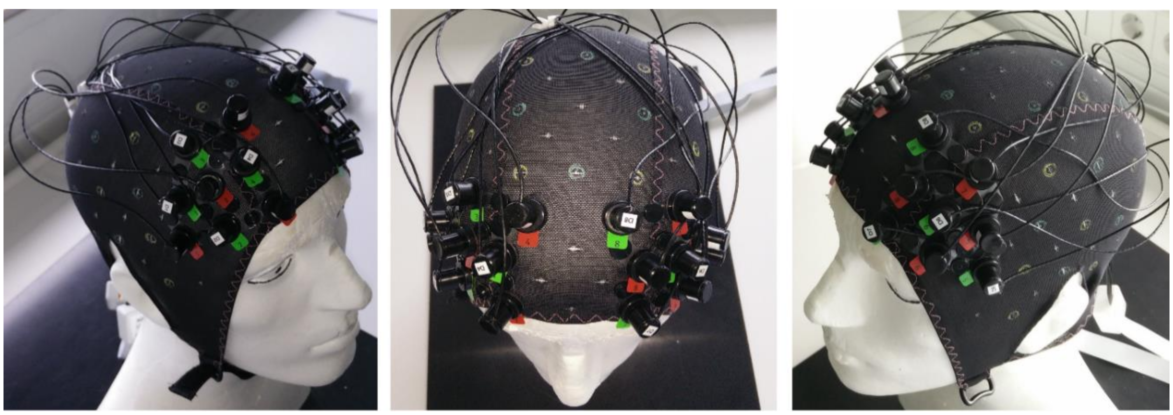
Figure 1: The CyberMotion Simulator (Venrooij et al., 2015). To see the simulator in action during this experiment go to this video.

NIRS

The Brain Products NIRSport Model 88 mobile imaging system was used to measure the hemodynamic response, using continuous-wave near-infrared light, on wavelengths of 760 and 850nm at a rate of 7.8125 Hz (developed by Brain Products GmbH, Gilching, Germany). By placing sources and detectors 30-35mm apart, sufficient penetration depth and signal strength could be achieved (De Winkel, Nesti, Ayaz and Bülthoff, 2017).

An EEG cap with the international 10-5 system was used to position the light sources and detectors on the scalp. Positions AF3, AF4, F5 and F6 on the international 10-5 EEG electrode placement system have been confirmed as relatively accurate localizations of Brodmann areas 9 and 46, which are attributed to the dorsolateral PFC structure, that is ascribed to being correlated to mental workload in a realistic control task (Ayaz et al, 2012; Cohen et al., 1997; Herff et al., 2014; Koessler et al., 2009). Sources and detectors were placed on and around these positions of interest to measure the hemodynamic response.

Photons from the sources travel back in a curved path into detectors placed a few centimeters apart. This path between source and detector was called an optode channel. In total there were 21 of these optode channels. In order to classify these different channels, they were named after their respective source-detector pair. The first optode was the source on F8 sending photons to detector F6. This channel was called ‘F8-F6’. This was done for the other optode channels, to finally get the following 21 optode channels: 'F8-F6', 'F8-AF8', ‘AFF6h-F6', 'AFF6h-AF8', 'AFF6h-F4', 'AFF6h-AF4', 'Fp2-AF8', 'FP2-AF4', 'AFF2h-F4', 'AFF2h-AF4', 'AFF2h-AFF1h', 'F5-F7', 'F5-AFF5h', 'AF7-F7', 'AF7-AFF5h', 'AF7-Fp1', 'F3-AFF5h', 'F3-AFF1h', 'AF3-AFF5h', 'AF3-AFF1h' and 'AF3-Fp1' . This setup gave a spatial resolution of around three centimeters on each prefrontal hemisphere. Pictures can be found in Figure 2, with red labels on the light sources and green labels on the detectors.



A. Right side of head B. Top side of head C. Left side of head

Figure 2: Setup of NIRS optodes on the head of a kind ‘volunteer’. A NIRS cap with the international 10-5 system was placed on a participant’s head. Red labels show the light sources, green labels the light detectors. Numbers on the labels made sure that the experimenter placed the sources and detectors in their matching slots. The sources and detectors were held in place by slots, placed in the cap beforehand. The cables from the optodes were connected to the NIRSport box. The cap, labels, slots and optodes were all provided by Brain Products GmbH, Gilching, Germany

Subjective mental workload

As criterion variable, the NASA Task Load Index (TLX) was used. The TLX was chosen as it is known to be relatively easy to use and has been acknowledged as a valid and reliable measure of MWL (Hart and Staveland, 1988; Hart, 2006; Rubio, Díaz, Martín and Puente, 2004). The widespread use also renders the results more generalizable to other research on MWL. The questionnaire was taken in online form directly after each experimental condition. The online version of the TLX that was used was designed by Vertanen (2017) and can be found on http://keithv.com/software/nasatlx/nasatlx.html. Grier’s 2015 meta-analysis of the NASA TLX used in other experiments can then be used to interpret how high the participants’ subjective MWL was in relation to other experiments.

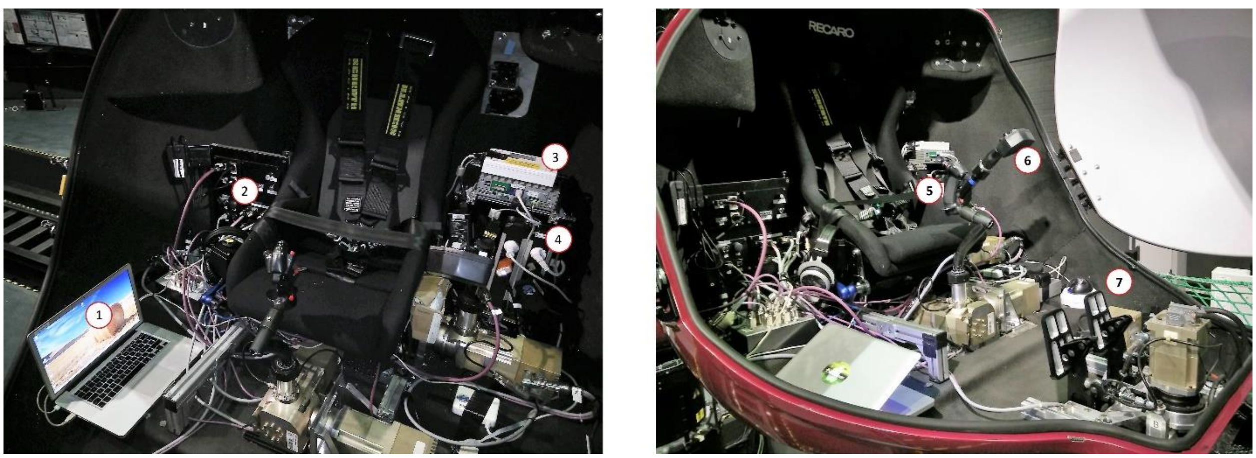


Figure 3: Setup inside the CMS cabin. For an experiment, the participant was sat in the chair in the middle of the cabin, with setup of Figure 2: ‘NIRS setup on head’, connected to the NIRSport box. The numbers in the image on the left mark materials associated with data gathering. The numbers in the image to the right mark the Wittenstein controls. These are as follows. 1: Laptop for storing experimental data. 2: CAN outlets. 3: NIRSport box. 4: Power outlets for setup inside cabin. 5: Cyclic stick. 6: Collective lever. 7: Foot pedals

Data collection setup

Figure 3 provides an overview of the data collection setup in the cabin. The NIRSport box was fastened inside the CMS cabin. This box was connected via cables to the NIRS optodes on the participant’s head. This box also received time synchronization signals with a cable from a CAN network. Three different values were communicated through this network. Value 1 signaled that a participant had started a flying task, value 2 occurred whenever the participant advanced to the next maneuver in the task and value 3 signaled that a participant had ended the task. These signals were used to couple events in the experiment with the NIRS measurements. All data collected in the NIRSport box was then sent via USB-cable to a laptop that was also secured in the cabin. The laptop was folded during experimentation, so the participant could not see what was on the screen. This laptop stored the data. Both the NIRSport box and the laptop were connected to power outlets in the cabin.

Exclusion of data

Technical problems in the form of a blown fuse and a connection failure between the measuring equipment, resulted in incomplete data for five participants. An unexpected personal situation and motion sickness also resulted in incomplete experiments for another two participants. All data from all these participants were excluded from analysis, leaving data from twenty-one participants.

Data preprocessing

During the experiment, the NIRS setup collected several data outputs. Firstly, two files were stored consisting of the recorded raw voltage readings. One contained the output for wavelength one (760nm), the other for wavelength two (850nm). Secondly, every time a task or maneuver started or ended a time synchronization marker was saved.

Data was preprocessed with MATLAB R2016a (The MathWorks, Inc., Natick, Massachusetts, United States). The first step in the preprocessing was to discard six unfinished flying tasks from four participants. These were stopped and started over, due to the participant losing control of the helicopter or hitting the ground, thus crashing. These tasks did not go as intended and were deemed more likely to measure overload-related task disengagement than task-related mental workload.

From the finished tasks, the raw voltage readings from the two wavelengths were converted to the associated changes in concentrations of oxygenated and deoxygenated hemoglobin in millimoles per liter. A third order bandpass filter was then applied with cut-off frequencies of 0.01 and 0.3 Hz, to remove artifacts from high frequency noise, heart cycle and respiration, similar to the filter used in the research of De Winkel, Ayaz and Bülthoff (2017). Timestamps were used to determine the start and end of each helicopter flying task. This was used to split the data from when the participant was and was not performing the helicopter flying task. To determine the hemodynamic response related to the experimental manipulations, it was necessary to assign a baseline measurement for the hemoglobin. To minimize confounding factors between the baseline measurement and the measurements during the experimental manipulations, baseline measurements took place in the experimental environment, when the participant was already seated in the CMS and not performing the task. Participants had periods of around 30 seconds before the task (re-)started, during these periods baseline measurements for HbO2 and HbR were made for the subsequent task, using the median concentration of each as baseline. After establishing the two baselines and performing the experiment, the hemodynamic response was calculated for both HbO2 and HbR respectively as: median concentration during a maneuver minus baseline. This gave a total of 5 maneuvers x 5 task repetitions x 4 task repetition blocks = 100 medians for the HbO2 and equally 100 medians for HbR per participant.

Forming a baseline in the CMS itself ensured minimal environmental changes between baseline and response during the task. Using medians instead of means made the measurements less sensitive to sudden noise-related spikes.

According to the American Red Cross, standard low and high hemoglobin concentrations range between 8.4 and 10.9 mmol/l (The American National Red Cross, 2019). This gives a difference of 2.5 mmol/l between general hemoglobin levels. Taking this as a broad criterion, measurements that showed a hemodynamic response larger than 3 mmol/l for HbO2 or larger than 1.5 mmol/l for HbR were unusually large and associated with noise. Furthermore, a visual inspection was performed to remove channels with bad or no reception. This happened when, for example, part of an optode lost contact with the skin or a dark hair moved in front of the optode. Visual inspections backed up the aforementioned criterions, with the unusually large measurements being associated with a lot of noise. As measurements for both HbO2 and HbR were made with the same channel, such measurements were removed for both HbO2 and HbR. This resulted in removal of 24% of the total measurements. Filters did not filter out significantly more in the trials with motion (25%), versus the trials without motion (23%), suggesting that this was not the main contribution of noisy data.

Data analysis

Data analysis was performed with IBM SPSS Statistics 25 (IBM Corp., Armonk, NY, United States, 2017). The data analysis first checked the NASA TLX scores, to give information about MWL between experimental conditions. With NASA TLX as a criterion for MWL, the main analysis followed. This analysis checked for noticeable differences in HR around area AF7, related to changes in MWL. A contrast was made here between conditions with and without motion. Finally, an explorative analysis was made. This checked the other measuring locations for possible MWL-related hemodynamic responses.

NASA TLX

First, ANOVA assumptions for the NASA TLX were checked. If the data passed the checks, a linear regression was performed with the *TLX score* as dependent variable and *weather*, *motion* and the interaction between *weather and motion* as independent variables.

Neuroimaging data around AF7

After analysis on the NASA TLX scores, the main neuroimaging data for the channels 'AF7-F7', 'AF7-AFF5h', 'AF7-Fp1' was checked for ANOVA assumptions. This was done for both the HbR and HbO2 responses individually. If data was not clustered, a linear regression was performed. If data was clustered around participants, a nested ANOVA design was chosen to fit that hemodynamic response with participant as a random effect.

Two main analyses took place. One with the median hemodynamic response in HbR as the dependent variable and one with HbO2 as dependent variable. Five predictors were added to this analysis. The *motion manipulation* of the motion simulator and the *MWL manipulation* (the weather effects) were fitted to test their effects on hemodynamic response measurements. *Order of the task* was fitted to test for potential learning or fatigue effects. *Maneuver* was fitted to see how hemodynamic response developed between different subtasks during the flying task. This also showed possible potential for differentiating between them, in more online situations in future. Lastly, *channel* was fitted to the model.

Finally, this regression analysis was again done, first with only non-motion measurements and second with only motion measurements, to see if there was any contrast in results. The same variables were fitted to these models, minus the *motion* manipulation.

Exploration of the other neuroimaging channels

As this analysis and experiment was relatively new, the measurements did not focus solely on the three AF7 channels, so as not to miss other potential MWL related hemodynamic responses. The eighteen other channels were tested for ANOVA assumptions, for both HbR and HbO2 individually. These were then again fitted into a nested ANOVA design if data was clustered around participants, with a random effect for participant. If data was not clustered, a linear regression was performed. The same five predictors were fitted as in the analysis around area AF7. However, this time, the interaction between *channel and the MWL-manipulation* was also added, to account for the big differences between the measuring locations.

There were a lot of channels in the interaction between channel and the MWL manipulation of this explorative analysis, resulting in many concurrent analyses. This brought the danger of the type 1 problem. To counteract this problem, a Holm-Bonferroni correction was performed on these eighteen channels. The method is similar to the Bonferroni correction, but has a higher power (Holm, 1979). The method consists of multiple ranked tests. There are as many ranks as there are null hypothesis tests, in this case: eighteen. The first rank starts with the hypothesis test with the lowest p-value. The p-value is tested with the significance criterion divided by the product of the total number of null hypothesis tests, minus previous ranks, 'k'. In formula form this is gives a new significance criterion in the form of: $\alpha/(n-k)$. So with a criterion of $\alpha = 0.05$, for the first rank this is: $0.05/(18-0) \approx 0.0083$, the second $0.05/(17-1) = 0.01$ and so on. When a null hypothesis is rejected the method advances a rank. The method stops advancing after all null hypotheses are rejected or when a null hypothesis is accepted. When this happens, all later null hypotheses are then also accepted. Data will be presented after Holm-Bonferroni correction has been applied. So, the p-value will be corrected with the following formula: corrected p-value = p-value *(n-k).

Results

First the results of the NASA TLX will be compared between experimental conditions. This gives a criterion measure on how MWL was correlated with the experimental manipulations. Then the main analysis of the area around AF7 will be presented. Finally, the explorative analysis of the other measuring areas will be presented. Four words will be used to give a short and consistent way of referencing the experimental manipulations. For MWL, the foggy, windy conditions meant to increase MWL will be called the 'increased MWL' conditions. The clear weather conditions meant to decrease MWL will be called the 'decreased MWL' conditions. When referencing the motion of the simulator, we will reference to the conditions where the motion was turned on as 'motion' conditions and the conditions with it turned off as 'motionless' conditions.

NASA TLX

The Shapiro-Wilk test and Q-Q plots showed normality for the increased MWL, motionless condition, $p = 0.60$, the increased MWL, motion condition, $p = 0.71$ and the decreased MWL, motionless condition, $p = 0.91$, but not for the decreased MWL, motionless condition, $p = 0.04$. As this violation of normality was relatively small, a linear model was still applied, as these can be relatively robust against smaller violations.

Regression showed that the increased MWL conditions scored on average 7.2 higher on the subjective mental workload scale than the decreased MWL conditions. Motion was not associated with changes in subjective MWL. Estimates showed that the decreased MWL conditions scored around 61.1 on the NASA TLX and the increased MWL conditions scored around 69.8. See table 1 for more details on the results of the regression.

Table 1: Results of the multiple linear regression analysis on the NASA TLX score for subjective MWL. These show a significant increase in NASA TLX scores, when the increased MWL conditions were introduced, versus the decreased MWL conditions. Simulator motion did not influence subjective MWL.

		95% Confidence Interval			
	B	Lower Bound	Upper bound	t	*p-value*
Intercept	61.1	56.4	65.7	26.111	**<0.005****
Increased MWL	8.7	3.44	14.0	10.512	**0.002****
Motion	1.2	-4.2	6.4	0.171	0.680

Neuroimaging around AF7

The means for both HbR (0.019 mmol/l) and HbO2 (0.035 mmol/l) HR's were not very pronounced and both positive. The Wald test showed clustering around participants, for both HbR concentrations, with $Z = 2.730$, $p = 0.006$, and for the HbO2 concentrations, with $Z = 2.723$, $p = 0.006$. For this reason, a mixed effects ANOVA was performed for both concentrations, with random effects for each participant. The mixed effects ANOVA's found no main effects for weather or motion. Order of the trials also did not influence HR's, suggesting no time related effects as the task went on. Maneuvers, however, did have an influence on the HR's, suggesting either a difference between these maneuvers, or a non-linear time-related effect. See Figure 4, for a better idea on the estimated HR's between maneuvers. See table 2 and 3 for more details on the ANOVA models.

If we did the analysis with only the motionless trials, no significant results were found for MWL manipulation either. HbO2 gave $p = 0.205$, with increased MWL having an estimated effect on HR of $B = 0.035$, std. error = 0.028. HbR gave a: $p = 0.202$, estimated effect $B = -0.019$, with a std. error = 0.014. With the motion trials, this was also not significant. Increased MWL had an estimated effect on the mean HbO2 with: $B = 0.026$, std. error = 0.028, $p = 0.341$. For HbR, this estimated effect was: $B = 0.014$, std. error = 0.015, $p = 0.330$.

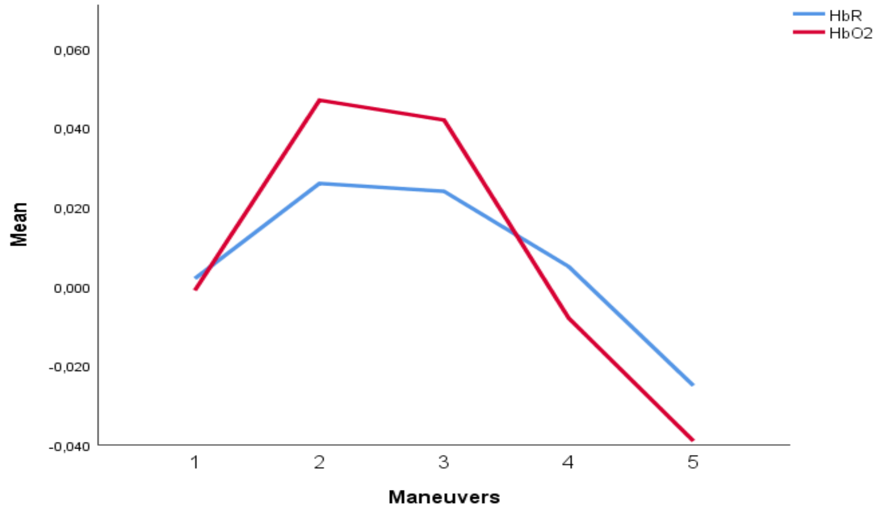


Figure 4: Figure showing mean HR for both HbR and HbO2, as found by the mixed ANOVA's. The second and third maneuver both showed significant increase in mean concentrations. Then a quick drop-off follows for HbO2, with a slower drop-off for HbR.

Table 2: Table showing the results of the mixed effects ANOVA for HbR around area AF7. These results show no effects from the MWL manipulation on the HR. However, the second and third maneuver show a significant increase in HbR, when compared to the other maneuvers.

		95% Confidence Interval			
	B Mmol/l	**Lower Bound Mmol/l**	**Upper bound Mmol/l**	**t**	***p-value***
Intercept	-0.025	-0.067	0.016	-1.227	0.222
Increased MWL	-0.002	-0.018	0.022	0.231	0.817
Motion	-0.004	-0.016	0.024	0.424	0.672
Maneuvers: 1st	0.027	-0.004	0.059	1.694	0.090
2nd	0.051	0.019	0.082	3.175	**0.002****
3rd	0.049	0.017	0.080	3.069	**0.002****
4th	0.030	-0.001	0.061	1.874	0.061
5th	0	0	0		
Trial Order: 1st	0.017	-0.014	0.048	1.066	0.287
2nd	0.001	-0.030	0.041	0.049	0.961
3rd	0.015	-0.015	0.047	0.966	0.334
4th	-0.015	-0.046	0.016	-0.932	0.352
5th	0	0	0		

Table 3: Table showing the results of the mixed ANOVA for HbO2 around area AF7. Again, the HR in the second and third maneuver are significantly higher than in the other maneuvers. No effects were found for the other experimental manipulations.

		95% Confidence Interval			
	B Mmol/l	**Lower Bound Mmol/l**	**Upper bound Mmol/l**	**t**	***p-value***
Intercept	-0.039	-0.118	0.039	-1.001	0.319
Increased MWL	-0.005	-0.033	0.043	0.245	0.806
Motion	-0.005	-0.033	0.043	0.244	0.807
Maneuvers: 1st	0.040	-0.020	0.010	1.315	0.189
2nd	0.088	0.029	0.147	2.903	**0.004****
3rd	0.083	0.024	0.142	2.746	**0.006****
4th	0.047	-0.012	0.107	1.563	0.118
5th	0	0	0		
Trial Order: 1st	0.031	-0.028	0.091	1.049	0.294
2nd	0.002	-0.057	0.062	0.079	0.937
3rd	0.032	-0.027	0.091	1.056	0.291
4th	-0.026	-0.085	0.033	-0.855	0.393
5th	0	0	0		

Exploration other channels

No relationship was found between an HR and the MWL manipulation for HbO2. The analysis of HbR likewise did not indicate a relationship between MWL manipulations and a hemodynamic response. The interaction between optode channels and MWL also did not yield significant results, after Holm-Bonferroni correction. Results are presented in table 4 and 5.

Maneuvers gave differences for both concentrations for all maneuvers. This indicated that here too an effect can be found between maneuvers. However, quite contrastingly with the AF7 channels, the means start off quite high, then drop steadily for the most part. See also Figure 5 for the estimated means for each maneuver.

Table 4: Table showing results for mixed ANOVA for HbR for the explorative channels. Only maneuvers were related to changes in HR.

		95% Confidence Interval			
	B	**Lower Bound**	**Upper bound**	**t**	**p-value**
	Mmol/l	**Mmol/l**	**Mmol/l**		
Intercept	0,006	-0,022	0,033	0,398	0,691
Increased MWL	-0,005	-0,035	0,024	-0,350	0,727
Motion	-0,002	-0,009	0,004	-0,723	0,470
Maneuvers: 1st	0,040	0,030	0,051	70,600	**<0,0005****
2nd	0,015	0,004	0,025	20,776	**0,006****
3rd	0,032	0,021	0,042	50,994	**<0,0005****
4th	0,018	0,008	0,028	30,433	**0,001****
5th	0				
Trial Order: 1st	-0,001	-0,011	0,009	-0,186	0,853
2nd	-0,002	-0,012	0,008	-0,405	0,686
3rd	0,004	-0,006	0,014	0,770	0,441
4th	-0,011	-00,039	0,000	-10,996	0,046
5th	0	-00,046	0,033	0,398	0,691

Table 5: Table showing results for mixed ANOVA for HbO2 for the explorative channels. Only maneuvers were related to changes in HR.

		95% Confidence Interval			
	B	**Lower Bound**	**Upper bound**	**t**	**p-value**
	Mmol/l	**Mmol/l**	**Mmol/l**		
Intercept	,008	-,046	0,062	0,298	0,766
Increased MWL	,007	-,064	0,050	-0,222	0,824
Motion	,004	-,017	0,008	-0,669	0,504
Maneuvers: 1st	,079	,059	0,099	7,665	**<0.0005****
2nd	,028	,007	0,047	2,687	**0,007****
3rd	,064	,044	0,083	6,214	**<0.0005****
4th	,034	,014	0,054	3,325	**0,001****
5th	0	0	.	.	.
Trial Order: 1st	-,001	-,022	0,018	-0,138	0,89
2nd	-,003	-,023	0,017	-0,308	0,758
3rd	,011	-,009	0,031	1,1	0,271
4th	-,019	-,039	0,001	-1,88	0,06
5th	0	0	.	.	.

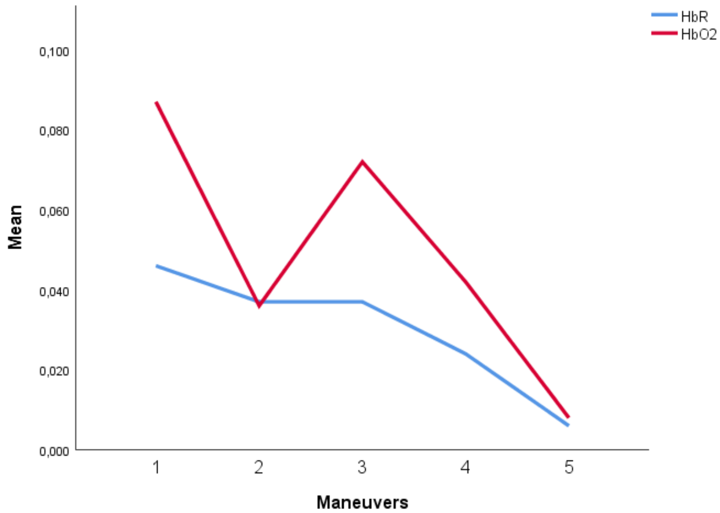


Figure 5: Means as estimated by the explorative mixed ANOVA's on HbO2 and HbR. Concentrations start off relatively high but drop off as the experiment progresses. HbR declines quite linearly, while HbO2 fluctuates around the second and third maneuver.

Discussion

A moving-base helicopter simulation was used to assess NIRS's capability to perform measurements on mental workload in a more ecologically valid, moving environment. Weather effects in the form of increased fog and wind were simulated to increase turbulence and reduce visual information to increase mental workload. Measurements were made, both with and without motion of the simulator, to see if there this affected measurements on potential mental workload related hemodynamic responses.

We hypothesized that NIRS would be able to detect mental workload related hemodynamic responses in the dorsolateral PFC, specifically around area AF7 of the international 10-5 system, irrespective of motion of the simulator. Results from the NASA TLX supported the assumptions that the simulation increased mental workload through the weather manipulation and not through the motion manipulation. No mental workload related hemodynamic responses were found in the NIRS data, though. While the setup did not manage to capture the expected effects, it did distinguish between certain maneuvers during the task. This gives interesting suggestions for future research into online measurements of mental workload with NIRS.

When diving deeper into the NASA TLX, the scores showed that participants experienced a high MWL overall in the experiment. When compared to Griers's meta-analysis of global NASA TLX scores in literature (2015), the estimated scores for the 'low' MWL conditions ranked around the 80th percentile and the 'high' conditions even scored somewhere between the 90th percentile and highest recorded scores. While there have been various experiments in more naturalistic settings, such as Ayaz et al.'s (2012) and Gateau et al.'s (2015), most of them had a big difference between the 'low' and 'high' MWL conditions. A bigger contrast between MWL states will probably give a bigger contrast in the hemodynamic response, as this has been shown to scale linearly (Fishburn, Norr, Medvedev & Vaidya, 2014).

The mean HR around area AF7 was positive for HbO2 as expected. However, for HbR it was also positive, which was not expected. Because a hemodynamic response is usually characterized by an eventual decrease in concentration, it would be expected to give a mainly negative hemodynamic response with activation (Tai and Chau, 2009). This did not make a good case for the current setup to have measured mental workload related hemodynamic responses, as was intended.

Three scenarios might explain this. Firstly, the set-up might not have focused on the correct place to measure task-related HR. Multiple studies found a correlation with mental workload in N-back tests and the HR in the dorsolateral PFC. However, the anterior medial cortex has also been found to correlate with mental workload (Ayaz et al., 2012). The current set-up focused on the dorsolateral PFC, since this was associated with MWL in many N-back tests. It might prove useful in future to include or focus on the medial frontal cortex, as this might associate more closely with this task's mental workload related activity. Secondly, the baseline might not have been ideal. The baseline was captured in the CMS, directly before each task. This was believed to give relatively little change in environment. However, these settings might also prove more intense, or participants might be more prone to visualize the task. Such factors might interfere with acquiring an accurate baseline. The third scenario is that the task simply did not succeed in eliciting a change in mental workload that was big enough to be captured by NIRS. This is possible, as previous research had relatively large differences in MWL in comparison to this experiment.

Neuroimaging analyses found very significant non-linear differences between maneuvers, distinguishing between different subtasks. Peaks were found around maneuver two and three, suggesting that something interesting happened around here. These peaks might be related to activation related to these maneuvers. Another option is that the differences in HR were time related. Due to its slower onset, halfway through the task hemodynamic response might have been at its maximum, where after it slowly declined again after reaching its peak. However, the peak would have been expected to persist, while task related activity persisted as well. That this did not happen would suggest a different or more complex cause, maybe more related to mental workload. These differences are at least interesting for future research into more 'online' measurements, as NIRS might apparently prove a useful tool for this. It is also important to realize that NIRS measurements might not be as straightforward as analyzing the mean HR and comparing them between tasks. Especially in more heterogeneous tasks, this might result in 'diluted' results, where differences between certain subtasks might cancel each other out when raked together into a single 'mean'. These results also indicate that this flying task was possibly a bit too complex for our current goals. The main goal was to differentiate between a lower mental workload state and a higher mental workload state. If the flying task had been reduced to the last 'hover' maneuver, for example, results might have been more pronounced or have had more power.

A very important drawback in this experiment was the amount of noise in the data irrespective of simulator generated motion. This meant that a lot of data was not usable and had to be filtered out. Simulator motion did not appear to be of significant influence, as there was almost no difference in how much data had to be filtered due to artefacting. Steps should be made to boost reliability. Both a non-translucent cap and tighter binding of the optodes may improve the signal-to-noise ratio. The use of an EEG-cap with the international 10-5 system assisted in linking specific channels to existing literature on brain anatomy and mental workload. Furthermore, it provided an accessible system to placing the optodes on the head. Optodes were fastened by locking them in rubber rings, fitted in the EEG cap. Other ways of fastening the optodes might decrease noise further. Also, the EEG cap was semitransparent as also visible in Figure 2 in the methodology. A non-transparent cap can probably minimize light artefacting further, especially in brightly lit environments.

Something to consider is that NIRS might require larger differences in mental workload, to be able to measure these differences. Research might use more proven methods of eliciting mental workload, to more reliably manipulate mental workload. This might make it easier to elicit larger differences.

Another point that might be considered in future is the fact that the cap could not cover the most prefrontal part of the cortex. Research such as Aghajani, Garbey & Omurtag's (2017) combined EEG and NIRS and used the EEG cap combined with a NIRS headband. This might also be considered for purely NIRS-based measurements to capture the most rostral part of the prefrontal cortex, together with the wider coverage possible with the EEG cap.

Finally, the amount of coverage that the current setup had, was limited by the amount of possible optodes. Only 8 LED's and 8 detectors could be placed on the head, with a maximum distance of 3.5 centimeters. More LED's and detectors can improve coverage and such a system might be considered, at the trade-off with more complexity in analyzing and interpreting the data.

Conclusion

The current set-up found interesting differences in hemodynamic responses between subtasks; these subtasks here being different maneuvers during the flying task. These differences raise two important points. Firstly, NIRS was able to differentiate between these subtasks, which is interesting on its own to follow up on. As mental workload may change drastically within a certain task NIRS might prove a useful methodology in providing some more 'online' insights in such changes. Secondly, it understates the importance of determining and distinguish such possible changes in mental workload in future. If such significant changes are not considered, when they occur within a task, they might distort overall results.

We did not find the expected mental workload related hemodynamic responses. This might be because the experiment did not measure on the correct position on the cortex, or due to problems with acquiring an accurate baseline. Another option was that the change in mental workload simply was not big enough to be measured by NIRS. Furthermore, differences in hemodynamic responses were found within the flying task, between subtasks called maneuvers. This could not be linked with mental workload, as we did not have a criterion variable, but drives home the fact that NIRS might be used for much more 'online' measurements than it currently has been. Furthermore, analyses might benefit from considering such possible differences distorting the data.

Certain suggestions can be made to improve on the current experiment. Future research might focus more on more ‘online’ measurements, or at least take within-task differences into account in the analysis, as NIRS distinguished hemodynamic response sizes between subtasks. Future researchers might also choose to focus more on the anterior medial cortex or parietal cortex. Furthermore, acquiring a baseline outside of testing conditions, might give different results. A less complex task, with bigger difference between MWL states may improve success when trying to capture mental workload related hemodynamic responses. Finally, the current set-up was limited to part of the prefrontal cortex by the amount of optodes possible with our system of 8 detectors and 8 LED’s, so a choice had to be made. A setup with more optodes might give more freedom in exploring task-related HR’s in the cortex, at the cost of more complexity.

Acknowledgements

I would like to extend my sincerest gratitude and respect to Dr. Ksander de Winkel, Stefano Geluardi and Carlo Gerboni for their enthusiastic and invaluable guidance through different stages of the project. Not only did they take time out of their busy schedules to set up most of the technical parts of the experiment, they welcomed me to Tübingen and gave me a home away from home. Furthermore, I would like to thank Dr. Rob van der Lubbe and Dr. Stéphanie van den Berg for guiding and encouraging me through the hardest parts of writing my thesis. Finally, I would like to thank my friends and everyone from the Max Planck Institute for Biological Cybernetic for helpful discussions, guidance and general good-humored banter. I am very grateful to have been part of this project and without these people, it would not have been possible.

References

Aghajani, H., Garbey, M., & Omurtag, A. (2017). Measuring mental workload with EEG+ fNIRS. *Frontiers in human neuroscience*, *11*, 359.

Aqil, M., Hong, K. S., Jeong, M. Y., and Ge, S. S. (2012). Cortical brain imaging by adaptive filtering of NIRS signals. Neuroscience Letters, 514(1), 35-41.

Ardila, A. (2012). The executive functions in language and communication. *Cognition and acquired language disorders (pp. 147-166).* Elsevier Inc.

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., and Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. Neuroimage, 59(1), 36-47.

Ayaz, H., Onaral, B., Izzetoglu, K., Shewokis, P. A., McKendrick, R., and Parasuraman, R. (2013). Continuous monitoring of brain dynamics with functional near-infrared spectroscopy as a tool for neuroergonomic research: empirical examples and a technological development. Front. Hum. Neurosci, 7.

Ayaz, H., Willems, B., Bunce, B., Shewokis, P. A., Izzetoglu, K., Hah, S., ... and Onaral, B. (2010). Cognitive workload assessment of air traffic controllers using optical brain imaging sensors. Advances in understanding human performance: Neuroergonomics, human factors design, and special populations, 21-31.

Berchtold, A. (2010). Sequence analysis and transition models. Encyclopedia of Animal Behavior, 139-145

Causse, M., and Matton, N. (2014). Using near-infrared spectroscopy to detect mental overload in flight simulator. Advances in Cognitive Engineering and Neuroergonomics, 11, 148.

Cain, B. (2007). A review of the mental workload literature. Defence Research and Development Toronto (Canada).

Callicott, J. H., Mattay, V. S., Bertolino, A., Finn, K., Coppola, R., Frank, J. A., ... and Weinberger, D. R. (1999). Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. Cerebral cortex, 9(1), 20-26.

Chance, B., Anday, E., Nioka, S., Zhou, S., Hong, L., Worden, K., ... and Thomas, R. (1998). A novel method for fast imaging of brain function, non-invasively, with light. Optics express, 2(10), 411-423

Dunn, A. K., Devor, A., Bolay, H., Andermann, M. L., Moskowitz, M. A., Dale, A. M., & Boas, D. A. (2003). Simultaneous imaging of total cerebral hemoglobin concentration, oxygenation, and blood flow during functional activation. *Optics letters*, *28*(1), 28-30.

Durantin, G., Gagnon, J. F., Tremblay, S. and Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. Behavioural brain research, 259, 16-23.

Everitt, B., and Skrondal, A. (2002). The Cambridge dictionary of statistics (Vol. 106). Cambridge: Cambridge University Press.

Fishburn, F. A., Norr, M. E., Medvedev, A. V., and Vaidya, C. J. (2014). Sensitivity of fNIRS to cognitive state and load. Frontiers in human neuroscience, 8, 76.

Fox, M. D., Corbetta, M., Snyder, A. Z., Vincent, J. L., & Raichle, M. E. (2006). Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences, 103*(26), 10046-10051.

Gałecki, A., and Burzykowski, T. (2013). Linear mixed-effects model. In Linear Mixed-Effects Models Using R (pp. 245-273). Springer, New York, NY.

Gateau, T., Durantin, G., Lancelot, F., Scannella, S., and Dehais, F. (2015). Real-time state estimation in a flight simulator using fNIRS. PloS one, 10(3), e0121279.

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T. P., Lin, C. T., ... and Makeig, S. (2011). Cognition in action: imaging brain/body dynamics in mobile humans. Reviews in the Neurosciences, 22(6), 593-608.

Grier, R. A. (2015, September). How high is high? A meta-analysis of NASA-TLX global workload scores. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting (Vol. 59, No. 1, pp. 1727-1731). Sage CA: Los Angeles, CA: SAGE Publications.

Harbluk, J. L., Noy, Y. I., Trbovich, P. L., and Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. Accident Analysis and Prevention, 39(2), 372-379.

Harrison, J., İzzetoğlu, K., Ayaz, H., Willems, B., Hah, S., Ahlstrom, U., ... and Onaral, B. (2014). Cognitive workload and learning assessment during the implementation of a next-generation air traffic control technology using functional near-infrared spectroscopy. IEEE Transactions on Human-Machine Systems, 44(4), 429-440.

Hart, S. G. (2006, October). NASA-task load index (NASA-TLX); 20 years later. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 50, No. 9, pp. 904-908). Sage CA: Los Angeles, CA: Sage Publications.

Hart, S. G., and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. Advances in psychology, 52, 139-183.

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., and Schultz, T. (2014). Mental workload during n-back task—quantified in the prefrontal cortex using fNIRS. Frontiers in human neuroscience, 7, 935.

Hill, A. P., and Bohil, C. J. (2016). Applications of Optical Neuroimaging in Usability Research. Ergonomics in design, 24(2), 4-9.

Hirshfield, L. M., Solovey, E. T., Girouard, A., Kebinger, J., Jacob, R. J., Sassaroli, A., and Fantini, S. (2009, April). Brain measurement for usability testing and adaptive interfaces: an example of uncovering syntactic workload with functional near-infrared spectroscopy. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 2185-2194). ACM.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics, 65-70.

Huang, D., Chen, S., Wang, S., Shi, J., Ye, H., Luo, J., & Zheng, H. (2017). Activation of the DLPFC reveals an asymmetric effect in risky decision making: evidence from a tDCS study. *Frontiers in psychology*, *8*, 38.

Huber, P. J. (2011). Robust statistics. In International Encyclopedia of Statistical Science (pp. 1248-1251). Springer Berlin Heidelberg.

IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Izzetoglu, K. (2008). Neural correlates of cognitive workload and anesthetic depth: fNIR spectroscopy investigation in humans (Doctoral dissertation, Drexel University).

Izzetoglu, M., Bunce, S. C., Izzetoglu, K., Onaral, B., and Pourrezaei, K. (2007). Functional brain imaging using near-infrared technology. IEEE Engineering in Medicine and Biology Magazine, 26(4), 38.

Jansma, J. M., Ramsey, N. F., Coppola, R., and Kahn, R. S. (2000). Specific versus nonspecific brain activity in a parametric N-back task. Neuroimage, 12(6), 688-697.

Kahneman, D. (1973). Attention and effort (p. 246). Englewood Cliffs, NJ: Prentice-Hall.

Kaller, C. P., Rahm, B., Spreer, J., Weiller, C., and Unterrainer, J. M. (2010). Dissociable contributions of left and right dorsolateral prefrontal cortex in planning. Cerebral cortex, 21(2), 307-317.

Keshmiri, S., Sumioka, H., Yamazaki, R., and Ishiguro, H. (2017). A Non-parametric Approach to the Overall Estimate of Cognitive Load Using NIRS Time Series. Front. Hum. Neurosci. 11: 15. doi: 10.3389/fnhum.

Koessler, L., Maillard, L., Benhadid, A., Vignal, J. P., Felblinger, J., Vespignani, H., and Braun, M. (2009). Automated cortical projection of EEG sensors: anatomical correlation via the international 10–10 system. Neuroimage, 46(1), 64-72.

Koo, B., Lee, H. G., Nam, Y., Kang, H., Koh, C. S., Shin, H. C., and Choi, S. (2015). A hybrid NIRS-EEG system for self-paced brain computer interface with online motor imagery. Journal of neuroscience methods, 244, 26-32.

Kwak, C., and Clayton-Matthews, A. (2002). Multinomial logistic regression. Nursing research, 51(6), 404-410.

Lodemann, P., Schorer, G., and Frey, B. M. (2010). Wrong molar hemoglobin reference values—a longstanding error that should be corrected. Annals of hematology, 89(2), 209.

Loft, S., Sanderson, P., Neal, A., and Mooij, M. (2007). Modeling and predicting mental workload in en route air traffic control: Critical review and broader implications. Human Factors, 49(3), 376-399.

von Lühmann, A., Wabnitz, H., Sander, T., and Müller, K. R. (2017). M3BA: A Mobile, Modular, Multimodal Biosignal Acquisition Architecture for Miniaturized EEG-NIRS-Based Hybrid BCI and Monitoring. IEEE Transactions on Biomedical Engineering, 64(6), 1199-1210.

Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., and Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. Biological psychology, 121, 62-73.

MATLAB and Statistics Toolbox Release 2016a, The MathWorks, Inc., Natick, Massachusetts, United States.

Mehta, R. K., & Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Frontiers in human neuroscience*, *7*, 889.

Nieuwenhuizen, F. M. (2013). The MPI cybermotion simulator: a novel research platform to investigate human control behavior. Journal of Computing Science and Engineering, 7(2), 122-131.

Pace, N. L., and Briggs, W. M. (2009). Stepwise logistic regression. Anesthesia and Analgesia, 109(1), 285-286.

Perfect, P., Jump, M., and White, M. D. (2015). Methods to Assess the Handling Qualities Requirements for Personal Aerial Vehicles. Journal of Guidance, Control, and Dynamics, 38(11), 2161-2172.

Petersen, S. E., Van Mier, H., Fiez, J. A., and Raichle, M. E. (1998). The effects of practice on the functional anatomy of task performance. Proceedings of the National Academy of Sciences, 95(3), 853-860.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Roy, R. N., Charbonnier, S., Campagne, A., and Bonnet, S. (2016). Efficient mental workload estimation using task-independent EEG features. Journal of neural engineering, 13(2), 026019.

Rousseeuw, P. J., and Leroy, A. M. (2005). Robust regression and outlier detection (Vol. 589). John wiley and sons.

Rubio, S., Díaz, E., Martín, J., and Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA TLX, and workload profile methods. Applied Psychology, 53(1), 61-86.

Ruocco, A. C., Rodrigo, A. H., Lam, J., Di Domenico, S., Graves, B., and Ayaz, H. (2014). A problem-solving task specialized for functional neuroimaging: validation of the Scarborough adaptation of the Tower of London (S-TOL) using near-infrared spectroscopy. Frontiers in human neuroscience, 8, 185.

Sheldon, M. R., Fillyaw, M. J., and Thompson, W. D. (1996). The use and interpretation of the Friedman test in the analysis of ordinal scale data in repeated measures designs. Physiotherapy Research International, 1(4), 221-228.

Strangman, G., Culver, J. P., Thompson, J. H., and Boas, D. A. (2002). A quantitative comparison of simultaneous BOLD fMRI and NIRS recordings during functional brain activation. Neuroimage, 17(2), 719-731.

Tai, K., and Chau, T. (2009). Single-trial classification of NIRS signals during emotional induction tasks: towards a corporeal machine interface. Journal of neuroengineering and rehabilitation, 6(1), 39.

The American National Red Cross (2019, July 28), Donors deferred for low hemoglobin. Retrieved from: https://www.redcrossblood.org/donate-blood/blood-donation-process/before-during-after/iron-blood-donation/donors-deferred-forlowhemoglobin.html

Unni, A., Ihme, K., Surm, H., Weber, L., Lüdtke, A., Nicklas, D., ... and Rieger, J. W. (2015, October). Brain activity measured with fNIRS for the prediction of cognitive workload. In Cognitive Infocommunications (CogInfoCom), 2015 6th IEEE International Conference on (pp. 349-354). IEEE.

Vertanen, K. (2017, March 17). NASA-TLX in HTML and JavaScript [Computer software]. Retrieved from https://www.keithv.com/software/nasatlx/

Vallesi, A. (2012). Organisation of executive functions: hemispheric asymmetries. Journal of Cognitive Psychology, 24(4), 367-386.

Waitzman, D.M., (2017). Chapter 20 - Oculomotor Systems and Control. *Conn's Translational Neuroscience (pp. 439-465).* Elsevier Inc.

Warner, P. (2008). Ordinal logistic regression. BMJ Sexual and Reproductive Health, 34(3), 169-170.

Wickens, C. D., Hollands, J. G., Banbury, S., and Parasuraman, R. (2015). Engineering psychology and human performance. Psychology Press.

Wilkinson, G. N., and Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. Applied Statistics, 392-399.

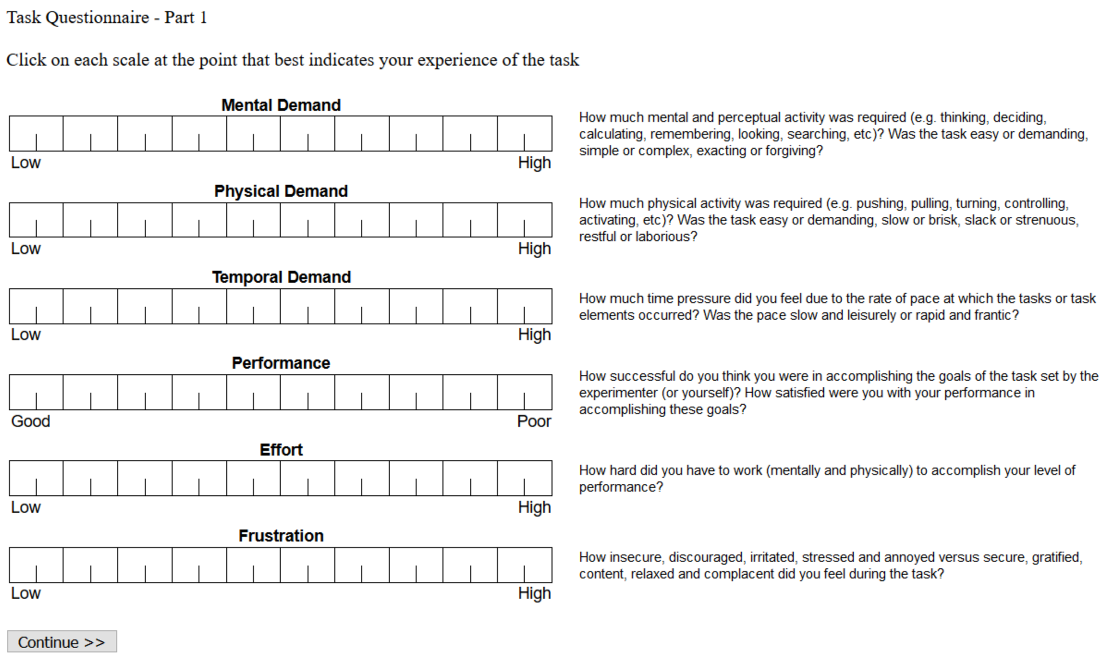
de Winkel, K. N., Nesti, A., Ayaz, H., and Bülthoff, H. H. (2017). Neural correlates of decision making on whole body yaw rotation: an fNIRS study. Neuroscience Letters.

Young, M. S., Brookhuis, K. A., Wickens, C. D., and Hancock, P. A. (2015). State of science: mental workload in ergonomics. Ergonomics, 58(1), 1-17.

de Winkel, K. N., Nesti, A., Ayaz, H., & Bülthoff, H. H. (2017). Neural correlates of decision making on whole body yaw rotation: an fNIRS study. *Neuroscience Letters*.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics*, *58*(1), 1-17.

Appendix A: NASA Task Load Index


Task Questionnaire - Part 1

Click on each scale at the point that best indicates your experience of the task

Mental Demand
Low High
How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Physical Demand
Low High
How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Temporal Demand
Low High
How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

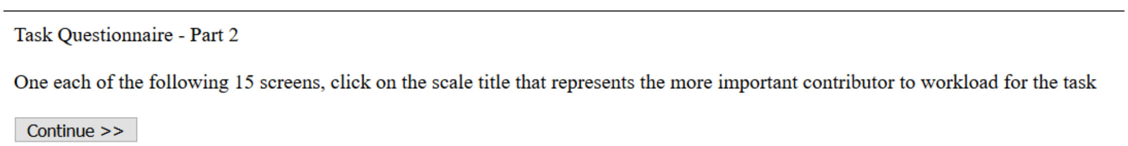
Performance
Good Poor
How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Effort
Low High
How hard did you have to work (mentally and physically) to accomplish your level of performance?

Frustration
Low High
How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Continue >>

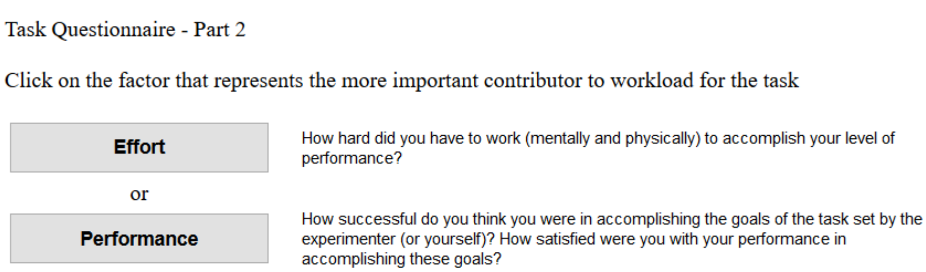
Figure A1: First screen of the NASA TLX. Participants could click where on the scale they felt each dimension of mental workload fell during the helicopter flying task.


Task Questionnaire - Part 2

One each of the following 15 screens, click on the scale title that represents the more important contributor to workload for the task

Continue >>

Figure A2: Second screen of the NASA TLX, explaining the second part of the test.


Task Questionnaire - Part 2

Click on the factor that represents the more important contributor to workload for the task

Effort
How hard did you have to work (mentally and physically) to accomplish your level of performance?

or

Performance
How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Figure A3: Third to eighteenth screen of the NASA TLX. Participants had to choose which of two shown dimensions contributed more to workload for the task. This screen is an example of a total of fifteen screens that were shown. Here comparing the dimensions ‘effort’ and ‘performance’. All other fifteen combinations of the six dimensions also passed by, each once.

Appendix B: Instructions for Participant in English

Max Planck Institute for Biological Cybernetics
Human Perception Cognition and Action
Spemannstr. 38, D-72076 Tübingen

MAX-PLANCK-GESELLSCHAFT

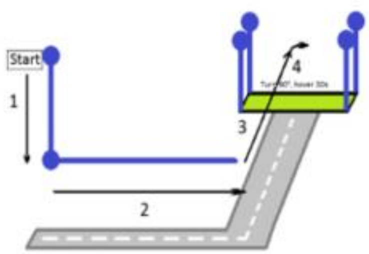
BIOLOGISCHE KYBERNETIK

Appendix B: Instructions for Participant in English

Near-InfraRed Spectroscopy (NIRS) is a technique to measure oxygen consumption in the brain using near-infrared light. By monitoring oxygen consumption, we can make inferences about the functioning of the brain. Because light is harmless and unaffected by motion or electrical signals, this technique is a promising neuroimaging tool for various situations. The goal of the present study is to determine how well NIRS performs when a participant is seated in a moving machine.

During the experiment, you will take place in a motion simulator and wear a NIRS system. You will also wear a headset that lets you communicate with the researchers. In the motion simulator you will control a simulated helicopter. To control the helicopter you will use: a cyclic stick to adjust the tilt of the helicopter, a collective lever to control height, and foot pedals to turn the helicopter left and right (i.e. control yaw). You will have to complete several trials along a track that has been laid out for you.

A trial goes as follows (see figure 1): you start by pressing a button on the collective lever; from there, the simulation starts with you hovering in the air in front of a globe; now:



1) **move your helicopter straight down** until you see another globe in front of you; *push the cyclic button*, next,
2) **move sideways to the right** following the line in front of you, until you see a transparent box in front of you; *push the cyclic button*,
3) **fly forwards, into the box**; *push the cyclic button* finally,
4) **hover, turn your nose right 90 degrees on the spot and** *push the cyclic button* **remain hovering for 10 seconds**.

The simulator will then be reset, and you can start a new trial by pressing the button on the collective lever.

There are four blocks of five trials each: Two blocks will have motion, two blocks will have no motion of the simulator. Then within these two blocks there will be either nice weather or bad weather. The experimenter will tell you the conditions of the current block in advance. There will be breaks between blocks. During the breaks, the experimenter will come by with a questionnaire on how much mental workload you experienced. Mental workload refers to how hard you had to work mentally to be able to complete the task. We will test whether the NIRS recordings can inform us of your mental workload. It is important that you complete the task to the best of your abilities.

The experiment will take around 3 hours. The data will be kept pseudonymously for 20 years at the Max Planck Institute. If at any time during the experiment you wish to have a break, you start to feel dizzy, get a headache, start sweating, or feel any other form of discomfort, please communicate this immediately to the researchers. The experiment will then be paused or stopped. If at any point you would like to withdraw your participation from the study, you are free to do so without penalty.

Joost Zijlstra, BSc, Dr. Ksander de Winkel &
Dr. Eng. Stefano Geluardi

Max Plank Institute for Biological Cybernetics
Human Perception Cognition and Action

Spemannstr. 41, D-72076 Tübingen
Tel: 07071 601 643
Email: joost.zijlstra@tuebingen.mpg.de

Appendix C: Instructions for Participant in German

MAX-PLANCK-GESELLSCHAFT

Max Planck Institute for Biological Cybernetics
Human Perception Cognition and Action
Spemannstr. 38, D-72076 Tübingen

BIOLOGISCHE KYBERNETIK

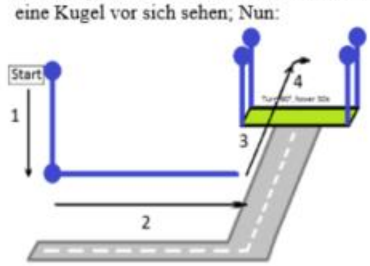
Anweisung für Versuchspersonen

Near-InfraRed Spectroscopy (NIRS) ist eine Technik, die mit Hilfe von nah-Infrarot Licht den Sauerstoffverbrauch im Gehirn misst. Durch die Messungen des Sauerstoffverbrauchs können wir Schlussfolgerungen über die Funktionsweisen des Hirns ziehen. Das Licht ist ungefährlich und nicht beeinträchtigt von Bewegungen oder elektrischen Signalen. Dadurch ist es ein vielversprechendes Mittel zur Bildgebung des zentralen Nervensystems. Das Ziel dieser Studie ist es, zu bestimmen, inwiefern NIRS durchzuführen ist, wenn der Teilnehmer in einer bewegenden Maschine sitzt.

Während des Experiments werden Sie in einem Bewegungssimulator sitzen und das NIRS System tragen. Außerdem werden Sie ein Headset tragen, mit dem Sie mit dem Wissenschaftler kommunizieren können.

In dem Bewegungssimulator werden Sie einen Simulationshelikopter steuern. Um diesen zu steuern, gebrauchen sie Folgendes: Einen zyklischen Stick, um die Neigung des Helikopters anzupassen, einen Hebel um die Höhe zu kontrollieren und ein Pedal, um die Spitze des Helikopters nach rechts oder links zu bewegen. Sie werden verschiede Testläufe auf einer Bahn durchlaufen, die für Sie ausgelegt ist.

Ein Testlauf verläuft folgendermaßen (siehe auch Figur 1): Starten sie die Simulation, indem sie auf den Knopf auf dem Hebel drücken; daraufhin wird die Simulation beginnen und abheben, sodass Sie eine Kugel vor sich sehen; Nun:



1) **Bewegen Sie den Helikopter gerade nach unten** bis Sie eine andere weitere Kugel vor sich sehen; als nächstes
2) **Bewegen Sie den Helikopter rechts zur Seite** und folgen Sie dabei der Linie vor Ihnen, bis Sie eine transparente Box vor sich sehen;
3) **Fliegen sie vorwärts, in die Box**;
4) zuletzt, schweben Sie den Helikopter in die Box, **Turmen Sie die Nase 90 Grad nach rechts, und schweben Sie dort für 10 Sekunden.**
Darauffolgend wird die Simulation zurückgesetzt und sie können einen neuen Testlauf beginnen, indem Sie auf den Knopf auf dem Hebel drücken.

Es gibt vier Blöcke in den fünf Versuchen: einer von diesen ist mit gutem Wetter; der andere ist mit schlechtem Wetter. Der Wissenschaftler wird Ihnen zuvor sagen, in welcher Lage Sie sich befinden. Es wird außerdem Pausen zwischen den Blöcken geben. In diesen Pausen wird der Wissenschaftler Sie bitten, Ihre mentale Belastung zu bewerten. Wir werden testen, ob die NIRS Aufzeichnungen uns über Ihre mentale Belastungen informieren kann. Es ist hierbei wichtig, dass Sie das Testlauf nach besten Kräften ausführen.

Das Experiment wird ungefähr 3 Stunden dauern. Die Daten werden pseudonymisiert aufbewahrt werden für eine Zeitdauer von 20 Jahren im Max-Planck-Institut. Wenn Sie während des Experiments eine Pause wünschen, Ihnen schwindelig wird, Sie Kopfschmerzen bekommen, zu Schwitzen beginnen oder sich in irgendeiner anderen Art und Weise unwohl fühlen, kontaktieren Sie bitte sofort den Wissenschaftler. Das Experiment wird dann gestoppt. Wenn Sie im Laufe der Prozedur das Experiment nicht mehr fortsetzen wollen, ist es Ihnen freigestellt, Ihre Teilnahme ohne jegliche Konsequenzen zu beenden.

Joost Zijlstra, BSc, Dr. Ksander de Winkel &
Dr. Eng. Stefano Geluardi

Max Plank Institute for Biological Cybernetics
Human Perception Cognition and Action

Spemannstr. 41, D-72076 Tübingen
Tel: 07071 601 643
Email: joost.zijlstra@tuebingen.mpg.de

Appendix D: Informed Consent in English

MAX-PLANCK-GESELLSCHAFT

Max Planck Institute for Biological Cybernetics
Human Perception Cognition and Action
Spemannstr. 38, D-72076 Tübingen

BIOLOGISCHE KYBERNETIK

Informed Consent

I, ___________________________, state that I am ______ years of age and that I agree to participate in a research study being conducted by Joost Zijlstra of the Max Planck Institute for Biological Cybernetics.

I hereby declare that I fully understood the instructions and risk as described on the 'Instructions for participants' page. I acknowledge that ________________________ has informed me that my participation in this study is voluntary; that I may withdraw my participation at any time without penalty, and that all data that I contribute will remain confidential and will be stored safely in pseudonymous form. The data will be saved at the Max-Planck-Institute for 20 years and deleted afterwards. I understand that my mental workload will be measured with near-infrared spectroscopy while controlling a helicopter in a motion simulator and that I will receive a complete explanation at the end of my participation. I understand that the study involves no serious risk.

_______________________________________ ____________

(Signature of Participant) (Date)

_______________________________________ ____________

(Signature of Researcher) (Date)

If you would like to receive by email (______________________________________):

[ ] a summary of the results at the conclusion of the study

[ ] a scanned copy of this informed consent form

Joost Zijlstra, BSc, Dr. Ksander de Winkel &
Dr. Eng. Stefano Geluardi
Max Plank Institute for Biological Cybernetics

Human Perception Cognition and Action

Spemannstr. 41, D-72076 Tübingen
Tel: 07071 601 643
Email: joost.zijlstra@tuebingen.mpg.de

Appendix E: Informed consent in German

MAX-PLANCK-GESELLSCHAFT

Max Planck Institute for Biological Cybernetics
Human Perception Cognition and Action
Spemannstr. 38, D-72076 Tübingen

BIOLOGISCHE KYBERNETIK

Einverständniserklärung

Ich, ______________________________, erkläre hiermit, dass ich ______ Jahre alt bin und bereit bin in einem Experiment, durchgeführt von Joost Zijlstra vom Max-Planck-Institut für Biologische Kybernetik, teilzunehmen.

Hiermit erkläre ich, dass ich die Anweisungen wie beschrieben bei die Informationen für Versuchspersonen verstanden habe. Ich bestätige, dass ____________________ mich darüber aufgeklärt hat, dass meine Teilnahme an dem Experiment freiwillig ist und ich jederzeit ohne Konsequenzen zurücktreten kann. Meine Daten werden vertraulich behandelt und in pseudonymisierter Form für 20 Jahre am Max-Planck-Institut gespeichert und anschließend gelöscht. Ich bin mir darüber im Klaren, dass in das Experiment meine mentale Belastung gemessen werde mit Near-Infrared Spectroscopy und daß ich am Ende des Experimentes eine umfängliche Erklärung erhalten werde. Ich bin mir darüber im Klaren, dass das Experiment keine ernsthaften Risiken beinhält.

___________________________________ ___________

(Unterschrift Versuchsperson) (Datum)

___________________________________ ___________

(Unterschrift Versuchsleiter) (Datum)

Falls Sie per email (__________________________________) bekommen möchten:

[ ] eine Zusammenfassung der Resultate des Experiments

[ ] Eine Kopie dieser Einverständniserklärung

Joost Zijlstra, BSc, Dr. Ksander de Winkel &
Dr. Eng. Stefano Geluardi
Max Plank Institute for Biological Cybernetics

Human Perception Cognition and Action

Spemannstr. 41, D-72076 Tübingen
Tel: 07071 601 643
Email: joost.zijlstra@tuebingen.mpg.de

Appendix F: Liability waiver in English

Max-Planck-Institut für biologische Kybernetik

Clarification for Test Persons

Last name, First name: ………………………………………….....

Address: …………….………………………………………………….....

By signing below you confirm that:

- You have been informed about insurance coverage.
- You have read and understood this clarification document and the "Safety Instructions motion simulator".

Participation in the demonstration of the motion simulator is completely voluntary. You can terminate participation at any time. Inappropriate or uncontrolled driving of the motion simulator represents a substantial safety risk for **all** persons taking part in the demonstrations and can result in serious physical injuries. Safety must have the highest priority **during any running** of the motion simulator. Therefore:

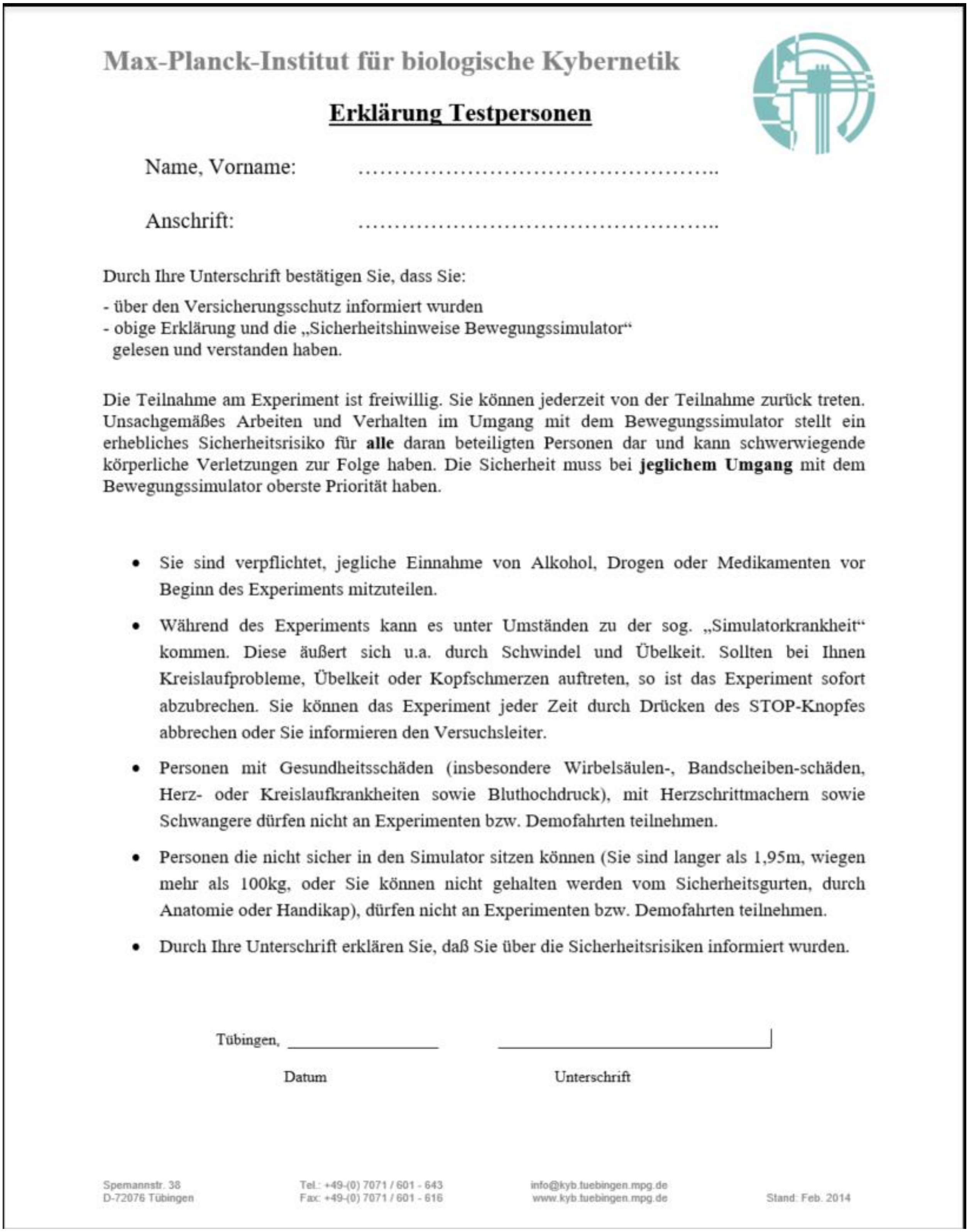
- You are required to declare any and all consumption of alcohol, drugs or medication before being seated in the simulator.
- Should problems with blood circulation, nausea or headache occur during the ride, the ride should terminate immediately. You can terminate the ride at any time by pushing the STOP button or informing the experimenter.
- People with health problems (in particular spinal column or intervertebral disc problems, heart or circulatory disease, as well as high blood pressure) or people with heart pacemaker are not allowed to ride. Pregnant women are also excluded from participating.
- People who are not able to sit in the simulator safely (longer than 1,95m, weighing more than 100kg or who cannot be held by safety belts by anatomy or handicap) are not allowed to ride.
- By signing below, you declare that you have been informed of the safety risks.

Tübingen, __________________ ________________________

Date Signature

Spemannstr. 38
D-72076 Tübingen
Tel.: +49-(0) 7071 / 601 – 643
Fax: +49-(0) 7071 / 601 – 616
info@kyb.tuebingen.mpg.de
www.kyb.tuebingen.mpg.de
Stand: Februar 2014

Appendix G: Liability waiver in German



Max-Planck-Institut für biologische Kybernetik

Erklärung Testpersonen

Name, Vorname: ……………………………………………..

Anschrift: ……………………………………………..

Durch Ihre Unterschrift bestätigen Sie, dass Sie:

- über den Versicherungsschutz informiert wurden
- obige Erklärung und die „Sicherheitshinweise Bewegungssimulator" gelesen und verstanden haben.

Die Teilnahme am Experiment ist freiwillig. Sie können jederzeit von der Teilnahme zurück treten. Unsachgemäßes Arbeiten und Verhalten im Umgang mit dem Bewegungssimulator stellt ein erhebliches Sicherheitsrisiko für **alle** daran beteiligten Personen dar und kann schwerwiegende körperliche Verletzungen zur Folge haben. Die Sicherheit muss bei **jeglichem Umgang** mit dem Bewegungssimulator oberste Priorität haben.

- Sie sind verpflichtet, jegliche Einnahme von Alkohol, Drogen oder Medikamenten vor Beginn des Experiments mitzuteilen.
- Während des Experiments kann es unter Umständen zu der sog. „Simulatorkrankheit" kommen. Diese äußert sich u.a. durch Schwindel und Übelkeit. Sollten bei Ihnen Kreislaufprobleme, Übelkeit oder Kopfschmerzen auftreten, so ist das Experiment sofort abzubrechen. Sie können das Experiment jeder Zeit durch Drücken des STOP-Knopfes abbrechen oder Sie informieren den Versuchsleiter.
- Personen mit Gesundheitsschäden (insbesondere Wirbelsäulen-, Bandscheiben-schäden, Herz- oder Kreislaufkrankheiten sowie Bluthochdruck), mit Herzschrittmachern sowie Schwangere dürfen nicht an Experimenten bzw. Demofahrten teilnehmen.
- Personen die nicht sicher in den Simulator sitzen können (Sie sind langer als 1,95m, wiegen mehr als 100kg, oder Sie können nicht gehalten werden vom Sicherheitsgurten, durch Anatomie oder Handikap), dürfen nicht an Experimenten bzw. Demofahrten teilnehmen.
- Durch Ihre Unterschrift erklären Sie, daß Sie über die Sicherheitsrisiken informiert wurden.

Tübingen, ________________ ______________________________

Datum Unterschrift

Spemannstr. 38
D-72076 Tübingen

Tel.: +49-(0) 7071 / 601 - 643
Fax: +49-(0) 7071 / 601 - 616

info@kyb.tuebingen.mpg.de
www.kyb.tuebingen.mpg.de

Stand: Feb. 2014

Appendix H: Safety Instructions for the Cybermotion Simulator in English

Safety Instructions CyberMotion Simulator (CMS)

- ***The instructions of the operating staff must be obeyed!***
- Maximal payload => **100kg**
- Smoking, the presence of animals, the carrying of bags, backpacks, sticks and other bulky items nearby the simulator is forbidden.
- People with health issues aren't allowed to use the simulator. This is particularly valid for people suffering from spinal or disc problems or suffering from high blood presure and/or therefore receiving medical treatment.
  Pregnant people and people with heart pace maker and furthermore people with heart and/or circulatory disorders are also excluded using the simulator.
- People which can't be securely held by the existing safety belts, or people who aren't able to hold on caused by anatomy or handicap are also excluded using the simulator.
- Intoxicated people are excluded using the simulator
- Loose objects (for example mobile phone, bunch of keys, long necklaces, coins, etc.) must be checked in prior to the use of the simulator.
- Earrings and scarves must be checked in prior to the use of the open seat setup.

Max-Planck-Institut für Biologische Kybernetik, September 2015

Appendix I: Safety Instructions for the Cybermotion Simulator in German

Sicherheitshinweise CyberMotion Simulator (CMS)

- ***Den Anweisungen des Bedienpersonals ist Folge zu leisten!***
- Die maximale Zuladung beträgt **100** kg,
- Rauchen, das Mitnehmen von Tieren, sowie das Mitführen von Taschen, Rucksäcken, Stöcken und anderen sperrigen Gegenständen am Simulator ist verboten.
- Personen mit Gesundheitsschäden dürfen den Simulator nicht benutzen. Dies gilt insbesondere für Personen, die an Wirbelsäulen- oder Bandscheibenschäden oder an Bluthochdruck leiden und/oder deswegen in ärztlicher Behandlung stehen. Schwangere Personen, sowie Personen mit Herz- und/oder Kreislauferkrankungen, sind ebenfalls von der Benutzung des Simulators ausgeschlossen.
- Personen, welche z.B. aufgrund ihres Körperbaus oder durch Behinderung von den vorhandenen Sicherheitsgurten nicht sicher gehalten werden und/oder sich nicht festhalten können, sind von der Benutzung des Simulators ausgeschlossen.
- Betrunkene und/oder berauschte Personen sind von der Benutzung des Simulators ausgeschlossen.
- Lose Gegenstände (z.B. Mobiltelefon, Schlüsselbund, lange Halsketten, Münzgeld, etc.) sind vor der Benutzung des Simulators abzugeben.
- Spitze Ohrringe und Ohrstecker sowie Schals sind vor der Benutzung des Simulators abzunehmen.

Max-Planck-Institut für Biologische Kybernetik, September 2015