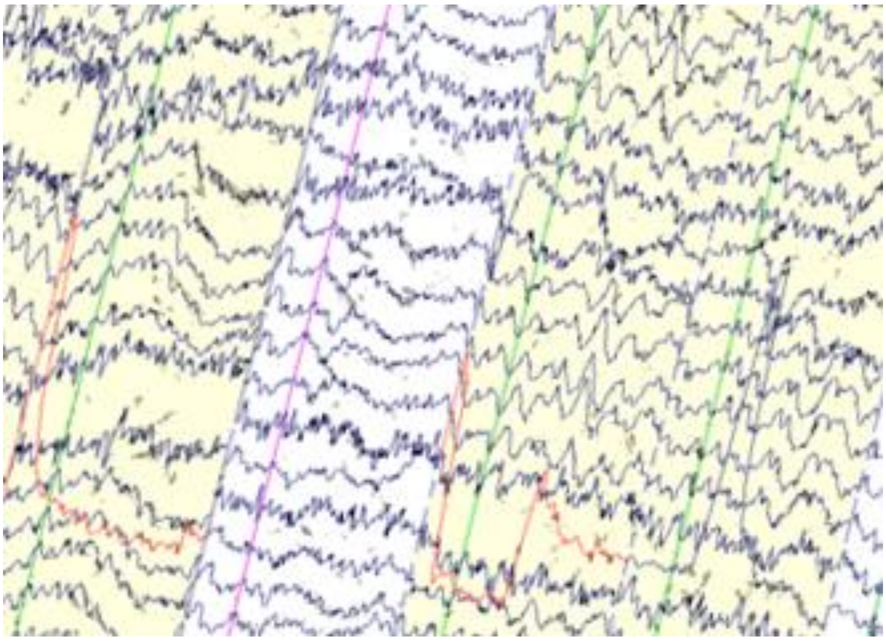


Assessing Cognitive Workload Between Different Tasks

Using EEG to develop and examine a method to measure variation of cognitive workload between different levels of difficulty

Master's thesis in Biomedical Engineering[1] and Learning and Leadership[2]

FANNY APELGREN[1] & IDA PETTERSSON[2]

DEPARTMENT OF COMMUNICATION AND LEARNING IN SCIENCE

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2020
www.chalmers.se

Assessing Cognitive Workload Between Different Tasks: Using EEG to develop and examine a method to measure variation of cognitive workload between different levels of difficulty

Fanny Apelgren and Ida Pettersson

Supervisor: Eva Lendaro, Department of Electrical Engineering, Chalmers University of Technology

Supervisor: Sheila Galt, Department of Communication and Learning in Science, Chalmers University of Technology

Examiner: Samuel Bengmark, Department of Mathemathical Science, Chalmers University of Technology

Master's Thesis 2020

Communication and Learning in Science

Chalmers University of Technology

SE-412 96 Gothenburg

Telephone +46 31 772 1000

Cover: The picture is taken from our own EEG data

Abstract

Assessing cognitive workload is an important tool, for example when evaluating different techniques for improving prostheses. Here, we have developed a method to compare how the cognitive workload differs if a prosthesis has sensory feedback or not. We have used electroencephalogram (EEG) and performed a pilot study on ten intact limb subjects. An *easy* and *hard* level were constructed by changing the weight of a force sensitive cube that were to be lifted back and forth over a barrier while counting sounds in an auditory oddball task. A third level consisted of only the oddball task. The difference in difficulty between the different levels were verified by measuring performance, and perceived effort. On a group level, these measurements all indicated that the *no task* condition was easiest, and the *hard task* condition was most demanding. Measurements of the number of lifts for different repetitions of the *easy* and *hard* conditions also showed signs of a learning effect during the performance of the *easy task*. The cognitive workload was measured by using the event-related potentials (ERP) technique and frequency bands. The results showed that the ERP component P3 was the only one that could indicate a significant difference between all three levels. A comprised measurement (consisting of the sum of ERP components N1, P2, P3, and LPP) and the alpha frequency bands (low-, high-, and broadband alpha) also showed a significant effect between some of the conditions.

Keywords: *Cognitive workload, Mental load, Learning, Electroencephalogram, EEG, Event-Related Potential, ERP, P3, Grasping task, Oddball Task, Dual-Task Paradigm*

Acknowledgements

The writing of this thesis has, as most other projects, been a rocky road. We have had moments on top where everything has gone our way and we felt like real scientist ready for our big breakthrough, but we have also stumbled around in dark valleys, unsure of what to do next. Nevertheless, we have finally made it, and here is our finished work! But we would have never made it without the support we have gotten from people around us.

We are very grateful for the Biomechatronics and Neurorehabilitation lab under Max Ortiz Catalan for the chance to do this, and for lending us both their time and equipment. Thank you everyone for meeting us and sharing your knowledge and experience, and for changing your plans to let us perform our measurements at the lab. Especially we would like to thank Shahrzad Damercheli who patiently helped us during all our measurements and training. And, of course, a huge thank you to our supervisor Eva Lendaro who guided us along the way, believed in us, and taught us how to think like scientists.

We are also very grateful for our never-ending source of support, feedback, ideas and hot chocolate (at least this would have been never-ending if hot chocolate could have been sent online): Sheila Galt. Thank you for everything!

We also had some other meeting to help us with our literature review and statistical analysis. So, thank you to Yommine Hjalmarsson and Serik Sagitov for sharing your expertise.

This project would also not have been possible without all the wonderful people who volunteered to spend hours with us while we poked them in the head with gel-filled syringes and asked them to move a cube hundreds of times. Thank you to everyone who somehow took part in our experiment, helped us practice, contributed your ideas in our mini-pilot study or offered to help, but were stopped by the arrival of the Covid-19 virus.

We are also very thankful for our friends and family who have supported us through this work, during panic and euphoria. Everyone who has taken time to listen when we tried to explain our work, or helped us when it was time to think about something else by sharing a hug, a lunch, or an ice cream.

Lastly, we would like to thank the two people who have been most important for this work: each other! As mentioned, this work has taken us over high mountains and through dark valleys and deep pits, with many treacherous rocks to stumble on. Sometimes, working together has been challenging, but it was by doing it together that we got out of those pit holes. We would never have reached as far alone. And, if we wouldn’t have done it together, we would have no one to share the amazing views from the top with!

Table of contents

Abbreviations

ABBREVIATION	MEANING	SHORT EXPLANATION
EEG	Electroencephalogram	A technique to record electrical signals arising from brain activity
EMG	Electromyography	A technique to record electrical signals arising from skeletal muscle activity
EOG	Electrooculography	A technique to record electrical signals arising from eye activity
ERP	Event-Related Potential	Small changes in recorded EEG as response to an event
HEOG	Horizontal Electrooculography	Electrodes placed at the side of each eye to measure horizontal eye movements
ICA	Independent Component Analysis	A technique to separate different components of a signal
NASA-RTLX	NASA Raw Task Load Index	A simplified version of a self-assessment questionnaire developed by NASA
VEOG	Vertical Electrooculography	Electrodes placed under and above one eye to measure vertical eye movements and blinks
PSD	Power Spectral Density	A signal's power content versus frequency

List of Figures

List of Tables

1 Introduction

Here we introduce this thesis work by presenting the background for why this research is needed. After that we give a short description of the method of this work together with our aims and limitation. We also present our research questions and how we contribute to the research field. Lastly, we give you as a reader a guide for the structure of this report.

1.1 Background

The loss of a limb would imply a major change in almost any lifestyle, and there are many dedicated scientists, engineers, doctors, therapists etc. around the world working to make this change as convenient as possible. One possibility is to use a prosthesis as a substitute for the lost limb. Most often these prostheses are strapped on with a socket and electrodes are attached to the remaining part of the limb [1]. The socket often chafe the skin and imply a lot of discomfort for the person wearing it [2]. These kind of prostheses are also unreliable and patients therefore often choose not to use them [1].

In 1990 the world's first *osseointegrated prosthesis* was implemented in Gothenburg, Sweden [3]. This means that the prosthesis is integrated in the bone of the amputee using a titanium rod. Apart from being a more secure and comfortable way of attaching the prosthesis than the conventional socket, this solution also opens up for the possibility to connect it to the muscles and nerves inside the remaining part of the limb. 29 years after the first osseointegrated prosthesis, in 2019, a Swedish man was the first in the world of getting an osseointegrated hand prosthesis with a *neuromuscular interface* [4], a so called e-OPRA [1]. This medical and technical achievement has been made possible by the collaboration between Chalmers Biomechatronics and Neurorehabilitation Laboratory (BNL), Centre for Advanced Reconstruction of Extremities at Sahlgrenska University Hospital and the company Integrum AB as part of their project "Natural Control of Artificial Limb Through an Osseointegrated Implant" [5]. By using the neuromuscular interface, the electrodes can pick up the signals from the muscles and nerves in the remaining part of the limb. That way, when the amputee execute the movement associated to move the hand, the hand moves [6]. By introducing *sensory feedback in the prosthesis, which reflects when pressure that is applied to the surface of the hand of the prostheses, the nerves in the limb are stimulated. This way* signals can also be sent from the hand to the brain and the brain can react to the stimuli given by the sensory feedback.

Before the implementation of the neuromuscular interface, prosthesis users had to rely only on visual feedback and could not feel how hard they pressed an object or even if they touched it at all [6]. With the addition of sensory feedback that gives the carrier a significantly better experience [4], Chalmers BNL hopes to further facilitate and improve the quality of life for people with amputated limbs or motor impairments.

Adding sensory feedback to a prosthesis intuitively seems to facilitate performing different tasks, like for example picking up a fragile object such as an egg. However, this needs to be investigated formally. One amputee who have received a prosthesis with sensory feedback have tried lifting a fragile object and they have broken or dropped the object less frequently with sensory feedback compared to when that feature is disconnected [7]. Also, it is possible that sensory feedback increase performance, but perhaps the effort is also increased. For this reason, there is a need for a quantitative and objective measurement of the mental effort, or *cognitive workload*, for performing a task, such as lifting an egg, with and without sensory feedback. Such a method could also be used to evaluate different stimulation paradigms, i.e. different ways to stimulate the nerves.

The first problem that arises when designing a method to measure cognitive workload is that there are currently only four people in Sweden with an implemented e-OPRA system [8]. This makes the sample size insufficient for reaching meaningful conclusions. Therefore, the conditions of lifting a fragile object with an e-OPRA prosthesis needs to be replicated with intact limb subjects as a complement. One way of doing this is to measure the cognitive workload for lifting the fragile object with your hand and compare this to when the sensory feedback is removed by using anesthesia on the hand and digits.

The second problem is that research that involve a physical intervention needs to be approved by the Swedish Ethical Review Authority [9]. However, this application normally takes 60 days to be approved [10] which makes this approach unsuited for the time limit of this project.

When measuring cognitive workload, there are several different options when choosing a method. These include pupil size measurements (e.g. [11]–[13]), heart rate variability (e.g. [14]–[16]) and breathing frequency (e.g. [16]). In this work, we will use two of the most common techniques to measure cognitive workload: *electroencephalogy* (EEG) using *event-related potentials* (ERP, e.g. [17]–[19]) and a self-assessment tool called NASA-RTLX (a task load index developed by the National Aeronautics and Space Administration [20], e.g. [15], [21], [22]) to adapt a method and test if that can be used to assess cognitive workload in this kind of task.

In 2019 a small study was made as part of Linn Berntssons master thesis at BNL. Her method was designed for testing amputees and were run with one amputee and compared three different conditions: no task, with sensory feedback and without sensory feedback. In the two last one the subject was instructed to lift a force sensitive cube back and forth over a small barrier. The first two conditions were also run with two intact limb subjects. The cognitive workload was evaluated using a combination of ERP measurements and the NASA-RTLX self-assessment tool. The results showed promise, but the method was tested with too few subjects to be able to draw any real conclusions. [23]

1.2 Brief description of this work

This report is part of a master's thesis at Chalmers University of Technology where both writers, Fanny Apelgren and Ida Pettersson, have studied Engineering Physics. We then moved on to a master's in Biomedical Engineering and Learning and Leadership, respectively. The thesis was written at the Department of Communication and Learning in Science (CLS) and the project was executed at Chalmers Biomechatronics and Neurorehabilitation Laboratory (BNL) at the Department of Electrical Engineering, under the Associate Professor Dr. Max Ortiz Catalán. The project has been supervised by Eva Lendaro (BNL) and Sheila Galt (CLS).

In this study, we measured event-related potentials (ERP) using EEG equipment for three different conditions, each recorded in three blocks. The participants performed a lifting task by moving a force sensitive cube back and forth over a small barrier at the same time as they performed a secondary task of listening to and counting sounds, known as an *oddball* *task*. The cube lit up when it was pressed too hard and the weight of the cube could be changed to vary the difficulty of the task between *easy* and *hard*. The participants were told that if the cube lights up, this indicates that the cube has been pressed to hard and “broke”, and that they should try to move the cube as many times as possible without “breaking” it. The third condition consists of merely the secondary task, i.e. counting sounds. This is called the *no task* condition. The EEG data were studied to examine if differences in event-related potentials components and the frequency bands could be seen.

During each condition the number of times the cube was lifted over the barrier and the number of times that it was “broken” were counted. The participant also reported the number of sounds that they counted in each block and filled out a self-assessment form, called the NASA-RTLX [20], to report the effort of each condition. The number of lifts and “breaks” per minute together with the difference between presented sound and counted sounds were studied as an indication of the subject’s performance, the result of the NASA-RTLX was used to measure perceived effort and the EEG data served as a quantitative measurement of the cognitive workload.

The different types of data and the experiment procedure that are used in this study were gathered from previous work on cognitive workload and recommendations from experienced scientists of the field. The force sensitive cube was designed with a few other similar models as an inspiration but was adapted for the criteria for this study. It has also been designed with the possibility for further development in mind, to enable use in other future studies.

1.3 Aims and limitations

The present study aims to develop and examine a method that can be used to measure the difference in cognitive workload with and without sensory feedback. The method by Linn Berntsson [23] have served as an inspiration, but we have mainly looked at other studies that have been tested with more subjects to develop our own improved methodology that is also adapted for intact limb subjects. Since anesthesia cannot be used without an ethical approval, we will test the method with other conditions to simulate the difference of with and without sensory feedback. Therefore, different levels of difficulty will be used as a substitute. The aim of this is to investigate how the variance in cognitive workload between the different levels of difficulty can be measured. We have also looked for signs of a learning process and the method has been tested using ten intact limb subjects. If the method can detect differences between different levels of difficulty, it could also be expected to be able to measure the difference between with and without sensory feedback, since these conditions are also believed to be different in difficulty. Therefore, the goal is for this study to serve as a pilot test in preparation for a future study where this methodology, or an adaption of it, will be used to investigate the difference in cognitive workload of performing a lifting task with and without sensory feedback, by using anesthesia.

With the limited timeframe of this work, we have done our best to process and analyze all the EEG data. However, there remains other ways to examine the data that has been recorded, this will be discussed further in the section about future work. Among other things, we will not examine the EEG data or the results from NASA-RTLX for different parts of each condition. Signs of a possible learning process within the conditions will only be examined by looking at the factors measuring performance.

1.4 Research questions

1) Will the subjects experience the expected difference in difficulty between the different conditions, on group and/or individual level, as indicated by…
   a) …the perceived effort, given by the scores on the NASA-RTLX?
   b) …the performance, given by number of lifts, success rate and accuracy of the oddball task?
2) Can the proposed method be used to measure differences in cognitive workload, on group and/or individual level as indicated by…
   a) …event-related potential (ERP) components? And if so, which components?
   b) …frequency bands? And if so, which frequency bands?
3) Which latency windows (for ERP components only) and electrode sites should be used to examine the differences in cognitive workload with ERP components and frequency bands?
4) Can a learning effect be observed during each condition by comparing the performance for each of the three blocks?

1.5 Contribution

The aim of this study is part of a larger goal, that ultimately comes down to improving the quality of life of amputees. By preparing for the future study with anaesthesia, this work is a step to provide quantitative evidence that adding sensory feedback in artificial limbs does lower the cognitive workload. This knowledge in turn will provide an incentive for the further development of prostheses. In addition to providing a method and a pilot study for this future study, this method might as mentioned also be used to evaluate different aspects of the prosthesis design, for example different stimulation paradigms.

This work has also contributed to the total knowledge about ERP experiments and discovered several conflicting opinions about the best procedures of the field. We have also discovered the lack of, and importance of, motivation and reasoning to explain why certain methods were chosen. Besides the results of this thesis the collected data could also be analysed further, and more aspects could be examined using for example ANOVA statistical analysis, which seems to be the most common procedure in ERP studies (e.g. [17], [18], [24]).

Studies using this or similar methods might also examine the learning process of receiving and learning to use a prosthesis. Even though the neuromuscular interface and the sensory feedback can be shown to decrease the cognitive workload, learning to live with a prosthesis will still demand practice and learning new strategies. The study of this progress can be an important step in the development of both prosthesis technology and the strategies used to teach someone to use a prosthesis.

1.6 Thesis outline

In the following chapters the thesis work will be described in further detail, starting with the theoretical background that lays the foundation of the work. After that follows the methods section where different parts of the method, from experiment procedures to data processing and analyses, are discussed. We describe possible approaches, discuss how they have been used in other studies and present how we have chosen to do and why. The results for the performance, perceived effort and EEG data are presented and thereafter discussed. Lastly, there is a conclusion and ideas for future work.

2 Theory

We will start by introducing the main concept of this work: *cognitive workload*. We will also give a background of the measurement methods: *electroencephalogram* (EEG) and *event-related potentials* (ERP). The latter will also be discussed further in the methods section (section 3). Here we will give a brief introduction of the technique together with how *ERP components* and *frequency bands* can be used to assess cognitive workload. We will then conclude the section by discussing how and why EEG measurements can differ between different individuals.

2.1 Cognitive workload

We are all aware that some tasks demand more cognitive resources than others. Most people have no trouble walking and talking at the same time, but when you are asked to solve some equations it might be harder to keep up an interesting conversation. This comes back to *attention* and *cognitive workload* (also known as mental workload or cognitive load).

There are different definitions to these, rather familiar, concepts and the relationship between attention and cognitive workload is also a matter of discussion. Rietschel et al. [19] states that “attention refers to the directed allocation of cognitive resources”. Similarly, Kantowitz [25] argues that cognitive workload is a subset of attention. Magill [26] elaborates this statement by saying that “attention refers to several characteristics associated with perceptual, cognitive, and motor activities” and that “a related view extends the notion of attention to the amount of cognitive effort we put into performing activities”. In this work there is no need to keep these interlaced concepts apart, so attention and cognitive workload will both be used in reference to *the cognitive resources demanded by a person to perform a certain activity or task*.

To get back to the question of why we can perform some tasks simultaneously while others cannot, we need to introduce what is known as *attentional reserve*, or attention capacity. This theory states that we have a certain amount of attention, or cognitive workload, and that this can be split to do several things. Each task demands some of the attention from our reserve and leaves the rest. In the example above, walking does not demand a lot of cognitive workload and leaves some attention that you can use for example for talking. Meanwhile, solving equations might not leave enough spare attention in the reserve for conversation, and perhaps walking and talking at the same time does not allow you follow a map to find your way in a new place.

That means that cognitive workload has an inverse relationship to the remaining resources of attentional reserve [27]. When the cognitive workload increases for a task, for example if you try to solve increasingly complex equations, the resources left for other tasks decrease.

Workload and attention seem to be closely related to performance and learning. Kantowitz [25] suggests a model where too low or too high workload leads to lower performance and this view is supported by Winnie et al. [28] who says that efficient learning happens at the optimal level of cognitive workload. As a further link to learning, Magill [4] suggests that a new task takes a lot of cognitive effort in the beginning, but that learning takes place and thereby the attentional demands decrease with practice. This is known as *the practice effect* and means that learning of a task can be indicated in difference ways. Either by a decrease of cognitive workload together with a stable level of performance, by an increase of performance with a stable level of cognitive workload, or by the combination of increased performance and decreased workload.

Something else that needs to be considered when looking at attention and cognitive workload is how it is balanced by the demands of the task at hand. If the challenge of the task is too low one will experience boredom, and frustration will emerge if the challenge is too much compared to the skill level. The area in between these two outer limits, where skill and challenge are perfectly matched, is usually called flow. This is the feeling that can make you keep up a task, for example a video game, for a long time. If you are bored or frustrated because the game is too easy or too hard you are likely to stop playing. So it is believed that both these conditions will decrease the attention of the task [29].

2.2 Electroencephalogram (EEG)

One way to measure cognitive workload that is commonly used is by the *electroencephalogram* (EEG). EEG is a clinical tool that measures the electrical activity of the cerebral cortex with electrodes attached to the human scalp. The cerebral cortex is the outermost layer of the cerebrum, which is the largest part of the brain, and is divided into left and right hemisphere. Each hemisphere is in turn divided into four lobes: frontal, temporal, parietal and occipital lobes, that are associated with different functions of the human body [30]. The brain and its different regions can be seen in Figure 1.

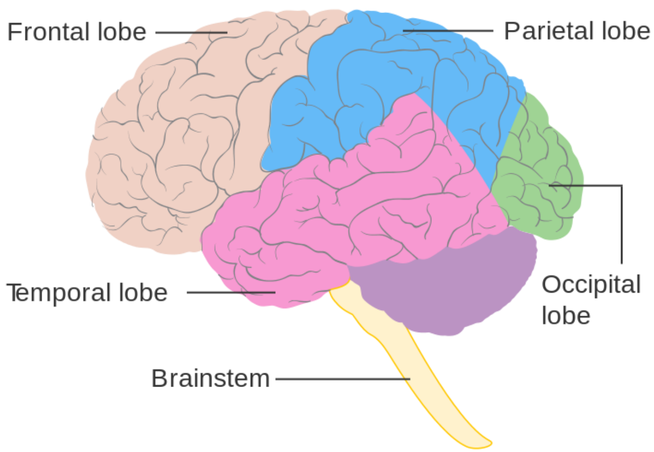


Figure 1. The brain and its four different lobes: frontal, parietal, occipital and temporal.

The electrodes that are used to measure the electrical activity of the brain usually consist of a metal disk or pellet. They can be attached to the head with stickers, but since the number of electrodes used for a measurement normally is more than 16, they are usually attached to a cap that can much easier be fitted to the subject's head. The electrodes pick up electrical activity from the brain in the form of electrical potentials for currents to flow from one electrode to a ground electrode. Since the recorded signals are in the range of 0 to 100 μV they typically need to be amplified by a factor of 1000-100000 before they are further processed [31].

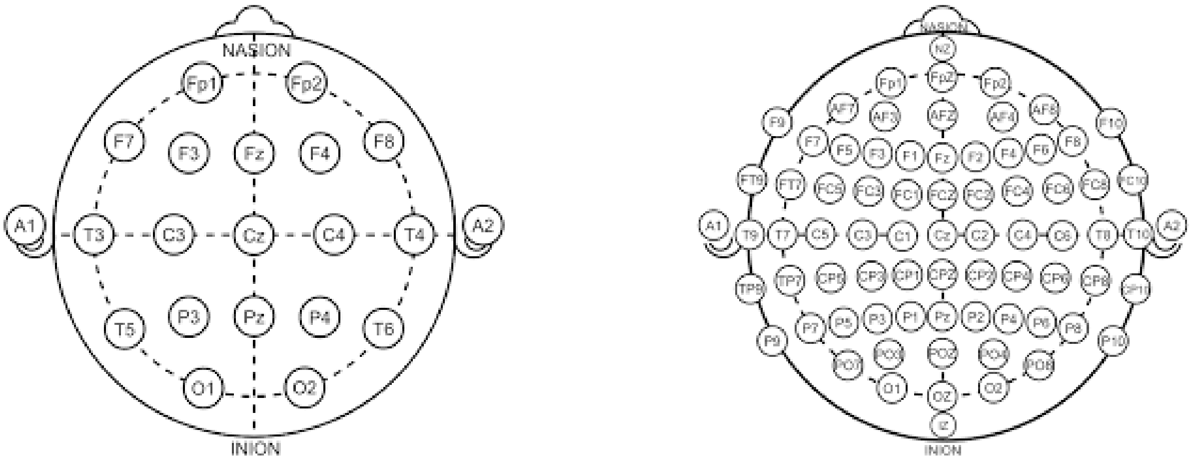
There are mainly four different characteristics for electrodes. They can be either wet or dry, and either passive or active [32]. Wet electrodes are generally Ag/AgCl electrodes, and one needs to put a conductive gel between the scalp and the electrode to get a good and stable electrical connection. This also helps lowering the impedance of the electrode-scalp connection. A lower impedance induces less noise which is important to get good quality of the measurement [31]. Dry electrodes instead consist of a single metal, often stainless steel, that act as a conductor and the electrode are put directly on the scalp. The difference between passive and active electrodes are that the active electrodes include a pre-amplification module after the conductive material. By that, the signal can be amplified before additional noise are introduced when the signal travels from the electrode to the system that measures the signal. This increases the signal to noise ratio. For passive electrodes there is no preamplification, which means that noise arising as the signals travel from the electrode to the measuring system will be amplified as much as the EEG signals. The different types of electrodes are combined, for example one can use active, wet electrodes [32].

2.2.1 Referencing

When creating an EEG amplifier, the ground electrode must be connected to a ground circuit for the EEG amplifier to work. This ground circuit is typically connected to other parts of the amplifier, which means that electrical noise is introduced at the site of the ground electrode. This means that there are noise present in the signal from the ground electrode that are not present in the signal from the other electrodes. To get rid of this noise, EEG recording systems use differential amplifiers. With the differential amplifier a reference electrode is used together with the operating electrode and the ground electrode to cancel out the noise. The differential amplifier records the potential between the operating electrode (O) and the ground electrode (G), as well as the potential between the reference electrode (R) and G. The amplifier then outputs the difference between these potentials O-G-(R-G) = O-R and since the noise from the ground circuit are the same for both O-G and R-G any noise generated at G will be eliminated in O-R. In other words, to get a single channel of EEG all three electrodes (operating, reference and ground) are needed. [31]

2.2.2 Electrode positioning

To get useful data that are comparable to other studies and possible to analyse it is important to position the electrodes in a correct way on the head. The most commonly used system to define the position of the electrodes is the 10-20-system [31]. Originally, this system used 21 electrodes, where two of them were placed on the earlobes and the rest were placed according to measurements of specific landmarks on the scalp. The landmarks used are the nasion (just above the nose, between the eyes), inion (the indent in the back of the head) and the left and right pre-auricular points (right in front of each ear), see Figure 2. An equator through the nasion, inion and the left and right pre-auricular points, together with a line between the nasion and inion and a line between the left and right pre-auricular points defines the measurements used to place the electrodes. The equator and the lines are then divided into sections with the first mark at 10 % and the following marks at 20 % intervals, resulting in the electrode positioning in Figure 2a. An extended 10-20-system with 128 electrodes can also be used, with marks on every 10 %, see Figure 2b [33]. Here we can also see that the electrodes are marked with letters and numbers. This is a way to indicate the location of the electrode. The latter gives the scalp region (F: frontal, T: temporal, C: central, P: parietal, O: occipital). The numbers indicate the distance from the center, where larger number are further from the central line. Even numbers are used for the right hemisphere and odd numbers for the left. The letter “z” stands for the number zero and is used instead of the number “0” to avoid confusion with the letter “O”.



a) 10-20 system

b) Extended 10-20 system

Figure 2. Electrode positioning according to the normal and extended 10-20 system. The letters stand for scalp region (F: frontal, T: temporal, C: central, P: parietal, O: occipital). The numbers represent the distance from the center, with even numbers to the right and odd numbers to the left. “z” stands for zero and is used instead of the number to avoid confusion with the letter “O”.

2.2.3 Artifacts and Noise

The electrodes do not only detect signals from activity in the brain, but also pick up other, non-neural, signals. Every time the subject moves, clenches their jaw, frowns, move their eyes, blink or something similar, this gives an electrical signal that can be picked up by the electrodes. Electrical signals from muscle movements are called electromyography (EMG). External sources, such as electrical equipment, can also emit electrical signals that can be picked up by the EEG electrodes. Another source for disturbance is the equipment itself, for example if the connection between an electrode and the scalp is instable.

All of these non-EEG, signals are called *artifacts* [31]. Some of the muscle movements, especially from the eyes since they are located close to the electrodes, can cause big disturbances of the recorded signal. Others, like electrical equipment or some muscle activations, are smaller and more regular. Both of these kinds of artifacts need to be handled to be able to see the subtle changes of the small, often below 100 µV, neural activity. How this can be done is discussed in section 3.4.5.

2.3 Event-Related potentials

Here we will briefly introduce the *event-related potential* (ERP) technique, which is the cornerstone of the method of this thesis. The different aspects of this technique will be discussed in further detail in section 3. We will also present the concept of *ERP components* and *frequency bands* as a way to measure cognitive workload. Lastly, we will discuss different reasons for why measurements of cognitive workload can differ between different individuals.

2.3.1 Basic concept

A common technique when measuring EEG is to use *event-related potentials* (ERPs). This is a way to single out certain activities in the brain. Raw EEG data is often hard to use, since it is a mix of all the neural activities in the brain. Even if you are told to focus on a certain task, your mind easily wanders. The ERP technique was first used in 1977 by Wickens et al. [34] and builds upon the idea that a certain stimulus, or event, can trigger a specific brain activity.

The ERPs are measured by presenting some kind of stimuli, for example sounds or flashes of light, repeatedly while measuring EEG. Each stimulus is time-locked to the EEG data and marked by a line at the appropriate time. Later, a short section of EEG data, a so called epoch, is extracted around every stimulus. The epoch begins a short time before the stimulus and ends a certain time after the stimulus. It is common to use 100-200 milliseconds pre-stimulus and 800-1000 milliseconds post-stimulus. The idea is that noise that is unrelated to the stimulus will cancel out when many epochs are averaged together and leave the EEG signals that are related to the stimuli.

The book “An Introduction to the Event-Related Potential Technique” by Steven J. Luck [31] is a commonly used reference in this work. This, together with articles using the ERP technique, has helped us make all of the decisions involved in conducting an ERP experiment.

2.3.2 ERP Components

Luck [29, p. 68] gives the following definition of *ERP components*:

“An ERP component can be operationally defined as a set of voltage changes that are consistent with a single neural generator site and that systematically vary in amplitude across conditions, time, individuals, and so forth. That is, an ERP component is a source of systematic and reliable variability in an ERP data set.”

These voltage changes can then be picked up by the EEG electrodes, with different weights depending on the relative location of the source and each electrode.

Here it is important to note the difference between ERP *components* and ERP *peaks*. The peaks in the ERP does also show voltage changes, but these changes do not necessarily reflect changes in a given component. For example, if the voltage of a positive peak is reduced it might reflect a reduction of an underlying positive component, but it might also be an increase of a negative component at the same latency, i.e. at the same time compared to the stimulus. There are some techniques to extract the components from the data, a common one being *independent component analysis* (ICA) that will be used and discussed in this work (see section 3.4.5.2). However, none of these methods can be completely trusted, and should be used with caution [31].

So, a single peak can never be assumed to represent a single component. Nevertheless, one can look at many different electrodes and study the latency of a peak. Since the time for a signal to travel the different distances from the source to each electrode can be closely estimated to be equal, the timing will coincide for one component at different electrodes.

To avoid having to investigate all electrode sites (since they can be many), and still be able to draw conclusions from an ERP waveform, Luck recommends to use the components that have been shown useful in earlier studies, either from other similar experiments or, if you are first in your field, from other fields [31].

As a way to facilitate discussions about ERP and comparisons between studies there is a conventional method for naming the different peaks of an averaged ERP waveform. These names start with a letter, either N or P, to denote whether a peak is positive or negative. After that follows a number, describing one of two things. In the first convention the number describes the ordinal position of the specific peak, i.e. the first positive peak would be called P1 and the third negative peak N3. This convention is depicted in Figure 3. However, this plot uses the old convention of plotting ERP waveforms with the negative axis directed upwards. In this work we will use the same naming convention but with the positive axis upwards, as is common in most modern ERP studies [31]. The other possible way is to name the peak according to latency (i.e. the time after stimulus onset), so that a positive peak occurring around 300 ms after the stimulus onset would be called P300. In some cases, peaks are also named to describe their function or location, such as the *error related negativity* (when the subject discovers that he or she did something wrong) or *late positive potential.* [31]

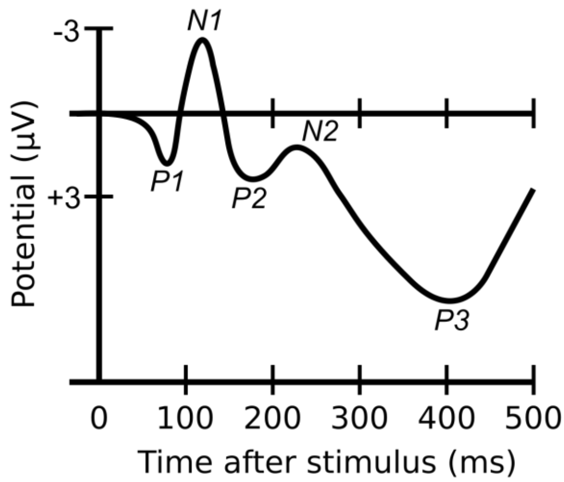


Figure 3. Depicting an example of an ERP waveform where each peak is named after the convention used in this thesis. The letter (P or N) stands for positive or negative (although note that negative is upwards in this plot) and the number stands for the peaks' ordinal position.

As mentioned, each component can be referenced to either by using the ordinal position of the peak (e.g. N1) or the latency (e.g. P200). The latter describes the latency at which the component is usually found, but this varies between different experiments and therefore this notion can be confusing. Luckily, the latency is often about 100 times the ordinal position, so that P1~P100, N2~N200 and so on [31]. However, some old conventions linger and P3 is still often referred to as P300 because it was first found about 300 ms post stimulus even though it is more common to arise later than that [31]. The latency also tells us something about the stage of the stimulus processing by the brain. That means that earlier components arise from perceptual processing in the brain while later components reflect later stages of the reaction, including evaluation of the stimulus [17]. Here we describe some components that have been shown to be an indication of cognitive workload, that were the most common in our literature research. We will use the naming convention based on ordinal position.

2.3.2.1 N1

The *N1 component* is specific to auditory stimuli and is characterized as one of the initial components in an auditory ERP, called long-latency auditory ERP components. It has been suggested that the N1 component signal the detection of acoustic change in the environment. The single-peak N1 component is evoked by short transient stimuli or by onsets of noise and has been shown to consist of three temporally overlapping constituents. The dominant contribution to the N1 component is most prominent at the fronto-central electrodes. [35]

The N1 component has been linked to cognitive workload in several studies [18], [36]–[39]. One of the examined studies could show that N1 varied between some of the levels but not all [24] and two failed to show significance for N1 [17], [19]. In these studies, N1 was found between 75 and 180 ms post-stimulus, where the studies that were successful of linking N1 to cognitive workload seems to have found it in the later region of that interval, see Figure 4.

2.3.2.2 N2

The *N2 component* is known for containing several different subcomponents: N2a, N2b, N2c. However, the basic N2 component (that will be discussed here) is said to be elicited by a repetitive, nontarget stimulus and it gets a larger amplitude if the stimulus is novel (not repeated). Depending on if the stimulus is task-relevant or task-irrelevant the N2 component appears with different latency, with later latency if the stimulus is task-relevant (the difference between task- relevant and irrelevant stimuli will be discussed more in 3.2). Also, if the stimulus is auditory a larger effect is seen in the central sites and if the stimulus is visual the effect shifts to be larger in the posterior sites instead. [31]

The N2 component has been examined in two of the studies that we have looked at [36], [37] and both showed that it successfully assess the cognitive workload. They found N2 in the interval 200 to 400 ms post-stimulus, see Figure 4.

2.3.2.3 P2

The *P2 component* is most prominent at the frontal and central scalp sites and is typically larger for stimuli containing simple, infrequent target features. At posterior sites, the P2 component often interferes with N1, N2 and P3 and therefore it is hard to distinguish at posterior sites. [31]

Two of the studies in our literature study could show a significant correlation between P2 and different levels [17], [18], but three other could not verify this correlation [19], [24], [37]. P2 was found between 166 and 270 ms post stimulus, see Figure 4.

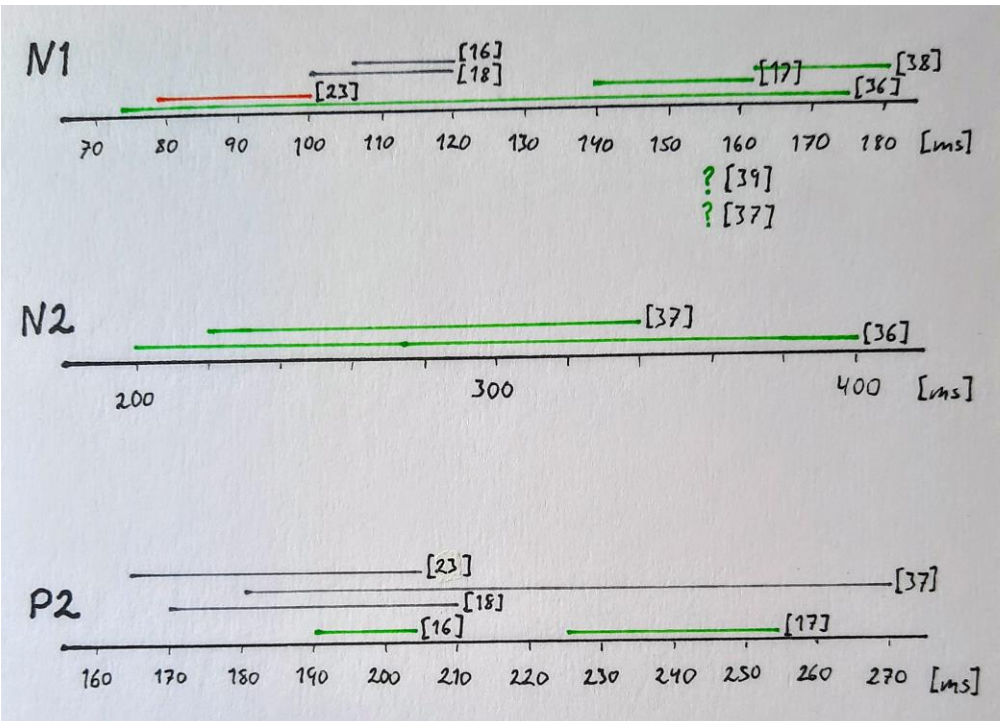


Figure 4. A sketch of the latency ranges where the different ERP components have been found according to our literature study. Each line is marked with the reference for the study it is taken from. Green lines indicate that there has been a significant difference between the different levels of difficulty. Orange means that a difference could only be seen between some of the tested levels, but not all. Grey lines mean that no significant differences could be shown. Studies that have not specified the latency range are marked with "?".

2.3.2.4 P3

The *P3 component* is the most examined ERP component when it comes to cognitive workload [40], and that shows in our literature study. There are also several other components that are closely related to P3, and sometimes hard to differentiate from it. The ones examined in the studies we have read are *novelty P3*, *P3a*, P3b and *early* and *late P3a*.

The P3 component is typically evoked by rare task-relevant events and it is said to reflect an updating of the context information, which often is assumed as an update of the working memory. There is also clear evidence that the amplitude of the P3 component can be influenced by the amount of attention allocated to a stimulus, which has been most clearly observed in dual-task experiments where the subject is to perform two tasks at the same time. The latency of the P3 component changes over the scalp and is shorter over the frontal areas and longer over the parietal areas. It also differs between individuals depending on how rapidly the subject can allocate their attentional resources, such that the latency is shorter for subjects with higher mental speed. [35]

As mentioned above, the P3 component can be divided into several subcomponents: mainly the P3a, P3b and novelty P3 component. These subcomponents are typically elicited by different task conditions and can be recognized by their different topographic distributions. The P3a subcomponent has a centro-parietal maximum amplitude distribution and is elicited by rare tones presented in a series of frequent tones without a task. If novel distracters (such as a dog barking) are used in a sequence of frequent tones a fronto-central P3 potential is elicited, which is called the novelty P3. The P3b subcomponent is elicited by task-relevant stimuli and has a parietal maximum amplitude distribution. Often, P3b and the classic P300 (P3) are said to be the same component. It is also found that the novelty P3 differs from the classic P300 component and that the P3a and novelty P3 are most likely variants of the same ERP that varies in scalp topography depending on attentional and task demands. [35]

Many studies have linked the P3 component and it's relatives to cognitive workload [17], [18], [34], [37], [39]–[42]. One study failed to show significance for P3 [29] and one could only show significance between some of the levels [36]. In the successful cases, P3 has been found in the interval between 270 and 517 ms post-stimulus, and more commonly in the earlier part of that interval, see Figure 5.

When looking at the related *novelty P3*, it has also been proven successful in assessing cognitive workload by several studies [19], [24], [43]. One has only shown a change of amplitude between some of the levels examined [44]. The novelty P3 component has been observed between 250 and 332 ms post-stimulus, see Figure 5.

Lastly, some studies have linked P3a to cognitive workload [38], [45], where one split the component into early and late P3a. Another study saw no correlation between different levels and neither P3a nor P3b [43]. The components were found somewhere in the range between 210 and 405 ms post-stimuli, see Figure 5.

2.3.2.5 LPP

The *late positive potential* (LPP) is commonly identified as a midline centro-parietal ERP with a strong connection to emotional stimuli such as pleasant and unpleasant pictures. It becomes evident at 300 ms, and can therefore be mistaken for the P3 component, but the LPP component often continues for latencies up to 2000 ms, even though it is maximal in the latency range of 300-1000 ms. LPP has also been shown to indicate reaction time to a stimulus by that the LPP amplitude increases when the reaction time increases. [35]

The LPP component has been shown to be an indicator of cognitive workload [17], [18], [21]. The findings have been within the interval 400 to 610 ms post stimulus, but two of these three studies found LPP close to the end of this interval, see Figure 5.

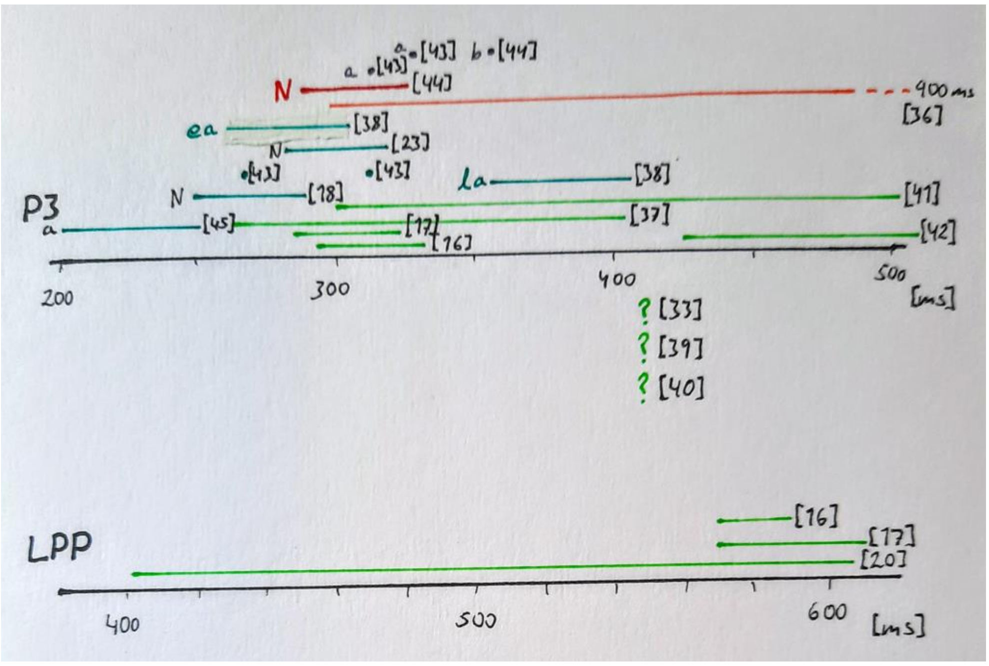


Figure 5. As the previous figure, this is a sketch of the latencies where each component has been found, and each study is marked with its reference number. This figure includes the same color coding as the previous one (green: significance, orange: partly significance and grey: no significance), but here darker colors are used to indicate different versions of P3. This is also indicated by letters where "a" is P3a, "b" is P3b, "N" is Novelty P3, "ea" is early P3a and "la" is late P3a. A dot indicates that no interval was given, only the latency of the peak. As before, "?" denotes studies where the latency has not been specified.

2.3.3 Frequency bands

The measured EEG signals often have an oscillatory, repetitive behaviour and therefore the collective electrical activity of the cerebral cortex is often called a rhythm. The EEG rhythms diverse between individuals and depends on things like the mental state of the subject, if they are awake or sleeping for example. Since the electrical activity arises from the activation of neurons in the brain, the rhythms can have different frequency depending on how synchronous the activated neurons are. The frequency range for the rhythms is approximately between 0.5 and 30-40 Hz and are often divided into five frequency bands, *Delta*, *Theta*, *Alpha*, *Beta* and *Gamma* [30]. The Alpha band is also sometimes subdivided into *Low-* and *High-Alpha*. The ranges of each band differ slightly between different studies. However, the differences in how the frequency bands are defined are relatively small (around 1 Hz). So, in this work we will discuss previous findings about a certain frequency band, such as Alpha, without consideration about the fact that the studies have used slightly different definitions of Alpha. We have decided to use the same ranges as was used by Rietschel et al. [46], which are presented in Table 1. Now we will present each of these frequency bands and their connection to cognitive workload, as shown by other studies that are part of the literature study of this work. We also describe the quotient *Theta/Alpha*.

Table 1. EEG frequency bands [46].

EEG FREQUENCY BANDS	
DELTA RHYTHM	<3 Hz
THETA RHYTHM	3-8 Hz
ALPHA RHYTHM	8-13 Hz
LOW-ALPHA	8-10
HIGH-ALPHA	10-13
BETA RHYTHM	13-30 Hz
GAMMA RHYTHM	>30 Hz

2.3.3.1 Delta

The *Delta rhythm* has a large amplitude and is mostly present during deep sleep. In normal adults it is normally not observed in the awake state other than that it is indicative of cerebral damage or brain disease[30]. It has also been shown that Delta rhythms are involved in motivational processes such as the necessity to satisfy the basic biological needs. [47]

Our literature study has shown that there seems to be no significant correlation between the Delta frequency band and cognitive workload. We found two studies that measured Delta in tasks of varying difficulty, but neither saw any significant results [46], [48].

2.3.3.2 Theta

The *Theta rhythm* mostly occurs during drowsiness and certain stages of sleep [31], but it has also been shown to correlate with a variety of behavioural, cognitive and emotional variables. The main domain seems to be memory and emotional regulations, but there are also indications that Theta activity occurs when performance of a learned task is increasing most rapidly and that it declines as tasks becomes familiar [47]. Especially at frontal scalp sites Theta activity can be facilitated by emotions, focused concentration and during mental tasks [49], meaning that it is expected to increase with increasing workload.

Theta is, together with Alpha (described below), the frequency band that has been shown to relate most to cognitive workload [40]. Several studies have shown that theta can show the difference in cognitive workload between different levels [11], [21], [40], [50]. However, our literature study has also shown that several studies have failed to show this correlation [14], [29], [45], [46], [48] and a few studies have seen statistical significance for theta between some levels, but not between all [24], [44]. This can for example mean that there is a difference between the easy condition compared to the medium and hard, but that no difference can be seen between the two latter conditions.

2.3.3.3 Alpha

The *Alpha rhythm* occurs during wakefulness over the posterior regions of the head and does normally have higher amplitude over the occipital areas. It is typically characterized by rounded or sinusoidal waveforms. The amplitude varies between individuals and in a given individual also from time to time but is normally below 50 µV in adults. It is commonly blocked or attenuated by attention and mental effort, especially visual attention, and are most prominent when the eyes are closed [51]. The amplitude of the Alpha frequency band is therefore expected to decrease with increasing workload.

As mentioned, Alpha and Theta has been shown to indicate cognitive workload [40]. Alpha is also the most studied frequency band in the literature that we have studied for this work, sometimes split up into sub-bands Low- and High-Alpha. Several studies have seen a significant difference in Alpha between different levels of difficulty [11], [14], [21], [29], [40]. One of the studies have, however, failed to show a significant effect [50]. When comparing High- and Low-Alpha, the upper frequency range seems to often yield significance [24], [44], [46], [48] while the lower range often only can show difference between some of the levels [24], [44].

2.3.3.4 Beta

Activity recognized as *Beta rhythm* are mainly found over the frontal and central regions of the head and is found in almost every healthy adult. The amplitude does normally not exceed 30 µV and it can be blocked by motor activity and tactile stimulation [51]. Beta activity normally increases with drowsiness and light sleep and also with mental activation [49].

In our literature study, there has been little evidence of a correlation between Beta and cognitive workload. Most studies that have examined beta have not been able to show a significant effect [14], [29], [46], [48], [50] while one has seen a difference only between some of the conditions [24].

2.3.3.5 Gamma

The *Gamma rhythm* consists of high-frequency oscillations and are said to be related to a state of active information processing [30]. Induced Gamma activity have been reported during sensory, cognitive and motor processing and may be related to sensory binding as well as sensorimotor integration [51].

One of the studies that we have read have seen evidence of a significant difference between different levels for the Gamma frequency band [46]. One study has seen effects between some of the conditions but not all [24]. However, two studies have also failed to show a correlation between gamma and cognitive workload [50], [52].

2.3.3.6 Theta/Alpha:

Besides the frequency bands, the quotient *Theta/Alpha* is also commonly used when assessing cognitive workload. There are several ways of calculating this ratio, often by using either frontal or parietal (see Figure 1) electrodes when measuring Alpha and Theta. Frontal Thetha/parietal Alpha [44] and frontal Theta/frontal Alpha [24] has both been used to indicate cognitive workload. Another study performed by Gentili et al. [45] showed that the Theta/Alpha ratio could be calculated from electrodes in the same area and still show significantly higher values for a higher level of difficulty.

2.3.4 Differences between individuals

As mentioned, cognitive workload is here defined as *the cognitive resources demanded by a person to perform a certain activity or task*. This means that the cognitive workload is not only correlated to the difficulty of the task, but also to the abilities of the individual. When the task demands are close to exceeding a person's ability, the workload is high, and the limits to boredom and frustration depend on both the task and the individual skill level.

Apart from this, ERP measurements also varies between individuals. Differences between different subjects can reflect biological differences such as skull thickness or cortical folding patterns [31]. Other factors that can affect the ERPs when measuring cognitive workload are age, lack of sleep, time-of-day, time since the last meal, time of year and geographic location (mainly because of difference in daylight), exercise (mainly affects older people), and the intake of common drugs such as caffeine, nicotine and alcohol [53].

3 Method

Here we present how we have constructed our method, by describing general theory for the different parts and discussing how others have chosen to do.

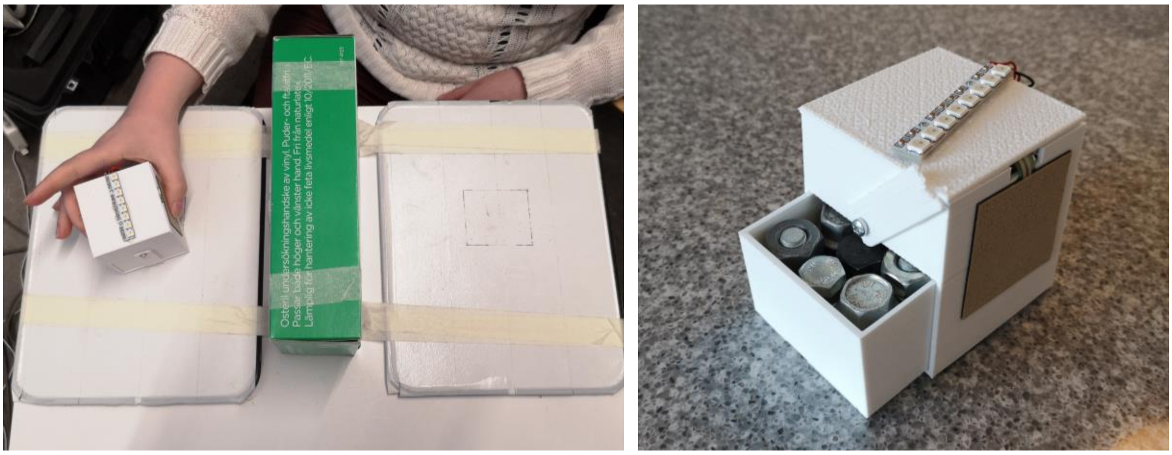
3.1 Grasping task

As mentioned in section 1.1, this work is a pilot study in preparation for measuring cognitive workload on intact limb subjects performing a grasping task with and without sensory feedback, where the latter condition will be done by using anesthesia on their hands and digits. Further, the conditions of this study are meant to mimic the conditions of with and without sensory feedback in prosthetic hands. A graphic illustration of the connection between the easy and hard condition of the different studies can be found in Table 2.

Table 2. A schematic illustration of how the different levels are meant to be represented in our study and the future studies with anesthesia and prosthetic hands, respectively.

	OUR STUDY	STUDY WITH ANESTHESIA	PROSTHETIC STUDY
EASY TASK	Lighter cube	Without anesthesia	With sensory feedback
HARD TASK	Heavier cube	With anesthesia	Without sensory feedback

So, the main task to be examined in this work is a grasping task. This is performed by lifting a force sensitive cube (described more in section 3.7) back and forth over a barrier as many times as possible, without pressing it too hard i.e. breaking it. If the cube is pressed too hard it is indicated by that a red LED bar light up. The weight of the cube can be increased by adding extra weights to the cube, in order to make it harder to lift it without pressing it too hard. In that manner there are two different difficulties for the grasping task: *easy* and *hard*. These are meant to represent the different conditions of *with* and *without sensory feedback,* that will be used in the future study with anaesthesia that this work is in preparation for. Pictures of the experimental setup and the force sensitive cube can be seen in Figure 6. The force sensitive cube and its design process is further described in section 3.7. The grasping task is comparable to the modified Box and Blocks test, i.e. the Virtual Eggs Test developed by Clemente et al. [54].



a) The setup for the grasping task, with two boards separated with a barrier. The cube was to be lifted back and forth over the barrier.

b) The force sensitive cube, with force sensors, LED bar and weights. The weights could be removed to reduce the difficulty of the grasping task.

Figure 6. The experimental setup for the grasping task, with a closeup of the force sensitive cube.

3.2 Dual-task paradigms and oddball tasks

Studies using the ERP (event related potential) technique are commonly performed by measuring ERPs of a secondary task that is performed simultaneously with a primary task of interest. This design is needed when it is not possible to directly assess the workload of the primary task, for example if there are no clear stimuli. The subjects are to primarily perform the primary task as well as possible and use remaining cognitive resources for the second task, doing it as well as possible under the circumstances. The secondary task in an ERP study can be for example to see flashes of light while performing a primary task of solving equations. Using ERP, the brain potentials related to the stimuli are measured. This way, the brain's responses to the secondary task stimuli are expected to decrease as the difficulty of the primary task increases, and this shows by a decrease in amplitude of the different ERP components presented in section 2.3.2. The ERP technique thereby uses the inverse relationship between cognitive workload and attentional reserve, mentioned in section 2.1. This use of two simultaneous tasks, a primary and a secondary task, to measure the cognitive workload of the primary task is called a *dual-task paradigm*. In some studies, the subjects are required to react to the stimuli in some way, for example by pressing a button or by silently counting, while other studies tell the subjects to ignore the stimuli.

A common dual-task paradigm is what is called an *oddball task*. Here, the stimuli contain common non-targets and rare targets, differentiated by for example pitch or colour. It is usual that the common non target represent 80 % of the stimuli. The ERPs are measured around the rare targets, since several ERP components are larger for a stimulus from a rare category than a common. The stimuli are usually either visual, auditory or somatosensory.

The dual-task paradigm is widely used, but also questioned. One argument is that adding a secondary task will affect the performance of the first task, and thereby change the variable under investigation [18]. To deal with this problem, it is often recommended to use task irrelevant stimuli, i.e. stimuli that the subject should ignore [25]. However, Castellar et al. [55] examined this and could not find evidence that the primary task, in this case a game, was affected by the secondary task of reacting to target sounds as fast as possible by pressing a button.

When applying a dual task paradigm, there are many different factors to consider. These include deciding if the subjects should ignore or react to the stimuli, what type of stimuli to use and the timing of the stimuli. We will continue by discussing these options.

3.2.1 The choice of stimuli

As mentioned, stimuli can be either visual, auditory or somatosensory. Since the primary task of this work (lifting a force sensitive cube) involves using visual and sensory feedback, we have chosen to use auditory feedback for the secondary oddball task. This so that the secondary task should interfere with the primary task as little as possible.

When using auditory stimuli, an approach that has become common is the *novelty oddball task*, which include *novel, complex* sounds (e.g. [18], [19], [39], [43], [45], [55]). This means a collection of complex sounds (e.g. a dog barking or a car honking) that are not repeated within each subject. When comparing different kinds of auditory stimuli, Dyke et al. [38] showed that complex sounds were better for measuring cognitive workload than simple sounds (e.g. a tone of a certain frequency). They could, however, not see any difference between if the sounds were repeated or not.

In novelty oddball studies, it is common to use 80 % common, simple sounds (e.g. a low pitch tone), 10 % rare, simple sounds (e.g. a high pitch tone) and 10 % novel, complex sounds (e.g. a person coughing or a mosquito buzzing) (e.g. [39], [43], [55]). The ERPs are usually measured around the novel, complex sound since this is a better way to elicit ERP components [38] and these sounds are most often task-irrelevant by either having the subjects react to the rare, simple sounds by pressing a button or count them (e.g. [39], [55]), or by asking the subjects to ignore all sound and only focus on the primary task (e.g. [14], [18], [19], [45]). This means that the novel, complex sounds are used for the ERP measurement but are not relevant for any of the tasks. However, according to a study made by Debener et al. [43] task irrelevance is not necessary when applying the novelty oddball task. They also found that the novelty P3 was actually larger for task relevant sounds. That is to say that it was more effective to let the subjects count the novel sounds, that were also used for ERPs, than to count the rare, task-irrelevant sounds. This is evidence against the common view, and all other studies that we have looked at, both before and after Debener's discovery, still use task irrelevant stimuli when measuring ERPs.

For this study, we apply the novelty oddball task using 80 % common, simple sounds, 10 % rare, simple sounds and 10 % novel, complex sounds, as described above. Henceforth, these sounds will be referred to as *common*, *rare* and *novel*, respectively. We choose to use 500 Hz as common sounds and 1500 Hz as rare sounds, since these sounds represented the broadest range of frequencies we could use that were deemed comfortable to listen to for the subjects. The novel, complex sounds were randomly chosen from 93 different audio clips and were only played once during each condition. The ERPs were measured by using the novel sounds, as recommended above [38] and the subjects were asked to count the rare sounds. This choice was made against what was shown by Debener et al. [43], since we decided to rather use the common approach of using task-irrelevant stimuli to measure ERPs. This will make it easier to compare the results of this study to others.

3.2.2 Secondary task: Counting, reacting or ignoring?

If the subjects are instructed to count sounds, this can also be used as an indication of cognitive workload. Since a harder task should decrease the attention available for counting, more errors should be made with a harder task than an easy one. However, Luck [31] raises an issue with this method. Since error could arise from missing a target, from mistaking a nontarget as a target or from losing count, it is impossible to tell if a correct number means that no error was made. For example, the combination of the two first errors would result in a correct number of counted targets. For this reason, the alternative of pressing a button in reaction of a target can be superior, since that allows both misses and false pushes to be considered. A problem with this approach is that it requires a movement that will result in artifacts.

We decided to ask our subjects to count the rare, simple sounds of the oddball task. This choice was made to avoid subjects getting bored of the task, since the primary task is very repetitive. This because boredom might affect the willingness to focus your attention on a task, as discussed in section 2.1.

3.2.3 Stimuli timing

If the interstimulus interval, i.e. the time between two stimuli, is too short there is a risk that the different epochs will overlap, which will mean that potentials resulting from one sound might affect the next one. On one hand, you want the interval to be as short as possible to maximize the number of epochs to draw data from. On the other hand, the ERP components are bigger the longer the interstimulus interval and if the stimulus are played too often that might be tiering for the subject. Also, if the interval is too long, so called *stimulus-preceding negativity* can occur, which means that the subject is anticipating a sound. This can be confusing when analysing the results. Luck recommends around 1000 ms interstimulus interval, and to use a temporal jitter of at least ±100 ms, since varying the interval also prevents stimulus-preceding negativity. This means that the interstimulus interval could vary randomly between 900 and 1100 ms, according to Luck. Also, since the epochs around each novel sound will later be averaged together (for more details, see 3.4.6), varying the interstimulus interval also helps to prevent regular noise, such as alpha waves, to show in the averaged ERP waveforms. When it comes to the duration of stimuli, Luck recommends 50-100 ms for simple sounds and 300-400 ms for novel sounds, with 5-20 ms rise and fall time. [31]

It seems like most of the earlier studies that have applied the novelty oddball task (e.g. [13], [16]–[18], [23], [37], [42], [44], [54]) have, however, used a longer time for the duration of the simple sounds. They have all used the same sounds, originally from[56], and to facilitate comparing our result to theirs we have decided to use the same source for our sounds.

Therefore, we have played the sounds in random order with a varied interstimulus interval between 960 and 1360 ms, as it follows Lucks recommendation and has been used by for example Debener [43] and Castellar Núñez [55]. The novel sounds are from the work of Fabiani et al. [56] and the duration is between 159 and 399 ms (mean 335,43 ms). The pure tones from the same source was 336 ms long, and as mentioned we chose to use 500 Hz for the frequent sounds and 1500 Hz for the rare. Rise and fall time are 10 ms for the pure tones, but vary for the novel, depending on their properties.

3.3 Experimental procedure

This thesis mainly resulted in a developed method to measure cognitive workload, that consists of the different parts described above: a dual-task paradigm consisting of an oddball task and a grasping task, where the cognitive workload is evaluated through EEG measurements and the self-assessment questionnaire NASA-RTLX. To test if the method could be used to measure cognitive workload, we performed a pilot study including 10 subjects.

3.3.1 Participants

A good average ERP waveform can be obtained either by using long trials or many trials. However, the long preparation time for each subject (about one hour with the subjects for our experiment) makes it unrealistic to examine many subjects. Normally each study uses about 10-20 subjects for ERP measurements [31]. We have measured ERPs for 10 subjects.

The 10 participants (six females and four males) was students at Chalmers University of Technology in the age 24 to 28, with mean age 25.5 and standard deviation 1.43. All had normal or corrected to normal vision and hearing. The subjects' handedness was evaluated through the *Waterloo Handedness Questionnaire*. This is made up of a series of questions of which hand one would use for performing certain tasks. The options were left always, left usually, both equally often, right usually and right always. The score is then added by assigning the options with values -2, -1, 0, 1, and 2, respectively. The score ranges from ±72 and this score would thereby indicate a strong preference for either the left (-72) or the right (+72) hand. According to this all subjects were right-handed, with a received score in the range 41 to 56, with mean score 49.2 and standard deviation 4.38. The subjects also read and signed an informed consent, which can be found in Appendix C, before the experiment.

To measure EEG we used an EEG system environment from *g.Tec Medical Engineering*, including *g.HIamp multi-channel biosignal amplifier* for 144 channels, *g.GAMMA EEG cap* with 128 *g.SCARABEO active Ag-AgCl electrodes* and *g.TRIGbox trigger pulse box*. The software used to collect the EEG data was *g.RECORDER*, a biosignal recording system from *g.Tec*. EEG was recorded at 2400 Hz from 128 electrodes according to the extended 10-20 system, which is described in section 2.2.2 and can be seen in Figure 2. Included in these 128 electrodes are four eye electrodes (EOG), two electrodes put on each earlobe and 122 scalp electrodes. As ground electrode we used the AFz electrode (between Fp and F in Figure 2). No online reference was used, instead the data was referenced offline. For a picture of a subject fitted with the cap and electrodes, see Figure 7.

The trigger pulse box was used to time-lock the audio stimuli from the oddball task described above in section 3.2, which was played for the subjects through in-ear headphones. Headphones were used so that the subjects should hear equally in both ears, compared to if speakers had been used where there is a risk that the speaker sound is heard differently in the ears. For localizing and digitizing the exact individual position of the electrodes in 3D for all subject we used *Polaris Krios System* from *Northern Digital Inc*.

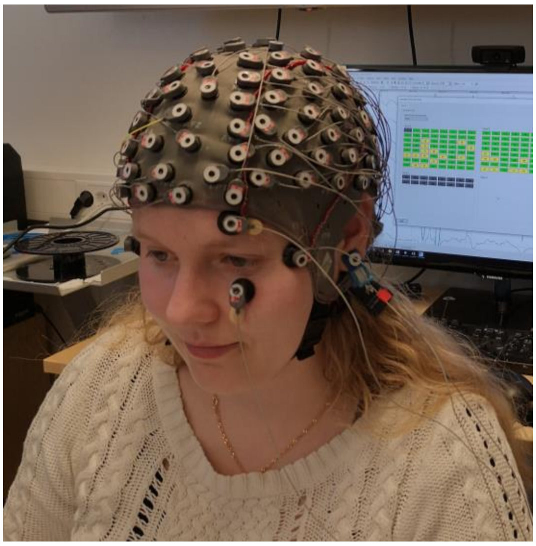


Figure 7. Person fitted with EEG cap with 128 electrodes. You can also see the EOG electrodes around the eyes.

3.3.2 Tasks

The developed method consists of three different conditions: *no task*, *easy task* and *hard task*, using combinations of the grasping task and the oddball task, described in section 3.1 and 3.2, respectively. *No task* means that the subject performs only the auditory oddball task while focusing their gaze at a plus sign on a computer screen. For the *easy* and *hard task* conditions the subject was to perform the grasping task at the same time as the auditory oddball task. The *easy* and *hard* condition represent a dual task paradigm, described in section 3.2, where the grasping task is the primary task and the oddball task is the secondary task. The *easy* and *hard* conditions are, as mentioned in section 3.1, meant to replicate the conditions of *with* and *without sensory feedback* that will be used in the future study with anaesthesia. The *no task* condition is to add another level of cognitive workload that can be used as a baseline to examine if we can measure the differences in cognitive workload between different levels.

3.3.3 Self-assessment using NASA-RTLX

To get an indication of how much workload the subjects themselves thought they put into each task, we used a self-assessment questionnaire, or task load index, developed by the *National Aeronautics and Space Administration* called *NASA-TLX* [57]. The NASA-TLX is commonly used to assess perceived effort (e.g. [14], [21], [39], [44], [45]). The questionnaire consists of six subscales which represent the variables: *mental*, *physical*, and *temporal demands*, *frustration*, *effort*, and *performance*. Each subscale is a twenty-step scale from 0 to 100 and the subjects were asked to put a cross on the step of each subscale that best represented their effort on each task. For each task, the values for all subscales were added together and divided by six to get the averaged NASA Raw Task Load Index (NASA-RTLX). This index is more commonly used in many studies because it is simpler to apply compared to the NASA-TLX which also includes an additional weighting process to weight the different subscales against each other [20].

We used the NASA-RTLX for each task and subject to get an indication of whether the perceived workload differed between the tasks. This will be used as an indicator to see if the subjects experienced the expected difference between the different conditions. Subjects were also told to mark the different blocks in each condition with 1, 2, or 3 if they experience a difference in effort. They were also told that they could mark with an X if they estimated the same value for all blocks. In the end, most of the subjects did not experience a difference between the blocks, so only conditions were examined. For subjects who marked a difference, we have used the mean value.

3.3.4 Procedure

Before the experiment, the subject was fitted with the EEG head cap and a connection was made between each electrode and the scalp using a conductive gel. Then the EOG electrodes were attached around the eyes using adhesive labels, and the reference electrodes, used for offline referencing, were clipped to the earlobes. The impedance for the connections was kept below 50 kΩ and were also controlled regularly between the measurements. Lastly the electrode positions were scanned. The participant also filled out the informed consent, a photo agreement and the Waterloo Handedness Questionnaire. They were asked to use their dominant hand for the grasping task.

During the experiment the subject was seated in a chair with an adjustable table in front of them. They got in-ear headphones through which the sounds for the auditory oddball task were played. Before the experiment started the subjects were informed about the tasks they were going to perform and got the possibility to ask questions about the procedure. We emphasised that the cube should be lifted as many times as possible *without* breaking it, and that the grasping task was the main task. The subjects also got to listen to one sound (or more if requested) of each type: *frequent*, *rare*, *novel* and *start/stop-sound*, so that they knew what to listen for. The start/stop sound consisting of three consecutive tones, were used to notice the participant that they could start respectively end doing the task. We also adjusted the audio to fit the subject’s preference. This could give rise to some differences between the individual results, since the intensity of a sound affects the amplitude of the reaction, or ERP components [31]. However, keeping the volume constant would have meant that the subjects would experience different volumes because of differences in hearing, which would also give raise to individual differences. If the sounds had been hard to hear or painfully loud, this would have contributed to exhausting the subjects faster. Therefore, the subjects got to adjust the volume so that they felt most comfortable.

The subject was also instructed to not blink excessively, to not frown, clench their jaws or keep unnecessary tension in any other muscles. This is to avoid artifacts and will be discussed further in section 3.4.5.

The time needed for each condition depends on how many epochs is needed to get a satisfactory ERP waveform. This in turn depends on what you are looking for in the data and how much noise there is, but Luck [31] recommends 10-50 epochs for larger components, such as P3, and 100-500 for smaller, such as P1. This because more measurements increase the signal-to-noise ratio and thereby makes it possible to study smaller components. Since stimulus duration, interstimulus interval and the percentage of novel sounds are already set (see section 3.2.3), the time for each condition depends on the number of epochs we chose to measure. We have chosen to use an algorithm that plays 600-720 stimuli in total. With novel sounds being 10 % of the sounds that gives us

60-72 novel sounds per condition. That way we also have some margin if some of the epochs needs to be rejected because of blinks or other artifacts, at least for large components. To keep the novelty of the novel sounds they were not repeated during a condition, which means that they were repeated a maximum of three times for each subject, with at least five minutes between the repetitions. Some of the sounds were not repeated at all. The algorithm randomizes the order of the played sounds as well as the interstimulus time, such that frequent, rare and novel sounds are mixed and played with different interstimulus times between each other.

Maximizing the number of epochs needs to be weighed against too long blocks. This will exhaust the subjects and might affect the number of subjects willing to participate. But more importantly it will affect the subject's ability to stay focused on the task, and longer times might therefore do more damage than good.

It is also important to insert enough time for rest between the measurements. This helps to keep the subjects alert and focused on the task. It can also reduce blinking and muscle artifacts during the measurement, since the breaks gives the subject time to blink and stretch. For this reason, each condition was divided into three blocks of about four minutes with at least one minute break between them. Between each condition there was also time for about five minutes break, or more depending on what the subject wanted. By letting subjects perform the same task for three hours, or until they were exhausted, Trejo et al. [58] have shown that fatigue will affect the measurement by increasing the amplitude of both the alpha and theta frequency bands and the P2 component. However, the same study showed that N1 and P3 was not significantly affected by the time of the task.

This means that each subject completed in total nine blocks, three for each condition. At the start and end of each block the special start/stop sound was played. After each block the subjects reported the number of rare sounds they had counted. During the longer break after each condition the subjects also filled in the NASA-RTLX questionnaire. All subjects started with the *no task* condition, and moved on to *easy task* and *hard task*, in that order. By doing so, the level of arousal will tend to vary between the conditions. To avoid this, Luck [31] recommends varying conditions unpredictably within each trial block. However, in the future study that this work is in preparation for, it will not be possible to switch back and forth between the conditions, since the conditions in that case will be with and without anaesthesia, respectively. It would be possible to use a random order between the different conditions, for example by inviting the subjects for two separate days, but we decided against this since the same order would make it easier to compare the different subjects' learning processes.

When the subjects performed the grasping task, the number of times they lifted the cube over the barrier was counted in order to get an indication of how well they accomplished the task during the different levels of difficulty. This was then divided by the time to compute number of lifts per minute, taking into account the fact that the total duration of the blocks shifted slightly. Since the task was to lift the cube without breaking it, we also counted the number of times they broke the cube (pressed it too hard such that it lit up). A success rate was then computed by subtracting the number of times the cube was broken from the number of total lifts. These measurements of performance will, together with the NASA-RTLX, be used to verify the differences between the different conditions. It will also be investigated for each block and compared to look for learning effects. It is expected that if learning takes place, performance would increase between the blocks.

An experiment procedure for this work can be found in Appendix A, with more information about preparation, execution and the work needed after each experiment. Also, at the end of the experiment the subjects also participated in another study. However, the procedures or results for this are not discussed in this work and since it was performed at the end it should not affect the result of this work.

3.4 Signal processing

Before analysing the EEG data, it needs to be processed to reduce the signal-to-noise ratio and obtain clean averaged curves to measure ERP components and frequency bands. This section describes common steps for signal processing and motivate our choices for this work.

The signal processing has been done using *EEGLAB* [59], which is a freely available *MATLAB* toolbox, and the plugin *ERPLAB* [60]. These are specifically designed to analyse EEG and ERP data.

3.4.1 Offline referencing

Even if the EEG equipment uses a reference site during the measurements, as discussed in section 2.2.1, this site needs to be specified and sometimes changed offline before analysing the data. This is called offline referencing or, if the reference site is changed, re-referencing. Since there are no electrically neutral sites on the head or the body in terms of neural activity, there are no perfect reference sites. This means that ERP measured at an active electrode will both reflect the EEG at the active electrode site and the reference site. Therefore, it is important to choose the reference site with caution, so it does not cancel out important information in the data. This means for example that a reference site near the site of interest is not a good choice. Also, reference sites that pick up much noise should be avoided to not get extra noise in the data. Which reference site that is the best depends on the application. [31]

Common reference sites used are one or both of the earlobes (e.g. [17]–[19], [21], [37], [39], [44], [46], [61]) or one or both of the mastoids (the bones directly behind the ears, e.g. [36], [40], [41], [55], [62], [63]) because they are convenient and not biased toward one hemisphere if the average of the earlobes or the mastoids are used. They also lie close to each other, which means that EEG data referenced to the earlobes are comparable to EEG data referenced to the mastoids. Often the average of the earlobes or mastoids are used, but the two earlobes or the two mastoids can also be linked together physically with a wire and used as reference. Another common referencing technique is to use an average of several electrodes as reference. [31]

The ears and mastoids can also be used as ground (e.g. [21], [38] used the earlobes and [43], [45] used the mastoids). Other possible sites to use as ground include the forehead (e.g. [37], [41], [42]) and, most commonly in our literature study, the FPz electrode (e.g. [18], [19], [40], [44], [46]).

These differences between studies are a bit problematic since depending on which electrodes that are used for referencing, the ERP waveform can change. This makes different studies hard to compare.

Based on Lucks [31] recommendation to use a reference cite that is not biased to one hemisphere, not introduces a lot of noise, not close to the place on the scalp where the effect of interest is largest and most important to use a reference that is commonly used in other papers, we choose to use average earlobes as offline reference.

3.4.2 Amplification

The EEG technique uses electrodes attached to the scalp to measure the electrical signals from the brain. These signals are very small, usually under 100 microvolts, and need to be amplified by a factor of 1000-100 000. As mentioned before, this is done either directly with the recorded signal (active electrodes) or afterwards with a computer (passive electrodes). The first alternative is better, since this reduces the risk of amplifying noise that can arise between the electrodes and the computer. This is also what is used in this work.

3.4.3 Filtering

Before anything can be said about the EEG data it needs to be filtered to take away some of the non-neural background noise. For example, the electrodes pick up the frequency of the surrounding electrical equipment. The power line frequency in Sweden is 50 Hz [64], which is higher than the frequencies that are commonly of interest in studies regarding cognitive workload. For this reason, the power line noise can be removed with a low pass filter. Luck [31] recommends using one with a half amplitude cut-off of 20-50 Hz. A half amplitude cut-off of 50 Hz means that 50 % of the 50 Hz signals will be cancelled. Hence, less than 50 Hz should be used to avoid the power line noise. Here one also need to take the Nyquist theorem into consideration. This theorem states that the sample rate should be at least twice as high as the highest frequency of the signal, or else information about the signal might be lost. We have used a sampling rate of 2400 Hz, which is more than enough with a half-amplitude cut-off of less than 50 Hz.

It is also common to see slow shifts in the EEG data that arise from changes in skin hydration or static changes in the electrodes. That is why a high pass filter should also be applied, of about 0,05-0,2 Hz. [31]

We have chosen to use a noncausal IRR Butterworth filter with high-pass 0,1 Hz, low-pass 30 Hz and a slope of 12 dB/oct.

3.4.4 Epoching

Because of the trigbox, every stimulus is time-locked and marked in the raw data. The act of taking out segments around each stimulus in the EEG data is called *epoching*. The length of the epoch varies between different studies, but is usually about 1 second, for example 200 ms pre-stimulus and 800 ms after.

Luck [31] recommends using a pre-stimulus that is at least 20 % of the total epoch time and to use even multiples of 100. The latter is because this tends to cancel out alpha oscillations. The pre-stimulus period is used as a baseline for the epoch, as a complement to filtering to reduce the effect of slow voltage drifts (discussed above). In this work we have chosen to use a pre-stimulus interval of 200 ms and a post-stimulus interval of 700 ms. This way, the pre-stimulus period is more than 20 % of the total epoch, and the post-stimulus time should cover all the components described in section 2.3.2.

3.4.5 Artifact management

So far, filters and the use of the pre-stimulus interval as a baseline for each epoch has reduced the effect of the small and continuous types of noise and artifacts discussed above. Also, when averaging the epochs together, random noise will tend to cancel out. However, artifacts that always have the same polarity will have to be tended to by other means [63].

It is always a challenge to measure ERPs when the task demands that the subject moves frequently, since every movement elicits artifacts in the data. This is why the subjects were told to try to avoid unnecessary tension, see section 3.3.4. The most common, and also typically biggest, artifacts are the ones related to the eyes, i.e. blinks and eye movements. These potentials are called electrooculogram (EOG). Dristelle [63] says that it is standard procedure to instruct the participants not to blink and to keep their eyes fixed, while Luck [31] problematize this, saying that keeping from blinking could yield other problems and affect the cognitive workload. Hence, we decided to ask the subjects to avoid excessive blinking, but that they should blink when they needed to.

The procedure of having the subjects focus on a fixed point is used in many ERP studies, but it is not always possible. Here it might be possible, although very unnatural, to ask the subjects to fix on a given point while moving the cube back and forth. In the future experiment, when using anaesthesia, the visual feedback will be the only way for the subjects to know if they are gripping the cube or not, when performing the grasping task. Compromising their ability to see the cube properly would then increase the difficulty of that condition in an unwanted way. For this reason, we wanted to examine if the method, including eye artifacts, can be used to measure cognitive workload.

Below, we will present how eye related artifacts can be detected and either rejected or corrected. Lastly, we will discuss how these techniques can be used to mine the EEG signals behind the artifacts.

3.4.5.1 Artifact detection and rejection

Unlike continuous noise from nearby electrical sources or such that can be removed by using filters, this approach does not work on large and transient artifacts like eye movements and blinks. However, the transient nature of these artifacts means that they do not affect all of the data, only the parts where the artifacts occur. For this reason, one possible way of dealing with these kinds of artifacts is by simply removing all epochs that contain them.

The eyes are electrical dipoles, with a positive potential at the front of the eye and a negative at the back [31]. That means that when the eyes are moved, the signals change. To detect these changes, electrodes are commonly placed around the eyes when measuring EEG, as shown in Figure 7. *Horizontal EOG* (HEOG) is measured with electrodes lateral to both eyes, near the temple, and *vertical EOG* (VEOG) is measured with electrodes above and below one eye. The VEOG electrodes are also used to detect blinks, since they produce responses that are opposite in polarity depending if they are measured above or below the eyes. This means that blinks can be recognized by the fact that the EEG signal from the electrode under the eye, VEOG2, are opposite in direction compared to the other channels, as can be seen in Figure 8. When the eyes are moved, a muscle is contracted. The eyes then usually look in the same direction for a little while before they move again. This creates a characteristic box-shaped output that can also be seen in Figure 8. [31]

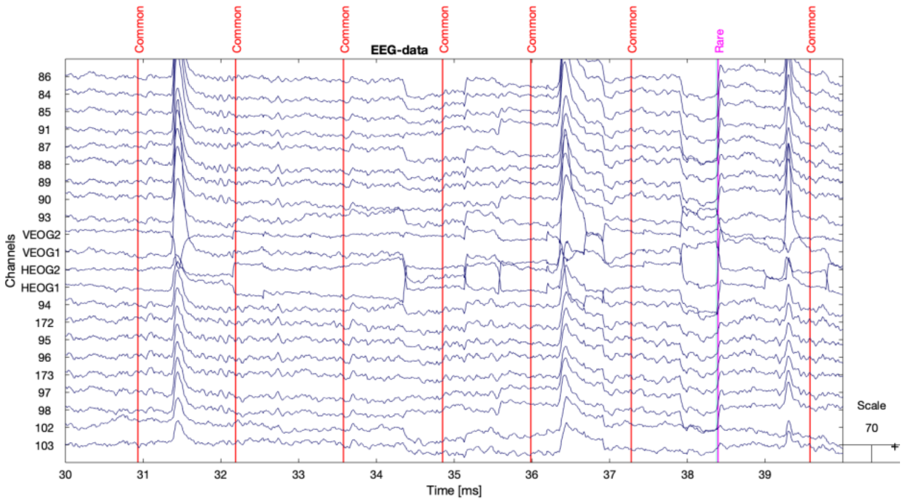


Figure 8. Eye blinks and horizontal eye artifacts. Notice the blinks at approximately 31.5, 36.5, and 39.5 s, where VEOG2 deflects in the opposite direction compared to the other channels. We can also see the characteristic box shapes of eye movements at approximately 32-34.5 and 35-35.5 s.

Thus, the artifacts arising from eye movements and blinks are fairly easily recognised. Therefore, it is possible to visually inspect all the data and manually remove all epochs that contain a blink or an eye movement. This is, however, very time consuming. A better alternative is to use some sort of algorithm to find these artifacts. One simple alternative is to use a *voltage threshold* to exclude all activity that are above or below certain limits. However, even after the use of high pass filters and pre-stimulus baselines there can be slow voltage shifts over an epoch. If the baseline moves from the mid-line this means that some large shifts might still be within the limits and, similarly, small shifts might end up rejected. [31]

A better alternative is to use a *moving peak-to-peak window*. This technique measures the voltage difference within a specified window width, for example 200 ms. If this voltage difference is larger than a given threshold the epoch is rejected. Then the window is shifted right (by e.g. 50 ms) and compares the maximum voltage difference in that area to the threshold for rejection. This procedure is repeated until all the data has been examined. [31]

A third method is the *step function*. Similarly to the moving peak-to-peak window it moves across the data in steps and looks at the data in short segments at a time. However, instead of measuring the maximum voltage difference, the step function method compares the average voltage in the first half of the window to the average of the other half. This is a better method to detect eye movements, and it can also detect blinks. Luck says that this might be slightly better overall compared to the moving window method, but he mainly stresses the fact that both are highly superior to simply using a voltage threshold. [31]

Both the moving window and step function methods need a specified threshold for rejection. This limit can be set equal for all subjects, but this is not recommended by Luck [31]. No humans are alike and neither are their EEGs. This means that a threshold that is right for one person might be too high to reject artifacts from another, or too low so that epochs without artifacts are rejected. Also, no algorithm is perfect, and there is always a possibility for mistakes. For this reason, Luck [31] recommends setting an initial, trial threshold and then visually inspect the data for each subject, to see if there are any epochs that should have been rejected but were not, or that was falsely rejected. If so, the threshold should be changed until a better fit has been found.

3.4.5.2 Artifact correction

Instead of rejecting data it is possible to use artifact correction, which removes only the part of the signal that corresponds to an artifact. This can be done by using a method called *independent component analysis* (ICA), which decomposes the EEG-data into a set of independent components. There are several different algorithms that can be used for this, but ICA is the most effective one [63]. By using ICA it is possible to identify components that isolate artifacts like blinks and eye movements and then simply subtract these components from the rest of the EEG data.

The theory behind ICA is often described in terms of audio recordings on a cocktail party, the so-called cocktail-party problem [65]. Imagine that you are in a room with several people talking simultaneously and that you record audio with several microphones placed on different locations in the room. Also imagine that you want to separate the signals in the recordings to find out what each person in the room is saying. When you listen to the different signals you will notice that they sound a bit different depending on where they are recorded, even if they are recorded simultaneously in the same environment. But, if you separate the signals into their independent sources you will notice that they consist of the same sources, each source weighted a bit different depending on where in the room the signal was recorded. The signals are a mixture of sounds from the different sources and can be written as weighted sums. This means that the signals x(t), where x is the amplitudes and t the time index, can be written as

$$\mathrm{x}_1 = a_{11}\mathrm{s}_1 + \mathrm{a}_{12}\mathrm{s}_2 + \cdots + a\mathrm{s}_\mathrm{n}$$

$$\mathrm{x}_2 = a_{21}\mathrm{s}_1 + \mathrm{a}_{22}\mathrm{s}_2 + \cdots + a_{2\mathrm{n}}\mathrm{s}_\mathrm{n}$$

$$\vdots$$

$$\mathrm{x}_n = a_{n1}\mathrm{s}_1 + \mathrm{a}_{\mathrm{n}2}\mathrm{s}_2 + \cdots + a_{\mathrm{nn}}\mathrm{s}_\mathrm{n}$$

where *s* represents the individual sources and a are the weighs. By assuming that the sources are statistically independent, ICA can separate them from the mixed signals x(t) and calculate the weights in the sum. However, as we can see in the equations above there are n^2 weights to calculate but only n independent sources and n signals. To solve this problem ICA therefore uses the vector-notation of these equations instead, written as

$$x = As$$

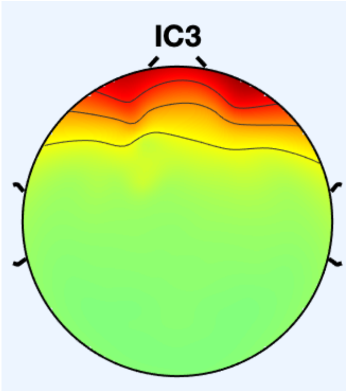
and assumes that both x and s are vectors of random variables. Both A and s are then estimated by assuming that s are statistically independent sources with non-gaussian distribution. ICA first estimates the mixing-matrix A and then computes its inverse W, the unmixing-matrix, where the independent sources can be calculated as

$$s = Wx.$$

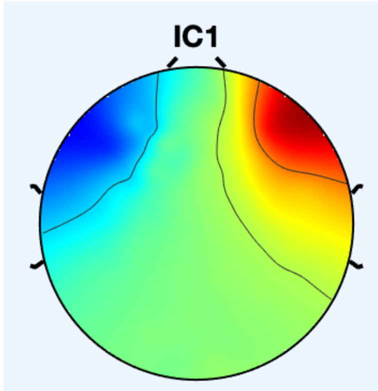
As in the cocktail party problem, EEG data can be described as a mixture of signals from different components of brain activity. Hence, we can apply ICA to the EEG data and get an estimation of what independent components that can be behind the measured signals from the different scalp electrodes. As can be seen above, ICA will calculate as many sources as there were inputted signals, i.e. it assumes that n signals will be generated from n independent sources.

Even though ICA can be used to remove many different types of artifacts, most studies (e.g. [55], [19], [11], [21], [38], [62], [63]) use ICA just for removal of eye blinks and horizontal eye movements. However, some studies (e.g. [24], [43]) also use ICA to remove other artifacts such as muscle artifacts, heart beat artifact, channel noise and power line noise. As described above in section 3.4.5.1 both eye blinks and horizontal eye movements have characteristic shapes that make them easy to detect, which means that independent components from eye artifacts are easy to recognize.

When visually inspecting the ICA decomposed data to label components as artifacts one often looks at the scalp topography together with the frequency spectra and time courses, to identify characteristics of certain artifacts [66]. Eye blinks are characterized by a scalp topography that shows high activity at the front of the scalp (see Figure 9a), together with the characteristic spikes that can be seen in Figure 8 and an activity power spectrum that decreases smoothly for all frequencies and lacks peaks. Horizontal eye movements are like eye blinks identified by a smooth frequency spectrum and in the time course they are characterized by the box-shapes shown in Figure 8. An example of the scalp topography corresponding to horizontal eye movements can be seen in Figure 9b. It typically shows high activity in the front, with opposite polarity on the left and right side of the head. [67]



a) A typical eye blink, where the signal is strong close to the eyes.



b) A typical horizontal eye movement, where the signal is strongly negative/positive by the left/right eye.

Figure 9. Scalp topography showing common eye artifacts: eye blinks and horizontal eye movements.

3.4.5.3 The choice of whether to reject or correct artifacts

As mentioned in 2.3.2, there are no way to mine the different components of an EEG measurement with absolute certainty. ICA is based on an underdetermined problem and performs an optimization process to conclude the underlying components. That means that ICA will not deliver the same result for the same data each time it is run. Therefore, many chose to reject all epochs containing too much noise, either by using an algorithm with a set threshold or by visual inspection ([14], [17], [18], [36], [37], [39], [41], [45], [46], [48], [50]).

However, despite its flaws, ICA is frequently used in ERP studies, since it can “save” data that would otherwise have needed to be rejected. In 2017, Drisdelle et al. [63] compared the convention of asking the subjects to keep their eyes fixed and thereafter rejecting epochs with eye movements to when eye movements were encouraged and later corrected using ICA. They found that ICA conserved the data and recommended this as a possible standard approach, instead of rejecting valuable data. The use of ICA was also recommended by Castellar et al. [29] in 2019. They specifically recommended it for studies evaluating cognitive workload by using computer games, and since our experiments have the need for eye movements in common, we deem ICA to be suitable for our study as well.

When using ICA for artifact correction, it is usually combined with rejecting epochs with remaining artifacts. After ICA has identified and removed the most commonly occurring artifacts, for example eye movements, the epochs that are deemed to contain too much noise, electrode cable movements or such can be rejected using either set thresholds (e.g. [19], [55], [62], [63]) or by visual inspection (e.g. [24], [38], [43]). Luck also recommends to remove portions of data with big voltage shifts before applying the ICA algorithm [31]. These shifts can occur if the subject makes a big movement and they are usually found in the beginning and end of a recording, before and after the actual task.

The identification of components that corresponds to artifacts is often done by visual inspection, but one can also use tools that automatically detects and labels artifactual components. One such tool is *ADJUST* [68] which combines stereotyped artifact-specific spatial and temporal features to identify artifactual components corresponding to eye blinks, vertical and horizontal eye movements and discontinuities. The dataset used for selecting the features used for the algorithm consists of EEG data from a dual-task study of 21 subjects. Another tool that can be used for automatic detection of artifacts is *ICLabel* [69], which contains spatiotemporal measures from more than 6000 EEG measurements and 20000 independent components. It labels the components into seven categories: brain, muscle, eye, heart, channel noise, line noise and other, where the category other includes all components that does not fit into one of the previous types. Through a comparative study it is shown that ICLabel perform better than or comparably to other automatic component detection tools and this with a tenth of the computation time.

Since the task of this work demands movements that will give rise to artifacts, we have decided to use ICA as a way to save data that would otherwise be useless. Note that ICA should be applied before epoching the data [31]. First the data was visually inspected and portions with artifacts that obviously arose from equipment failure was removed. Generally, this meant cutting from the beginning and end of each block, but one big deviation was also removed within one of the blocks. We also found out that one channel looked faulty for two conditions (showing big, regular fluctuations) so this channel was removed from analysis. For more information about how the signal processing affected the data for each subject and condition, see Appendix D.

We then used the ICA algorithm (runica) available in EEGLAB [60] and later ICLabel with a threshold of 85 % for eye- and muscle artifacts. This means that any components where ICLabel estimated the likelihood that it was linked to eye- or muscle activity as greater than 85 % was removed from the data. Before removing the components labelled as 85 % certain eye- or muscle artifacts, these components were also visually inspected to be sure that they were fine to remove. The ICA-corrected data was then epoched and checked with a moving window peak-to-peak algorithm to reject epochs with remaining artifacts that exceeded 100 µV compared to the mean of the pre stimulus interval (same threshold as [63]). We chose to use a set threshold for all subjects, since Luck's recommendation of using individual thresholds mainly applies to eye artifacts. Since these should be removed with ICA, the set threshold was used to find noise or artifacts originating from moving electrodes or unstable connections to the scalp. If more than half of the epochs around the novel sounds were rejected, the subject should be removed from further analysis, according to a recommendation from Luck [31]. For our data, 1-22 (mean 4,7) ICA components were removed from each subject and condition. After the following artifact detection one subject had to be rejected because of excessive noise. For the others 0-40,3 % (mean 7,5 %) of the epochs around the novel sounds were removed from each subject and condition, leaving 37-72 (mean 60,7) novel sound epochs for analysis. For more information about this, see Appendix D.

3.4.6 Averaging

The process of averaging epochs is meant to increase the amplitude of any effect that is related to the stimuli, while decreasing artifacts that are assumed to be completely random with respect to the time-lock of the stimuli.

Each epoch is sorted into a bin, according to what stimulus it is extracted around and what condition it is measured for. Using this information, averaged ERPs can be constructed for each bin. Here, we want to measure the ERPs elicited by the novel, complex sounds, so an average is constructed where epochs for novel, complex sounds are sorted into bins for each condition. After that, a *grand average* can be constructed for each condition. That is the average of the data from all the subjects performing the different tasks (*no task*, *easy task* and *hard task*).

3.5 Measurements to indicate cognitive workload

As seen in section 2.3.2 and 2.3.3 there are two common ways to measure cognitive workload with EEG: ERP components and frequency bands. However, in neither case it is easy to determine what band or component to focus on. For ERP components the ranges in which the components have been found also varies greatly, so it is not easy to know how to focus the search. In this section we will discuss how to measure the amplitude of ERP components and frequency bands and how we have chosen what factors to look at for measuring cognitive workload. The differences are also often seen at different sites of the scalp, so we will also discuss how we chose which electrodes to analyse.

3.5.1 Measuring amplitudes of averaged ERPs

When the averaged ERP waveforms have been obtained, we want to be able to measure the amplitude of different peaks, to obtain an indication of underlying ERP components, discussed in 2.3.2. However, there are a few different ways to measure the amplitude of an ERP peak. Here we will discuss the ones presented by Luck in “An introduction to the Event-Related Potential Technique” [31]. We will also discuss which ERP components that should be studied and how the latency window for measuring amplitude of different ERP components can be chosen.

3.5.1.1 Choosing a method to measure amplitude

One way of measuring the amplitude of a peak is to measure the biggest voltage within a given time interval. This is called *peak amplitude*. However, this technique is sensitive to the level of noise in the signal [31]. Since the level of noise depends on the number of epochs in the averaged waveform, it should only be used to compare conditions with an equal number of epochs [31]. This usually means having to use a subset of the epochs for some of the conditions and subjects, and thereby not use all available data. Peak amplitude also builds on the fact that there is something special about the peak voltage, while this is in fact not true since the peak only is an indication of underlying components (as discussed in section 2.3.2).

Another possibility is to measure the *mean amplitude* over a given interval. Mean amplitude is not very sensitive to noise and can therefore be used to compare conditions and subjects using all available epochs. It is not affected by smaller changes in latency between different conditions or subjects but is sensitive to the choice of latency window.

A third option is to measure something called *signed area amplitude*. This means that different sections of the ERP waveform integral are signed as either positive or negative [31]. That way one could for example chose to measure only the area above the baseline within a given interval. Since portions of data that are within this interval but below the baseline does not affect the result, this method is less sensitive to the choice of latency window compared to using mean amplitude. That also makes it superior when there are bigger latency differences between different subjects. This approach seems promising but has the drawback that it is relatively new and not frequently used. The tools in ERPLAB [70] are not yet adapted for this method. Another drawback is that it is usually bigger than the actual value, since noisy waveforms tend to have larger values than clean [31].

Considering the properties of the different measurement techniques, we have chosen to measure mean amplitude. Even though signed area amplitude is better in many aspects we have not found a formal comparison between the two. Mean amplitude is more commonly used (e.g. [19], [36]–[38], [44], [45], [62], [63]) which makes it more reliable, and using this method will make it easier to compare our results to others. It also makes it easier to be able to perform the measurements in ERPLAB.

3.5.1.2 The choice of which ERP components to examine

As discussed in section 2.3.2, several components have been shown to indicate cognitive workload, and the interval for where they have been found often varies greatly between different studies, as can be seen in Figure 4 and Figure 5. According to our literature research the components N1, N2, P2, P3, and LPP can all be used to measure cognitive workload. Since we are using complex, novel sounds together with simple tones we will most likely see the *novelty* P3, as discussed in section 2.3.2.4.

Since this study is a pilot study in preparation for future work, we decided to use our own data to determine what components to examine. Therefore, we computed an average across all subjects, conditions and channels (electrodes) and chose components to examined based on which ones were visible in that waveform.

3.5.1.3 The choice of latency window and electrode sites

In the studies that we have read, it is rarely described why certain electrode sites were chosen, but common ones to use are the central channels Fz, Pz and Cz (e.g. [18], [19], [36], [37], [41]).

Here, we wanted to investigate which of the 121 available channels (electrodes) would be best for our purpose. We again used the waveform averaged over all subjects, conditions and channels and used this to choose appropriate latency windows. These latency windows were then used to determine which channels to use for the analysis. The mean amplitude within each latency window of the average over all subjects and conditions were measured. By comparing this between the 121 channels, we could see at which electrode sites each peak was most prominent. A similar method was used by Dyke et al. [38].

We have also chosen to use clusters of electrodes (as e.g. [62]). By using the mean ERPs as captured by several electrodes, we should obtain a waveform with bigger signal-to-noise ratio than would be given by just one electrode. This also seems like a good way to take advantage of our 121 electrode channels and it makes the method sturdier against problems with some electrodes. If only one electrode is used and this does not end up having a good connection for some of the subjects, a lot of data would be useless. We also wanted to try using clusters of two different sizes, where the smaller clusters consisted of five electrodes and the larger of seventeen. Each cluster was centered around the area where the mean amplitude for each peak was largest.

Now, it remained to measure the amplitude of the ERP waveform of each subject, by using these clusters. When measuring mean amplitude, the approach that seems most common is to use the grand average across all subjects and conditions to determine what latency window to use for each electrode site (e.g. [19], [36], [38], [44], [45], [63]), or in our case electrode cluster. A time window is chosen for each component of interest at each electrode site and this latency window is then used for all subjects at that site. The width of this window varies between different studies, as can be seen in Figure 4 and Figure 5.

We have chosen to follow a procedure presented by Handy [71] (and used by for example Miller et al. [18] and Deeny et al. [17]) to use small latency windows centered around the peaks in the grand average with a width of between 15 and 40 ms. We therefore chose to use an interval of 40 ms, centered around the latency of the peak amplitude of the relevant component in each cluster. This latency window was then used to measure each component for each subject and condition at the small and large electrode cluster. Another option would have been to select latency windows for each subject as well, but since we have not seen this method in the literature, we decided against it.

Luck argues that using your own data to decide what latency windows to use creates bias that increases the chances of your experiment yielding the result you are looking for [31]. He instead recommends using latency windows from previous, similar studies. As seen in Figure 4 and Figure 5, this varies between studies, and it is not easy to decide which ones to use. However, Luck says that you can use your own data for deciding measurement parameters if you perform a follow-up experiment, where you use the same latencies [31]. Since this work is a pilot study for a future experiment, we have chosen to use our own data for determining the latency windows. This will give the following work good references to use when deciding their methods. The same argument also holds for using our own data to decide which electrode clusters to use.

3.5.1.4 A compiled measurement for ERP

The theory for ERP components state that the amplitude of the components should decrease as the difficulty of the primary task increases, as discussed in section 3.2. This is because the brain's reaction to the sounds decrease as the primary task demands more attention. This means that we could construct a *compiled measurement* for ERP, by using the amplitudes for all assessed ERP components. We therefore constructed a compiled measurement by adding the measurements for all measured mean amplitudes, to see if that could be used to better assess the differences in cognitive workload between the different conditions.

3.5.2 Measuring frequency bands

Here we will motivate which frequency bands were measured and how we chose to measure these.

3.5.2.1 The choice of which frequency bands to examine

According to our literature study, described in section 2.3.3, the most successful frequency bands for assessing cognitive workload is *Theta* (3-8 Hz) and *Broadband Alpha* (8-13 Hz) sometimes divided into *Low-Alpha* (8-10 Hz) and *High-Alpha* (10-13 Hz). We have therefore decided to measure these frequency bands.

For the quotient *Theta/Alpha*, there are several options for combining frontal and parietal measurements, as discussed in section 2.3.3.6. We have decided to use our measured values for theta and broadband alpha, where the scalp region of these will be determined by where they are most prominent, as will be discussed further in this section. This means, using the same electrodes for measuring Theta and Alpha, and take the quotient Theta/Alpha in that area.

3.5.2.2 Choosing a method

A common way to analyze the frequency characteristics of EEG data is to compute the *power spectral density* (PSD) for the different frequency bands mentioned in section 2.3.3 (e.g. [11], [14], [44], [72], [73]). Puma et al. [11] and Holm et al. [73] both conducted studies on cognitive workload during multitasking. The tasks in these studies included controlling events on a screen by joysticks and clicking buttons. In the studies carried out by Jaquess et al. [14], [72] the participants were to perform different scenarios in a flight simulator. The mentioned studies have in common that the participants are performing some kind of movement of the hands (i.e. a motor task). Common between the studies is also that they calculate the PSD and analyse the frequency bands alpha and theta. These studies all use approximately the same procedure for spectral analysis and achieve significant results, which shows that this method can be used for studies of different types. Therefore, we have chosen to follow the procedure for spectral power analysis carried out in these studies.

The EEG data first went through the same processing steps as when computing ERP values (described above) i.e. offline referencing (section 3.4.1), filtering (3.4.3), ICA (3.4.5.2), epoching with baseline correction (3.4.4), and artifact rejection (3.4.5.3). To transform the data into the frequency domain and calculate the PSD we used Welch's periodogram with 1 s Hamming windows and 50 % overlap and then calculated the absolute band power for the theta (3-8 Hz), broadband-alpha (8-13 Hz), low-alpha (8-10 Hz) and high-alpha (10-13 Hz) frequency bands. This was done over all three blocks at once for each condition. Lastly the spectral power for each frequency bandwidth was divided by the spectral power for the data's entire frequency spectrum and natural log transformed before it went through statistical analysis.

3.5.2.3 The choice of electrode sites

As described for the calculation of ERPs in section 3.5.1, we wanted to investigate which of the 121 electrodes available in our measures that were the best to use for measuring the different frequency bands. This was done by calculating the average band power where all subjects and conditions were averaged together. The result for each electrode was plotted over the scalp and resulted in a scalp distribution. From this a small and a large cluster of electrodes corresponding to the area of highest band power for each frequency band were extracted. As for the ERP clusters, these areas were centered around the electrode corresponding to the highest band power for each frequency band. When measuring the band power for each frequency band, the absolute band power was averaged for the electrodes of the clusters.

3.6 Statistical analysis

The statistical analyses of the results of the measured values for ERP components and frequency bands has been performed using a two-way ANOVA. The ANOVA analysis gives the significance, or p-value, of the results in a comparison. The p-value denotes the probability that the results are due to random errors. If p = 0.05, that means that there is a 5 % probability that the null hypothesis is true, i.e. that there is no real difference between the compared conditions. A smaller p-value thereby means that you can reject the null hypothesis with more confidence.

This means that there is always some chance that the null hypothesis is falsely rejected, i.e. that you say that there is a difference between the conditions even though it is just by chance. This problem grows when performing *multiple comparisons*. If several factors are measured, the probability of at least one of them showing a statistically significant effect increases, compared to if just one factor is examined.

In our case, one of our goals is to examine which factors should be used to assess cognitive workload (ERP components and frequency bands), and how this should be done (latency windows and electrode sites). This means that we need to face the problem of multiple comparisons. However, since this study is more exploratory, this is not seen as a problem. However, multiple comparisons is still a factor that needs to be considered when analysing the results of this study.

The two-way ANOVA was computed using the MATLAB functions *anova2* and multcompare. The ANOVA analysis returns a table showing, among other things, the p-value for comparisons across rows and columns in the table of interest. If the p-value is less than 0.05, this means that there is a significant difference between some of the rows/columns. However, the anova2 function does not say which rows or columns. To examine this, we have used *multcompare* that returns a comparison between the rows/columns. In our case, it will tell us if there is a significant difference between all conditions or only some of the conditions, for example *no task* might be significantly different from *hard task*.

3.7 Design of force sensitive cube

For the grasping task we needed an object for the subject to lift, which could be imaginarily broken repeatedly. The object is said to be broken when it is squeezed harder than a given limit and when the object is broken it should give some feedback to the subject. We decided that the feedback need to be visual, because auditory feedback can interrupt with the auditory stimuli from the oddball task and sensory feedback would possibly interrupt the lifting task by for example making the subject drop the object. We also had in mind that the object should later be used in a study where sensory feedback is removed using anaesthesia, and therefore using sensory feedback for when the object was broken was no alternative. Talking about sensory feedback, we also concluded that the sides of the object need to be fixed, such that the cube cannot be clamped. This to make gripping the object comparable between subjects with intact limbs and users of hand prostheses, which do not have finger perception in the same way as intact limb people.

To clarify the characteristics needed for this object, design requirements together with their priority were set up.

3.7.1 Design requirements

The design requirements were divided into soft and hard criteria, depending on their level of priority. Here, hard criteria are the ones that need to be fulfilled, whereas the soft criteria are wanted but not needed. The hard criteria were:

- Size of object not bigger than 7cm x 7cm x 7cm
- Object can measure how hard the object is pressed
- Set limit for when the object is pressed too hard and give visual feedback
- Sides of the object should be fixed
- Level of difficulty for lifting the object should be easy to change manually
- High friction on the gripping sides

Further, soft criteria set up were:

- Ability to save and send data wirelessly to a computer
- Changing level of difficulty for lifting the object automatically during the experiment
- Object can measure vertical lifting force
- Electronics in object battery powered

Figure 10. Final version of the force sensitive cube. On the sides are force sensors with sandpaper to gain higher friction. A LED bar on the top shows when the cube is broken i.e. pressed to hard. The cube has a box that can be drawn out and filled with weight. All electronics are fitted inside the cube together with a 9V battery.

3.7.2 Design process

The design process for the object resulted in four different versions, all of them were a development of the previous version. A description of all versions can be found in Appendix B. Firstly, it was decided that the object should be in the shape of a cube because it is an easy shape to construct and manage. We also decided that the level of difficulty for the cube should be given by its weight, which easily can be changed manually. Early in the process we also decided that we wanted the cube to be computerized, to make it easy to set and change limits for when the cube was broken. This also gives the opportunity to save measured data if wanted. For this we used a commercial microcontroller of the make Arduino because it is easily accessible and easy to implement. To measure when the cube is broken, we decided to use force sensitive sensors. The final version of the cube can be seen in Figure 10 and the material used are listed below.

- LED-bar of 8 LEDs, *NeoPixel 8 LEDs WS2812*
- Resistor of 560 Ω
- 2x force sensor, *Interlink FSR 406*

- 3x resistor of 10 kΩ
- Button to reset *Arduino Micro*
- 3D-printed cube 63x63x68 mm
- *Arduino Micro*
- 9V battery
- Coupling board, *Luxorpart 45x34 mm*
- Cables
- Sandpaper, grit size 120
- Bluetooth transmission module, *Velleman HC05*

The cube had two force sensors, one on each side for the subject to grip, and to increase the friction and make it easier for the subject to lift the cube without slipping we added sandpaper on the force sensors. To prevent the cube from gliding on the surface of the table, glue from a glue gun was put at the bottom of the cube.

During the whole design process, the most challenging criterion was to keep the size of the cube within our set limits and at the same time fit all equipment inside the cube, as well as adapt the cube so that the level of difficulty could easily be changed. We also wanted the cube to be as light as possible when no weights were added, such that the change of weight between the two levels (see section 3.1) should be as big as possible. To easily be able to change the weight of the cube a box was included inside the cube and it could be pulled out and filled with weights. We ended up with a cube that weighted 109 g in its basic form, which corresponded to the easier level of difficulty, and 404 g when weight was added (corresponding to the harder level of difficulty). The force limit for breaking the cube was set such that the cube just could be lifted without getting broken on the hard level, which equalled 7.8 N. For information about how the cube was calibrated to get the force limits in Newtons, see Appendix B.

To show if the cube was broken a visual feedback in the form of LED lamps was used. Early in the design process we used a single red LED, but after testing that version of the cube in a pilot test it was clear that it was hard for the subject to see when the LED lit up. Therefore, we changed the single LED to a LED bar of 8 LEDs, which made it much easier for the subject to see when the cube was broken. This also gives the possibility to show a grading of how hard the cube is pressed for the subject all the time if wanted.

The number of times the cube was broken was counted both manually and automatically by the cube. To be able to display on an external computer when and how many times the cube was broken, a Bluetooth transmission module was included in the design.

The variables that the cube can measure is listed below in Table 3, and a sketch of the measured force can be seen in Figure 11.

Table 3. Variables measured by the cube.

VARIABLE	FUNCTION
$count$	Counts number of times the cube is broken
F_1, F_2	Normal forces on the force sensors when the subjects grip the cube

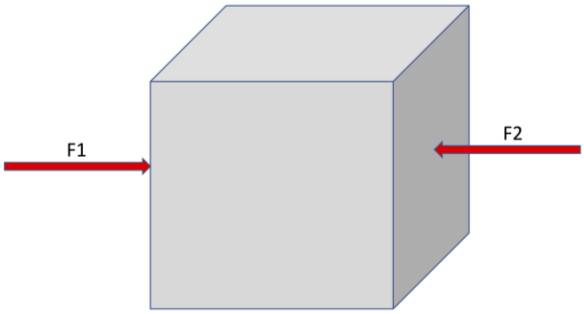


Figure 11. Forces measured by the cube. If F_1 and F_2 surpassed a certain threshold, the cube would light up to indicate that it had been "broken".

4 Results

In this section we will present our results. We will start by mentioning some details in the execution of the experiment that did not go as planned, and how we have handled these problems when analysing the results. Then we present the results for the performance (number of lifts, success rate and oddball test accuracy) and perceived effort, since these indicate if the difficulty between the levels differed as planned. Then we move on to the results for ERP measurements, including latency bands and channel clusters, and frequency band, including channel clusters. Here, we also present the results of the conducted statistical analysis to see which methods gave significant results.

4.1 Notes on the execution of the experiment

There were some smaller problems during the execution of the experiment, that needs to be taken into consideration. First, the LED bar on the cube did not always lit up as planned when the cube was pressed too hard. Sometimes there was a very dim light that was hard to see for both the subject and the experiment leaders. This means that the counted number of breaks might not be accurate, and some subjects did not realise that they broke the cube and thereby did not alter their lifts after breaking it. Therefore, subjects were told how many breaks had been counted after each block and reminded to pay attention to the LED bar. The LED bar was also sensitive to the battery level, so the battery was changed often. The cables were also a bit unstable, so the cube was tested after each block, as an extra precaution. If it did not light up, the number of breaks within the block were marked as unknown. In some blocks we also had technical problems that meant that the number of lifts or breaks were not counted. More information about the data considered for each subject can be found in Table 4.

Table 4. A summary of the data that was gathered for each subject, excluding the EEG data. "X" marks that data was collected and "-" marks missing data.

		ODDBALL TASK			NUMBER OF LIFTS PER MINUTE			
		No task	Easy task	Hard task	Easy task		Hard task	
SUBJ.	NASA-RTLX	# Counted sounds			Time	# Lifts	Time	# Lifts
1	X	X	X	X	X	-	X	X
2	X	X	X	X	X	X	X	X
3	X	X	X	X	X	X	X	X
4	X	X	X	X	X	X	X	X
5	X	-	X	X	X	X	X	X
6	X	X	X	X	X	X	X	X
7	X	X	-	X	-	X	X	X
8	X	X	X	X	X	X	X	X
9	X	X	X	X	X	X	X	-
10	X	X	X	X	X	X	X	X

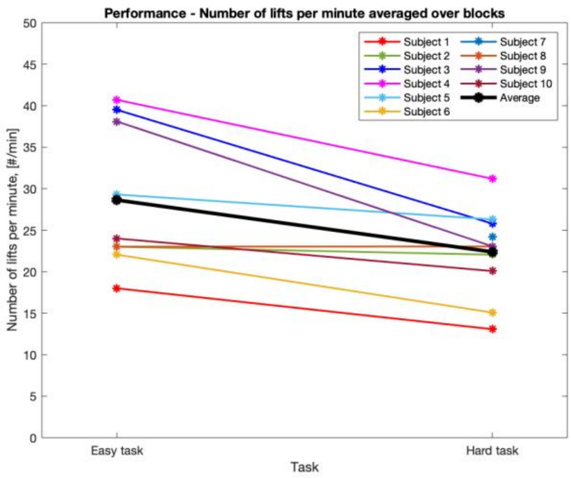
This meant that all parameters were not measured for all subjects. We have chosen to use the available data as much as possible. This means for example that if the number of lifts per minute was missing for one of the blocks, the remaining two were still used, and lay the basis for the overall average for that subject and condition.

We also noticed during the experiment that some subjects sometimes hit the barrier, since they did not always lift the cube high enough. These misses were, however, not counted since the problem was only seen some way into the experiment period.

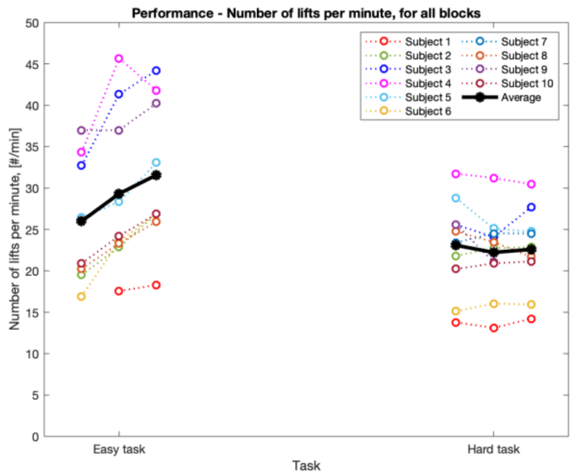
4.2 Performance

When the subjects performed the grasping task, the number of times they lifted the cube over the barrier was counted in order to get an indication of how well they accomplished the task during the different levels of difficulty. The number of lifts per minute for each subject and level of difficulty, together with the mean number of lifts for the two levels are shown in Figure 12a. In Figure 12b we can see the number of lifts for each block as well. Note that the trend is that the subjects lifted the cube less times per minute in the *hard task* than the *easy task*. When looking at the individual blocks we can see that subjects on average lifted the blocks more for each block during the *easy task*, but performed more or less the same number of lifts per minute for each block in the *hard task* condition.

By subtracting the number of times the cube was broken (pressed too hard) from the number of lifts and computing the percentage of successful lifts, we got the success rate for each subject. This is presented in Figure 13. As for the number of lifts per minute, we observe that the success rate decreased when the level of difficulty increased which again indicate that the *hard task* is more difficult than the *easy task*. From this we can tell that not only did the subjects perform less lifts per minute in the *hard task* condition, they also had lower success rate.



a) The number of lifts per minute for the easy and hard condition for each subject.



b) The number of lifts per minute for each of the three blocks within the easy and hard condition.

Figure 12. These plots show the subjects' performance in the grasping task, counted as number of lifts per minute. The black line represents the average as taken over all subjects. Note that the subjects generally perform fewer lifts in the hard task than in the easy task. There is also a trend within the easy task, that subjects lift more every block. This pattern is, however, not seen for the hard task blocks.

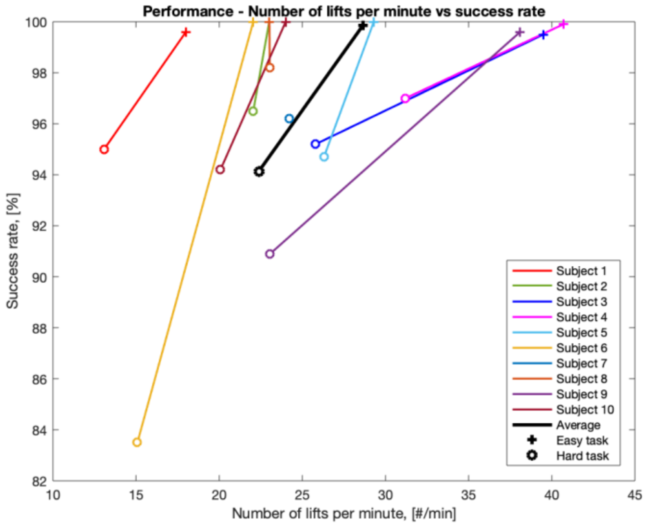
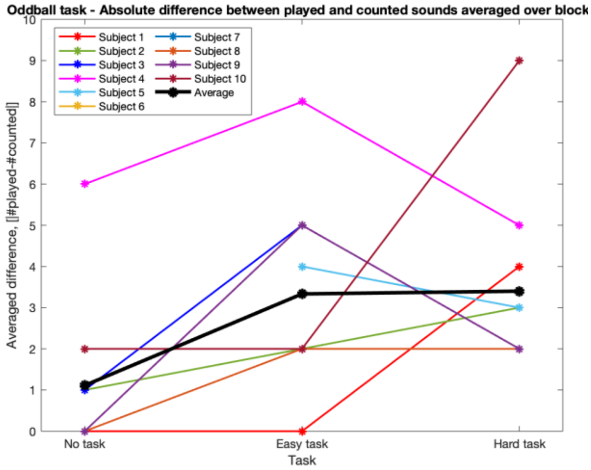
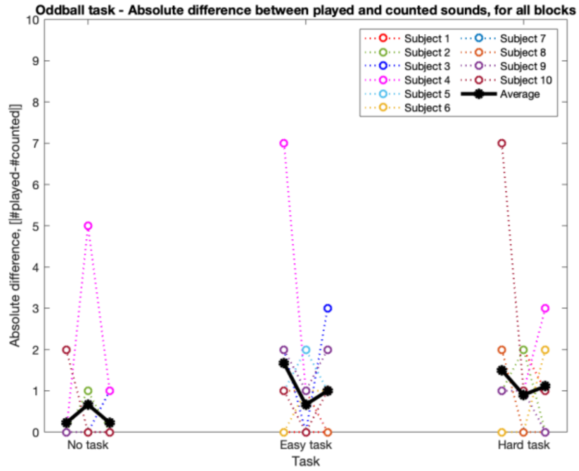


Figure 13. Plot of the success rate as a function of the number of lifts per minute. The black line represents the average as taken over all subjects. Note that almost all subjects performed less lifts per minute with a lower success rate for the hard task (o) compared to the easy task (+).

For the oddball task the subjects were told to count the rare sounds at the same time as they performed the primary task, i.e. the grasping task for the easy and hard condition. During the *no task* condition the task only consisted of counting sounds and focusing their glance at a plus sign at the computer screen. The absolute value of the difference between the number of rare sounds counted by the subject and the number of played rare sounds are presented in Figure 14a, for all conditions and subjects. The mean value for each condition is also plotted, in black. Note that the general trend is that the subjects miscounted more in the *easy* and *hard* condition than in the *no task*, but with only a small change between *easy* and *hard*. This again indicates that an increased level of difficulty impairs the subject's performance of the task, although we can see a bigger variation for each subject. Some subjects performed better in the *hard task* than in the *easy* and *no task*. In Figure 14b, depicting the accuracy of the oddball task for each block in the different conditions, we can see no clear pattern.





a) The accuracy of the oddball task for the three conditions.

b) The accuracy of the oddball task for each of the three blocks within each condition.

Figure 14. These plots show the performance of the subjects as measured by the accuracy of the oddball task, the absolute value of the difference between counted and played sounds. The black line represents the average as taken over all subjects. Note that subjects generally performed best during the no task condition, while there is only a slight increase between easy and hard. There seems to be no clear pattern between the different blocks of each condition.

4.3 Perceived effort

After they performed each condition the subjects got to estimate the effort to perform the tasks by filling in the NASA-RTLX form, and the average task load index was computed as described in section 3.3.3. The resulting average task load index for each condition and subject is plotted in Figure 15, together with the mean task load index for each condition. We observe that most of the subjects estimated that they had to put more effort into the *hard task* than into the *easy task* and *no task*. Most of them also estimated that their effort was lowest in the *no task* condition. Again, the average for each condition indicates that the *hard task* is harder than the *easy task* and that *no task* is easiest to perform. As for the oddball task we see some variation in which task the subjects thought they had to put in most effort.

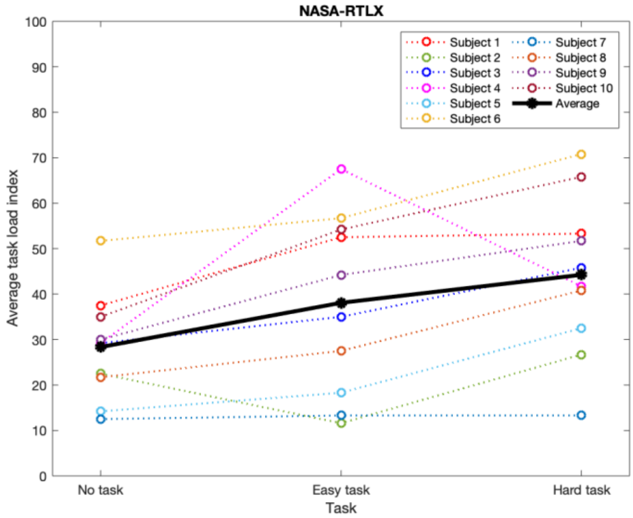


Figure 15. Average task load index from the self-assessment questionnaire NASA-RTLX. The black line represents the average as taken over all subjects. Although there are individual differences, note that the average line shows that the subjects as a group experienced less effort for the no task condition and most for the hard task condition. The easy task is in between.

4.4 ERP components

Note that Subject 9 was removed from the ERP component analysis, due to too much artifacts in the data.

The grand average wave, averaged over all subjects, conditions and electrodes can be seen in Figure 16. Here we can clearly see peaks that seems to be corresponding to N1 (peaking at around 100 ms), P2 (around 225 ms), and novelty P3 (around 310 ms), all marked in the figure and with latencies similar to those found in the literature, see section 2.3.2. The N2 component, between P2 and novelty P3, is not very prominent so we decided to focus on measuring N1, P2, (novelty) P3 and LPP. Since there is no risk of confusion between different versions of the P3 component, we will henceforth denote the novelty P3 as P3, when the extinction is not necessary. Using the grand average over all subjects, conditions and electrodes (shown in Figure 16), we also decided on a latency window around each component to be examined. The latency windows chosen were:
N1: 80-150 ms, P2: 200-250 ms, P3: 250-350 ms, and LPP: 500-650 ms and are also marked in Figure 16. These clusters were then used to determine which electrode clusters that were to be used for measuring each component, by measuring the mean amplitude

over these latency windows for the average over all subjects and condition for each electrode.

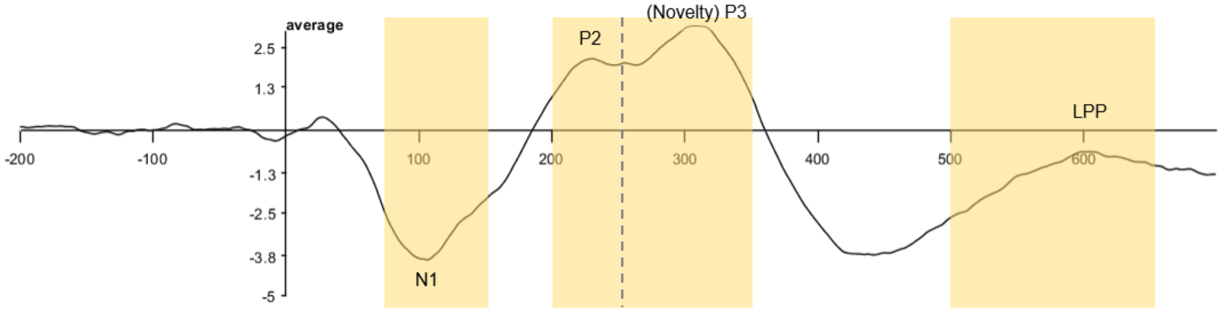


Figure 16. The ERP waveform averaged over all subjects, conditions and channels (electrodes). We can see the components N1 (around 100 ms), P2 (around 225 ms), (novelty) P3 (around 310 ms) and LPP (around 600 ms). The latency windows that were used to determine which clusters of electrodes to use for each component is marked with yellow, and the border between the latency windows for P2 and P3 is marked with a dotted line.

Figure 17 shows how the mean amplitude of the average across subjects and conditions vary over the scalp for each of these latency windows. From this, we have constructed two clusters for each component, one small (five electrodes) and one larger (seventeen electrodes) around the areas where the amplitude is highest for each component. A description of which electrodes that were used for each cluster can be found in Appendix F. As expected from the theory (2.3.2), N1, P2 and P3 are all most prominent in the fronto-central region of the scalp. P3 was expected to be centro-parietal, but is in our study more prominent in the occipital region.

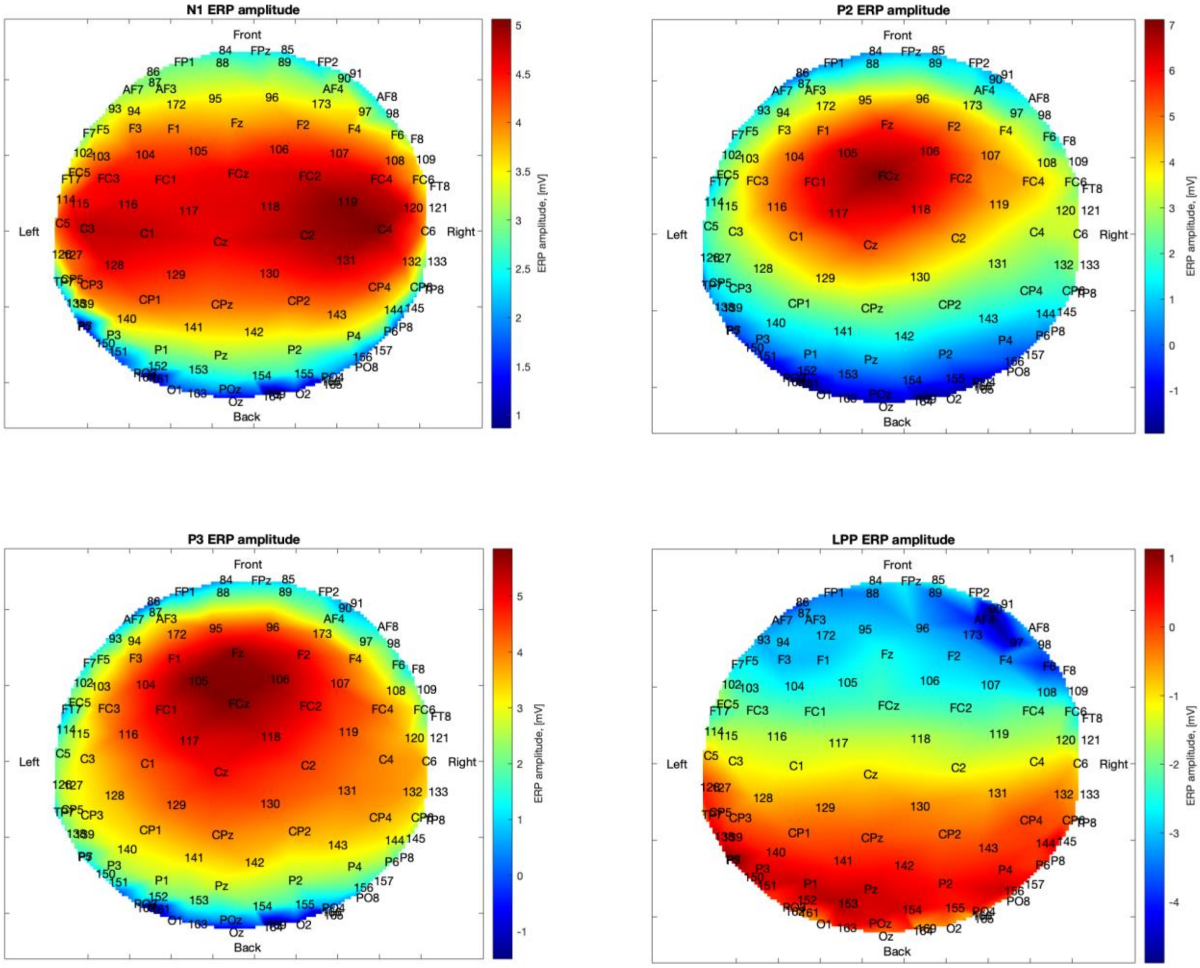


Figure 17. Plots showing the electrodes as distributed over the scalp, as shown from above with the nose facing upwards in the figure. The colors represent the distribution over the scalp of the mean amplitude of the grand average over subjects and conditions for ERP components N1 (80-150 ms), P2 (200-250 ms), (novelty) P3 (250-350 ms), and LPP (500-650 ms). Note that N1, P2 and P3 are all quite symmetric, and most prominent over the frontal/parietal parts of the scalp. LPP is also more or less symmetric, but is most prominent over the occipital areas.

The grand average over all subject at the different clusters can be seen in Figure 18. Since the clusters are chosen around the area where each component is largest, we can see that each component is most prominent in its respective clusters. Here we can see that the amplitude is generally largest for the *hard task* condition, second largest for the *easy task* condition and smallest for the *no task* condition, just as expected.

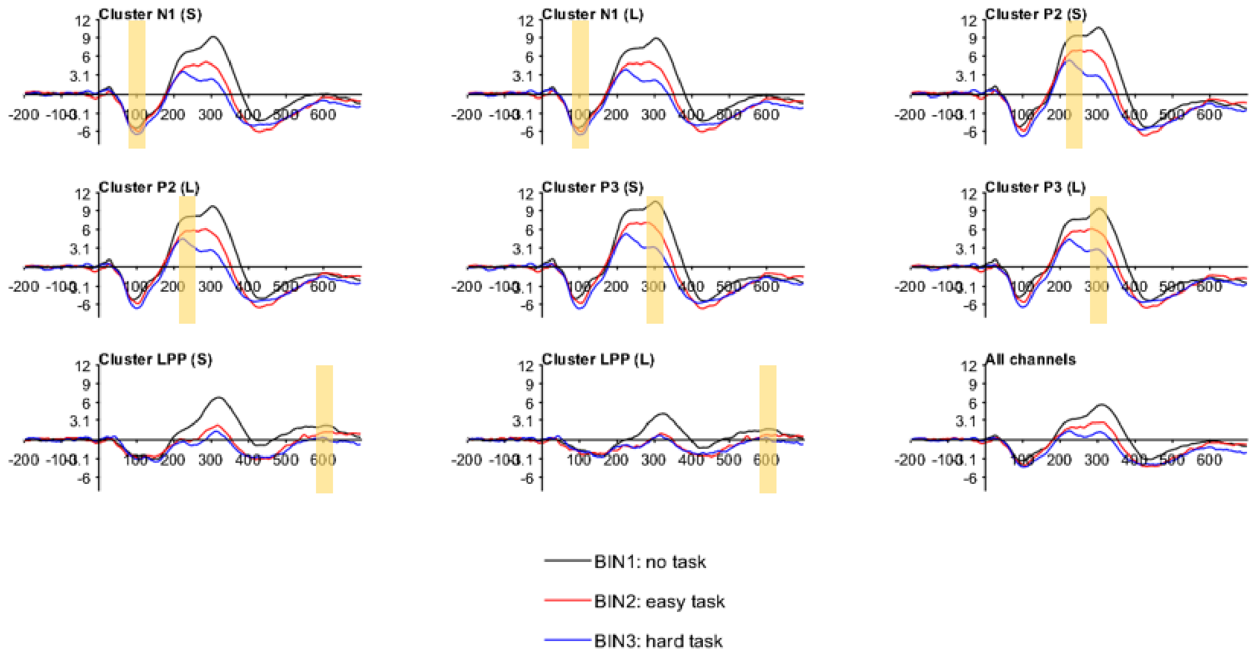


Figure 18. The grand average over all subjects as shown in the different clusters for the different conditions, small (S) and large (L). The clusters have been chosen to best show each component: N1, P2, P3 and LPP. The latency windows that has been used to measure the mean amplitude of each peak within each latency window are marked with yellow. Notice that the amplitudes of the peaks are generally largest for no task and smallest for hard task. The easy task is generally in the middle. In the bottom right corner, we can also see the different conditions when measured over all electrodes.

For each cluster the peak amplitude of the average over subjects and conditions was measured for the relevant component. This resulted in latencies (for the S/L clusters): N1 (102/103 ms), P2 (232/231 ms), P3 (298/300 ms), and LPP (602/600 ms). A latency window of 40 ms centered around these latencies was then used to measure the mean amplitude of each component in the small and large clusters for each subject. The latency windows are marked with yellow in Figure 18. The resulting amplitudes for each condition are shown in Figure 19 (small clusters), and Figure 20 (large clusters). Here we can see that there is some variance between different subjects, but that the mean amplitudes of components P2, P3 and LPP all decrease with increasing task difficulty, in accordance with what was discussed in section 3.2. However, the opposite is seen for the amplitude of N1, that increases with increasing task difficulty. Finally, we constructed the compiled measurements by adding the amplitudes of all measured components: N1, P2, P3, and LPP.

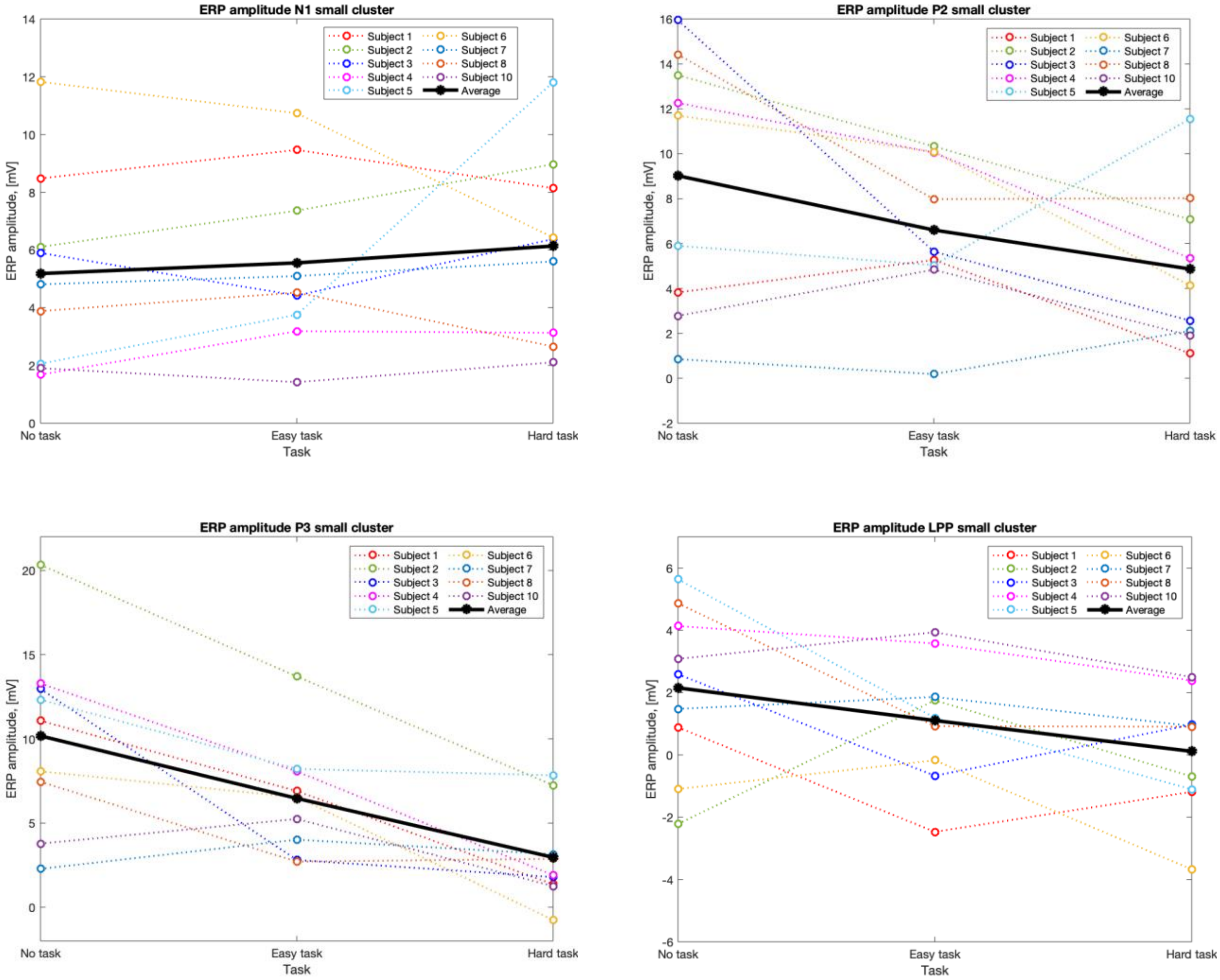


Figure 19. The amplitude of each ERP component (N1, P2, P3, and LPP) for each condition, as measured in the assigned small clusters for each component. The black line represents the average as taken over all subjects. Note that, although there is some variance between different subjects, the mean amplitude for components P2, P3, and LPP all decrease with increasing task difficulty. For N1, the change is in the opposite direction. Note that Subject 9 was removed from the ERP component analysis, due to too much artifacts in the data.

After performing a two-way ANOVA analysis to calculate the p-value for each component and cluster, and the compiled measurement, we could see that only P3 shows a significant effect between all three conditions. The compiled measurement showed a significant difference between *no task* and *hard task.* The results for the small clusters generally gave a smaller p-value than the large clusters, but not always. The resulting ANOVA tables and pairwise comparisons for each component can be found in Appendix E.

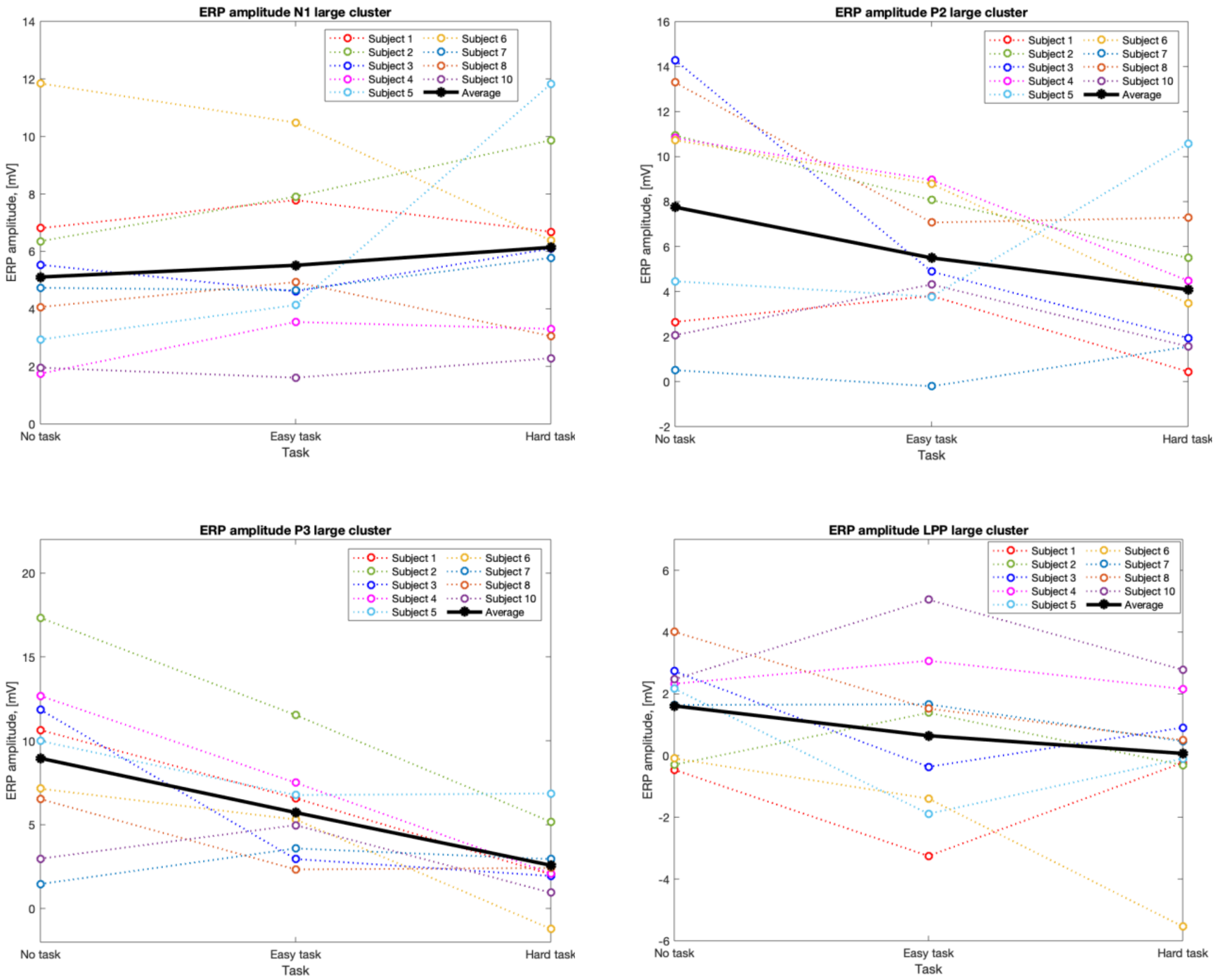


Figure 20. The amplitude of each ERP component (N1, P2, P3, and LPP) for each condition, as measured in the assigned large clusters for each component. The black line represents the average as taken over all subjects. As for the small clusters, there is some variance between different subjects, but the mean amplitude for components P2, P3, and LPP all decrease with increasing task difficulty and N1 changes in the opposite direction. Note that Subject 9 was removed from the ERP component analysis, due to too much artifacts in the data.

4.5 Frequency bands

Note that Subject 9 was removed from the frequency band analysis, due to too much artifacts in the data.

As mentioned in section 3.5.2 we measured the power of the frequency bands in small and large clusters of electrodes. These clusters were constructed by averaging the power over all subjects and conditions and plotting the result for each electrode as a scalp distribution, the result can be seen in Figure 21. We notice that the power scalp distribution for Low-Alpha, High-Alpha and Broadband-Alpha (all mentioned as the Alpha frequency bands from here) are approximately the same, therefore the same clusters are used for all Alpha bands. We constructed one small cluster of five electrodes and one large cluster of seventeen electrodes (sixteen for Alpha, for symmetry reasons) for all examined frequency bands according to the scalp distribution. We then measured the absolute PSD for each condition averaged over the electrodes in each cluster. The result can be seen in Figure 22 and Figure 23. More information about the clusters can be seen in Appendix F. In compliance with the literature (see section 2.3.3), we see that the power of the Alpha bands decrease for an increasing level of difficulty. However, the Theta power seems to also decrease which is not in compliance with the literature.

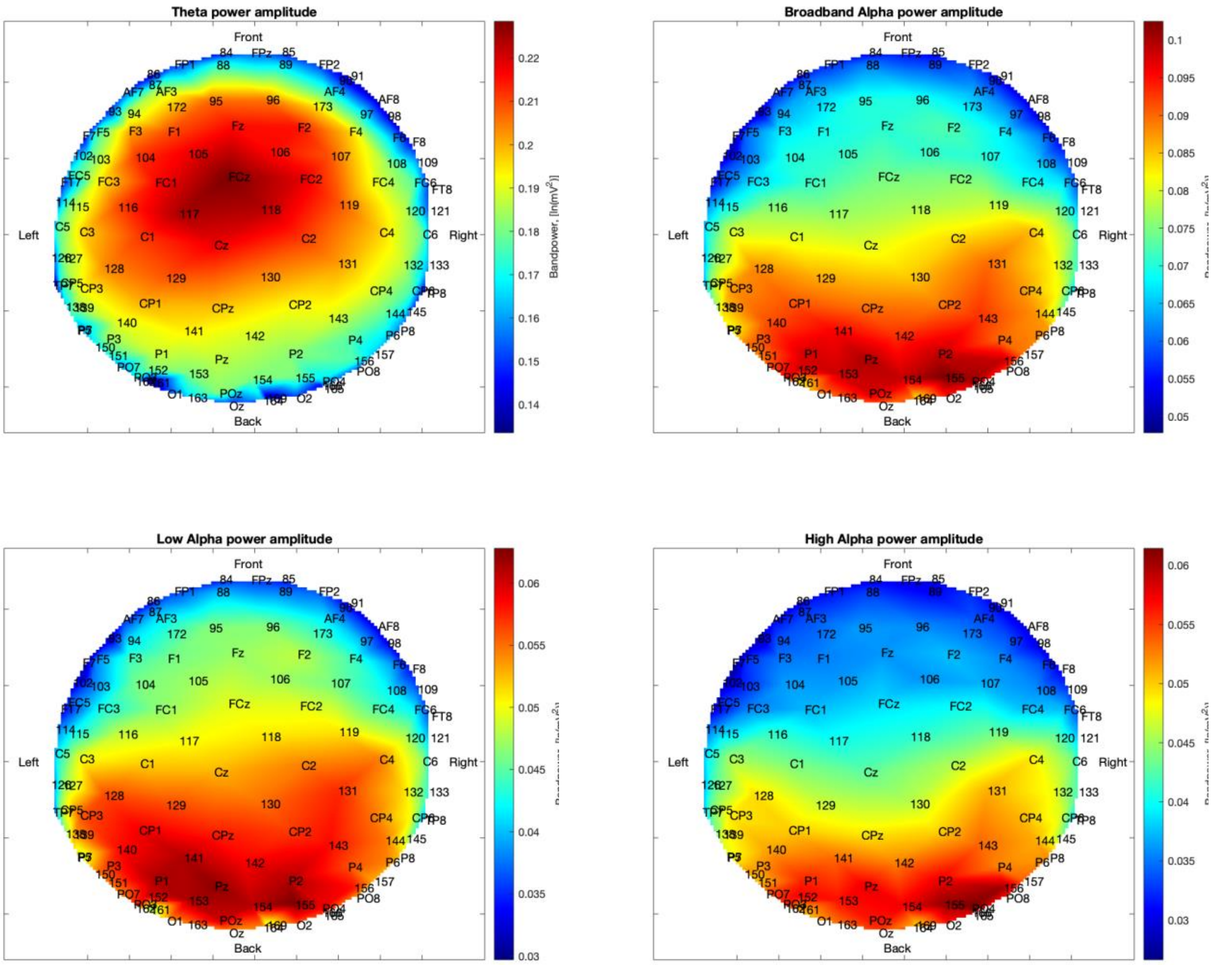


Figure 21. Scalp distributions of the spectral power density, as shown from above with the nose facing upwards in the figure. The colors represent the distribution over the scalp of the mean amplitude of the grand average over subjects and conditions for the frequency bands Theta (3-8 Hz), Broadband-Alpha (8-13 Hz), Low-Alpha (8-10 Hz), and High-Alpha (10-13 Hz). Note that all are quite symmetric, and that the Theta band is most prominent at the frontal/parietal regions of the scalp, while the alpha bands are most prominent over the occipital areas.

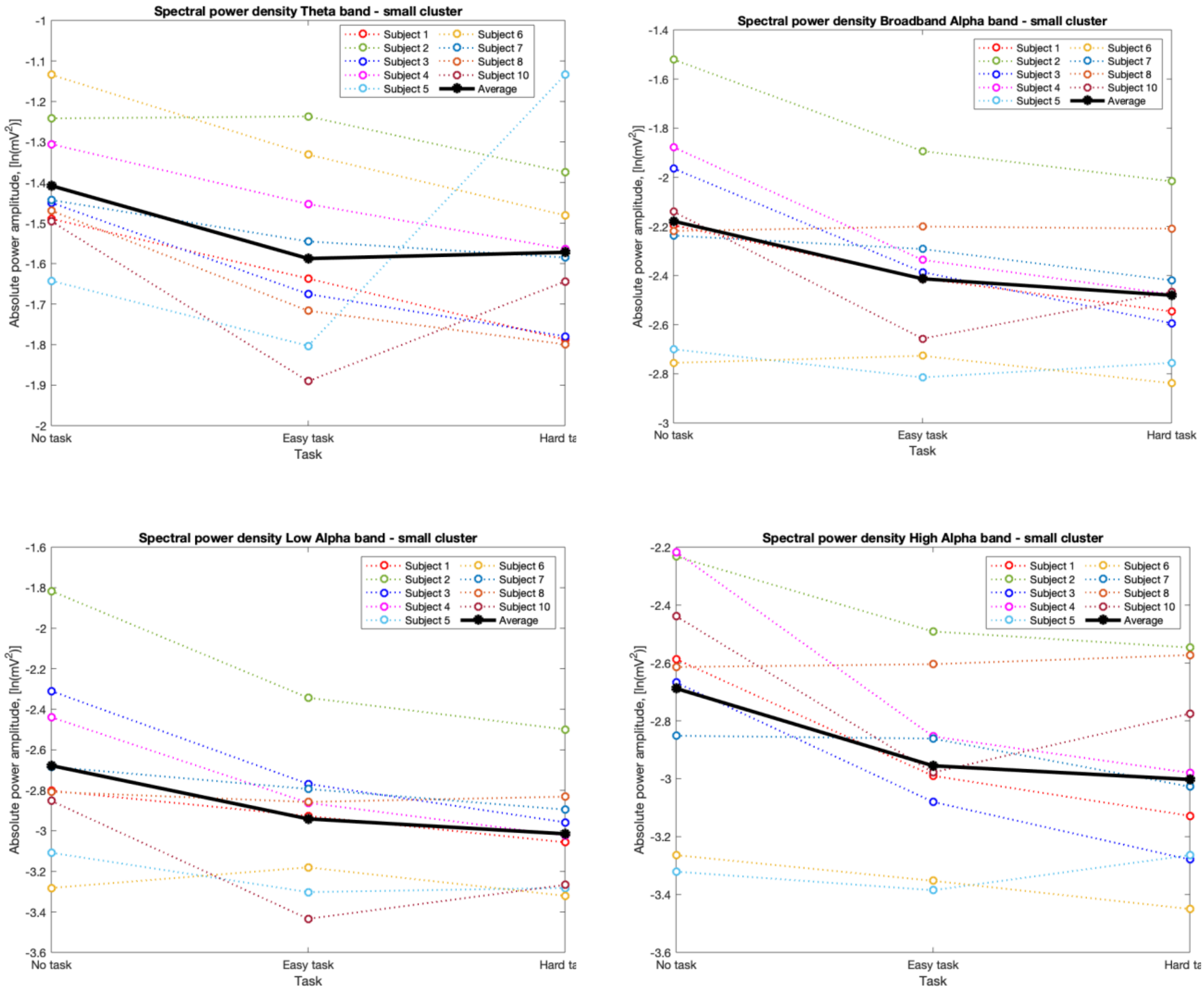


Figure 22. The absolute spectral amplitude of each frequency band: (Theta, Broadband Alpha, Low-alpha and High Alpha) for each condition, as measured in the assigned small clusters for each component. Note that the mean amplitude for all alpha bands (broadband, low and high) decrease with increasing difficulty, as would be expected. Theta, however, decrease between no task and easy task, but then increase slightly for the hard task while it would be expected to increase with increasing workload. Note that Subject 9 was removed from the frequency band analysis, due to too much artifacts in the data.

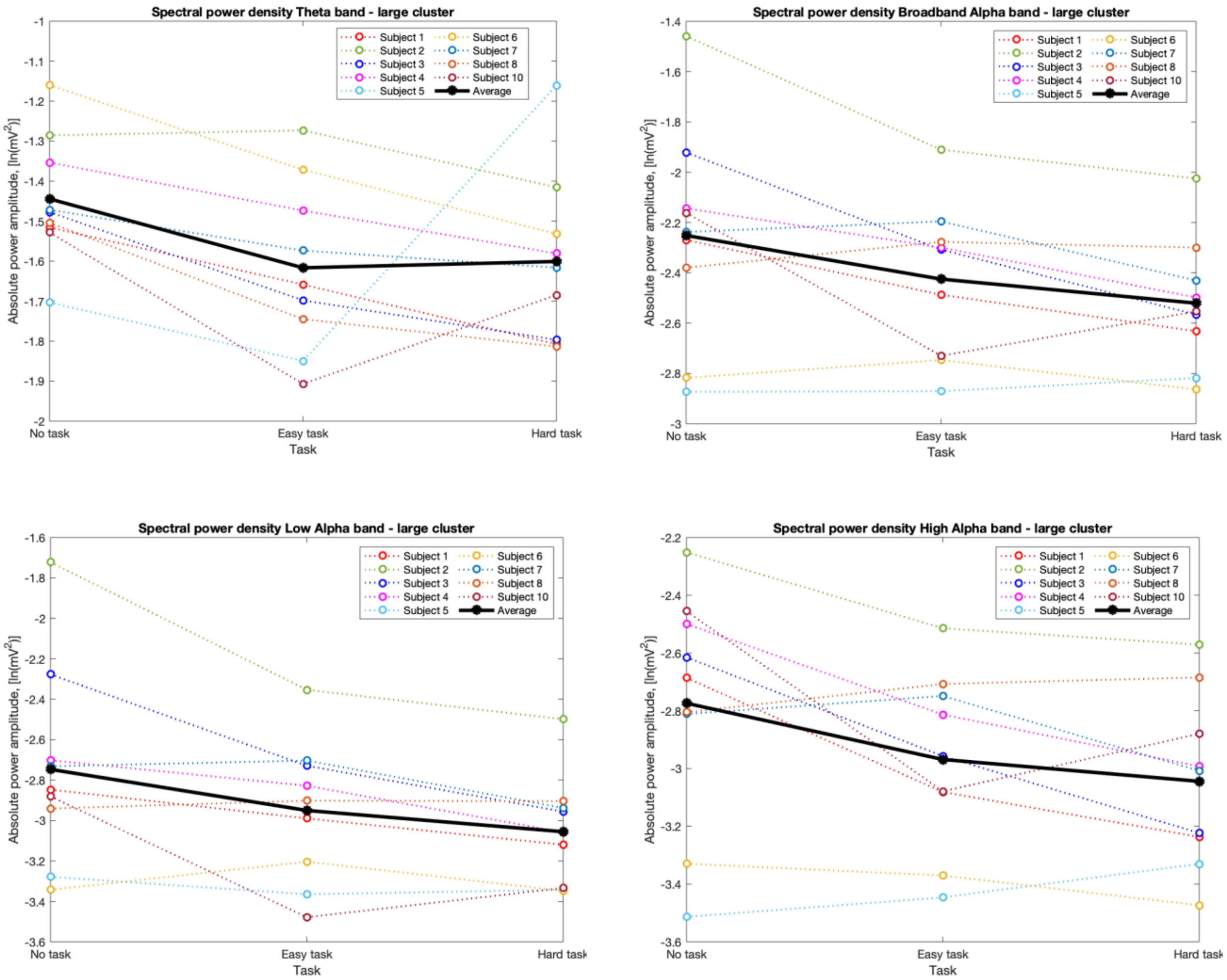


Figure 23. The absolute spectral power amplitude of each frequency band: (Theta, Broadband Alpha, Low-alpha and High Alpha) for each condition, as measured in the assigned large clusters for each component. Note that the mean amplitude for all alpha bands (broadband, low and high) decrease with increasing difficulty, as would be expected. Theta, however, decrease between no task and easy task, but then increase slightly for the hard task while it would be expected to increase with increasing workload. Note that Subject 9 was removed from the frequency band analysis, due to too much artifacts in the data.

A two-way ANOVA analysis revealed that none of the frequency bands showed a significant difference between all conditions. However, all Alpha bands showed significance between some conditions. For the large clusters, the Alpha bands showed a significant difference between *no task* and *hard task*, while *no task* differed significantly from both *easy task* and *hard task* for the small cluster measurements. The full ANOVA tables and pairwise comparisons can be seen in Appendix E.

5 Discussion

Here we will discuss our results, starting with thoughts about the experiment execution. After that we move on to discuss that factors measuring performance and perceived workload seems to indicate that we achieved the wanted difference in difficulty between the conditions. Lastly, we will address the EEG measurements (ERP components and frequency band) and discuss what can be said about the different measurements and cognitive workload. As discussed in section 2.1, learning can be indicated either by a decrease in cognitive workload or an increase of performance, and this will also be discussed.

Since the individual variance for all measurements is quite large, we will only discuss the mean values here. However, we will finish with a discussion about individual differences.

5.1 Experiment execution

As mentioned in section 4.1, there were some flaws involved in the execution of the experiment. However, we still chose to use all available data. This gives more data for the analysis, but also adds to the uncertainty of the results. This, however, only applies to the data measuring the performance.

One flaw was that it was sometimes hard to see when the cube was broken. This was mainly because the cube lit up and went out again very fast sometimes, so fast that it was hard to see. This was partly resolved by reminding the subject to look closely when they did not see as many breaks during a block as the experiment leaders. This also made it hard to count all the breaks and some might have been missed. This means that the counted number of breaks, and thereby the success rate, cannot be completely trusted. With the exception of the very first block of the grasping task, before subjects learnt that they needed to look more closely at the LED, we estimate the error to be below 20 % between the number of breaks counted by us, and the number of breaks seen by the participants. So, all in all, this is not believed to have been a big problem and we still believe that analyzing this data might give some valuable information about the subjects' performance. It is also believed to not have had a big effect on the subjects' effort. It would be possible for the subject to lower the attention when they did not see the cube break, but this is believed to be counteracted by the fact that the subjects now had to focus more to see when the cube broke. However, this problem should be fixed in future experiments using the cube.

As described in section 3.7 the cube was equipped with a Bluetooth-module, with the purpose to be able to send data (such as the number of times the cube was broken) from the cube to a computer. However, when the battery got low the Bluetooth-module was the first thing of the cube's equipment that got instable and stopped working. It was mainly because of the instable behavior of the Bluetooth-module that we also counted the number of times the cube broke manually, but also to get an indication of if the automatic counting worked as we wanted. During the measurements we recognized that the automatic counting of the number of breaks implemented in the cube counted more broken cubes than the experiment leader did. The reason for this we found out to be that the cube counted every time the value from the force-sensors reached above the set force limit. The problem with this is that since the cube updates these values with a baud rate of 9600 (i.e. 9600 times per second) and lights up the cube only for the time the value is above the force limit, this might be for a very short time, so short that the subject and the experiment leader might not see it. This can of course be solved by lighting up the cube for a longer time every time it is broken and count the number of times the cube lights up instead, which would solve the problem mentioned above that it was hard to see when the cube was broken.

During the measurements we also noticed that he output of the cube varied a bit over time and between each measurement session, which means that the force needed to break the cube could have small variations between subjects. However, since the variation was small this didn't seem like a big problem, but still the variation was there. Therefore, more investigation on how the sensors are used needs to be done to increase the robustness of the values the cube outputs and to make sure that all subjects get equally hard tasks. This problem could perhaps be solved by computing a baseline for the output from the sensors when no force is applied to the force sensors and subtract this from the force limit used, before each measurement session. During the calibration of the cube (see Appendix B) we also noticed that if a heavy weight was applied to the force sensors, it took some time for the force sensors to adjust the output values after the weight was removed.

Another problem was the fact that some subjects hit the barrier in some of the lifts, when they did not lift the cube high enough. This should be added to the evaluation of the success rate in future experiments. It is also possible that losing the sensory feedback (in the future study using anesthesia or in using a prosthesis without sensory feedback) might cause the subject to drop the cube, so the number of drops per minute could also be worth counting. However, this was not a problem in our case.

5.2 Performance

The performance has been evaluated using the number of lifts per minute, the success rate of the lifts and the accuracy of the oddball task. It was also collected for each of the three blocks for each condition. That means that this data can be used to compare the different conditions, but also to see differences to indicate learning within each condition. It is also interesting to note the variance in the number of lifts per minute and the success rate of each participant, so this will also be discussed.

5.2.1 Comparison of conditions

As can be seen in Figure 12a, the general trend was that the number of lifts per minute decreased between the *easy* and *hard task*. This was the case for most subjects, to some degree, with the only exception being subject 8. The differences in performance of the subjects as a group indicates that there was a difference in difficulty between the different conditions, as planned.

The same conclusion can be drawn by looking at Figure 13. Here we can see that the subjects generally not only lifted the cube less times per minute for the *hard task* compared to the *easy task*, but also had a lower success rate. The accuracy of the oddball task also increased with increasing difficulty, although the difference between the *easy* and *hard* condition is small. This could indicate that the main problem was adding the lifting task to the oddball task, while the difference between *easy* and *hard* did not affect the ability to count the sounds as much.

Although a greater difference between the *easy* and *hard* grasping task might have been desired, such a change of the method might not improve the results. The reason is that increasing the difficulty of the *hard task* or lowering that of the *easy task* would have increased the risk of some of the subjects reaching boredom or frustration. This would affect their ability to keep focused on the task, as discussed in section 2.1 and 2.3.4.

5.2.2 Assessment of a learning process within each condition

In Figure 12b we take a closer look at lifts per minute by looking at the blocks within each condition. There is some variance between the different subjects, but the average over all subjects might give some information. For the *easy task*, it is quite clear that the subjects in general lifted more as a function of time, here grouped in consecutive blocks. Although we do not know if the workload decreased between the blocks (since we have not analyzed the EEG data for each block), this increase might be indicating a learning process. However, this change cannot be seen during the *hard task* condition. One explanation for this might be that the subjects were tired by then, and perhaps the harder task was more tiresome as well. Another might be that the learning had been saturated and the performance was by then limited by the difficulty of the task or that the easy level represented the optimal level for learning as suggested by Winnie et al. [50], see section 2.1.

For the oddball task, no pattern can be seen between the blocks in each condition, see Figure 14b. Perhaps this can be explained by the fact that this is a risky way of assessing cognitive workload, as discussed in section 3.2.2, since there are several different factors that can affect the count.

5.3 Perceived difficulty of the different conditions

As can be seen in Figure 15 the subjects as a group experienced an increase of difficulty between the conditions *no task*, *easy task*, and *hard task*. The general trend of rising perceived difficulty indicates that the difficulty of different conditions did increase as planned, although it is again arguable if greater differences between *easy* and *hard* would have been desirable, as discussed in section 5.2.1.

It is also worth mentioning that it was impossible to hide our intentions regarding the differences between the levels from the subjects. They all new that adding the grasping task to the oddball task was meant to increase the difficulty, and that adding weight to the cube would further raise the effort of the task. This might have affected their perceived effort.

5.4 ERP components

As discussed in section 3.2, the general idea of using a dual task paradigm is that the amplitude of the components should be greater for an easier task than a hard one. This is because the brain should pay less attention to the sounds of the oddball task if the primary task is more demanding. For both the small and large clusters, Figure 19 and Figure 20, we can see that this is in fact the case for the average of components P2, P3 and LPP. Their mean amplitude is highest for the *no task* condition, lower for the *easy task*, and lowest for the *hard task*. However, for the N1 peak, a slight effect in the opposite direction can be seen.

We also saw that, out of the four ERP components that were measured, only P3 passed below the 5 % threshold and showed a significant difference between all of the conditions. The compiled method (the sum of all mean amplitudes) also proved to give $p<0.05$, but only when comparing the *no task* condition to the *hard task*. The fact that P3 showed the most promise for measuring cognitive workload is not surprising, since it has previously shown to be very effective for this purpose, as discussed in section 2.3.2.4. It is also the biggest component, and therefore needs less epochs to yield a sufficient signal-to-noise ratio, as mentioned in section 3.3.4.

When comparing the small and large clusters, the smaller does generally generate smaller p-values, which might indicate that they are more probable to give significant results. Since the small clusters are more concentrated around the highest peak amplitudes, this was expected. However, with the problem of multiple comparisons (discussed in section 3.6), one needs to be careful when drawing conclusions from this, and therefore these differences only offer an indication of which cluster size might be better than the other. It is also worth considering that a larger cluster makes the method less sensitive to faulty channels, for example if one channel shows large amounts of noise because of a bad connection to the scalp.

For the N1 component, that showed an unexpected pattern, this might be explained by the difference between peaks and components discussed in section 2.3.2. Since the P2 and P3 components both give rise to large, positive peaks close to the latency of the N1 component, it is expected that they would increase the overall amplitude of that latency window. This can explain the fact that the negative N1 amplitude seems to have decreased with increasing workload, since the positive P2 and P3 increased.

The compiled method shows promise to be a useful tool when assessing cognitive workload. As discussed in section 3.5.1.4, it should be a way to emphasize the effect of all components' increasing amplitude for increasing workload. However, with the problem of P2 and P3 overlapping with N1 to create a reverse effect for this component, the method was not as effective as one might have expected.

There also is the factor of how possible fatigue affects the results. As mentioned in section 3.3.4, the three hour study by Trejo et al. [58] showed significant effects for P2, but not for N1 and P3. This, together with our shorter total time and the many breaks where the subjects could decide when they were ready to start again, should mean that fatigue is not a large effect in our measurements. However, for P2 and thereby also the compiled measurement, it is still a factor. Also, we cannot know that our tasks would imply the same level of fatigue as the one studied by Trejo. We would therefore recommend future studies to counterbalance the order of the different conditions across subjects.

5.5 Frequency bands

As discussed in section 2.3.3, several studies have shown that the Alpha power in general decreases when a more difficult task is performed and that the Theta power increases. This means that the Theta/Alpha power ratio is expected to increase with the level of difficulty. In Figure 22 and Figure 23 we can see that the Alpha power does in fact increase for a more difficult task for our measurements for Broadband Alpha (8-13 Hz), Low-Alpha (8-10 Hz), and High-Alpha (10-13 Hz). Averaged over all subjects the alpha power is highest for no task, lower for easy task and lowest for hard task. The two-way ANOVA also showed that the effect was significant for the Alpha bands between some of the conditions. The *no task* differed significantly from both *easy task* and *hard task* for the small clusters, while the large cluster measurements only showed a significant effect between *no task* and *hard task*. We also noticed that neither the Theta power nor the Theta/Alpha ratio follows the pattern from the theory. Neither are these effects significant. If we look closely on the graph of the Theta power we can see a slight increase between the *easy* and *hard task*, however between the *no task* and *easy task* condition we see a clear decrease in power. A possible explanation to this could be the placement of the electrode sites we used for these measures. As explained in section 2.3.3.2 the Theta power is expected to be most prominent at frontal scalp sites. However, in our scalp distribution we got a higher Theta activity at more central sites and the cluster we used for our measures was therefore more centrally distributed on the head. Possibly the result for both Theta power and Theta/Alpha power ratio would be slightly different if measured at only frontal scalp sites. Also, since the studies in our literature review only used one electrode to calculate the power of the frequency bands and the Theta/Alpha ratio, the fact that we are using clusters might also be a reason that we do not get the same results as previous studies. Therefore, a comparison between measuring frequency bands in clusters and measuring them from only one electrode is recommended to study further in a future study. However, the Alpha powers are expected to be most prominent in the occipital areas, which are included in our chosen clusters for the Alpha power measurements. This together with that the resulting Alpha power decreased for higher level of difficulty as expected shows that the clusters we chose to measure the power for were good choices.

As for fatigue, there is a risk that this affected these results, since both Alpha and Theta has been shown to be affected by fatigue in the three hour study by Trejo et al. [58]. With the same reasoning as for the ERP components in section 5.4, the effects here should be smaller because of our experiment design, but we cannot say for sure how big they are. Therefore, this is another reason to recommend a future study to counterbalance the order of the different conditions across subjects.

Another thing that might have an effect on the result for the frequency bands are the heterogeneity of the grasping task. The tasks performed in the studies discussed in section 3.5.2.2 are all motor tasks, but with a more varying behavior than ours. This might be a reason that we do not get the same results as presented in those studies. An alternative way to analyze the data is to compute event-related spectral perturbation (ERSP) instead of PSD, which calculates the event-related changes in spectral power. By computing ERSP for all relevant epochs and averaging the data over all epochs only the frequency content related to the event (in our case the oddball task) will be left, and this might lower the effect of the heterogeneity of the grasping task on the measuring of frequency bands. This kind of analysis have been carried out by Aliakbaryhosseinabadi et al. [61] who performs a study were the participants are to perform 90 trials of ankle dorsiflexion. They showed that attention to a task can be classified from EEG time-frequency features gotten from ERSP. Since their task also have a high heterogeneity it indicates that ERSP might be a good way to analyze the frequency bands for our task in a future study.

5.6 Differences between individuals

All of our results vary between different individuals, and there is always at least one subject that does not follow the pattern shown by the average. These differences could arise from physical or physiological differences, as discussed in section 2.3.4. Here we want to discuss these differences further.

When looking at performance, some difference could be explained by the different abilities of the subjects. However, bigger differences could indicate that the subjects interpreted the task differently and put different weight on the two statements of the task: “lifting the cube as many times as possible” and “without breaking it”. We can see in Figure 13 that the subjects differ in both number of lifts per minute and success rate. Subjects in the upper left corner of this plot have played it safe (many lifts with big success rate) the lower right corner would indicate high risk-taking (many lifts and low success rate). Here we can see that most of the subjects have played it rather safe, or have been successful in their risk-taking (performing many lifts with high success rate). The latter is also a sign that these subjects might have had abilities that made them better suited for the task, while the subject in the lower left corner of the plot have had some trouble (lifting few times with low success rate).

So, we conclude that there was a difference in how the subjects interpreted the task. However, all subjects seemed to understand that they should not break the cube, although there were individual differences between risk taking and abilities. There also might have been some differences in motivation to perform the task as well as possible.

However, even though the room for interpretation might be decreased, it can never be eliminated. One possible way would be to use gamification theory and introduce some kind of points, such that the subject was informed that each lift gave a point but each break took away a number of points. However, there are always differences between people, for example how willing they are to take risks. So, as we see it, these differences always have to be taken into consideration when performing these kinds of experiments, as well as the differences in abilities, engagement to the task and differences and the causes for differences in ERP data, discussed in section 2.3.4.

When looking at the perceived effort, see Figure 15, there were also some subjects that differed from the average pattern by experiencing a decrease of workload between the *easy* and *hard task*. A possible explanation is that to Subject 4, who represents the most extreme case, the mere idea of having to perform two tasks simultaneously might have been difficult. But once having practiced this with the *easy task* it felt easier or similarly easy to do it again, even though the difficulty of the primary task was increased. To explain the results for Subject 2, there is a possibility that they felt that they learnt how to perform the oddball task successfully during the *no task* condition. When they later moved on to adding an easy grasping task, this might not have been seen as an increase of difficulty. Similar patterns can be seen for the accuracy of the oddball task in Figure 14 for Subject 4 and 9 (who made more errors in the easy task then the other two conditions) and Subject 1 and 10 (who made the same number of errors in the *no task* and *easy task* conditions).

6 Conclusions

We will present our conclusions by answering our research questions.

1) *Will the subjects experience the expected difference in difficulty between the different conditions, on group and/or individual level, as indicated by…*
   a) *…the perceived effort, given by the scores on the NASA-RTLX?*
   b) *…the performance, given by number of lifts, success rate and accuracy of the oddball task?*

On a group level, we could establish that both perceived effort and performance reflected the expected difficulty between the different conditions, where the *no task* condition was easiest and the *hard task* hardest, just as planned. However, the individual variance was too big to draw conclusions from individual measurements.

2) *Can the proposed method be used to measure differences in cognitive workload, on group and/or individual level as indicated by…*
   a) *…event-related potential (ERP) components? And if so, which components?*
   b) *…frequency bands? And if so, which frequency bands?*

Several of the proposed EEG measurements shows promise to be an indication of cognitive workload. For ERP data, the (novelty) P3 and the compiled measurement (the sum of the mean amplitude of N1, P2, P3, and LPP) both showed a significant difference between the conditions while N1, P2, and LPP did not show significant effects. However, only P3 was able to show a significant difference between *easy task* and *hard task*. This is important for the future study comparing workload *with* and *without sensory feedback*, since these conditions where here simulated with the *easy* and *hard* conditions. However, we also have no idea of knowing if the difference between our *easy* and *hard* task are smaller or larger than the difference between *with* and *without sensory feedback.* The compiled measurement shows promise to be a new way of assessing cognitive workload and might be of use if more subjects were examined.

For the frequency bands, we could see some promising result for indication of cognitive workload. The Alpha power for both Low-Alpha (8-10 Hz), High-Alpha (10-13 Hz) and Broadband Alpha (8-13 Hz) frequency bands showed a significant difference between some level, such that the power decreased for higher demanding tasks. For the large clusters there was a significant effect between *no task* compared to *hard task,* while the small clusters showed significance between *no task* compared to both *easy task* and *hard task*. The Alpha frequency bands thereby show promise to be of use, but more subjects would be needed to see a significant effect between all levels. There was, however, no significant effect for either Theta (3-8 Hz) nor the quotient Theta/Alpha.

None of these measurements seems reliable for individual analysis.

3) *Which latency windows (for ERP components only) and electrode sites should be used to examine the differences in cognitive workload with ERP components and frequency bands?*

The latency window for each ERP component (for small/large clusters) were 40 ms wide and centered around: N1 (102/101 ms), P2 (232/231 ms), P3 (298/300 ms), and LPP (602-600 ms). They N1, P2 and P3 components were all measured in the fronto-central region of the scalp, although the exact locations of the clusters differ between them. LPP was measured in the occipital region. When comparing the p-values of the small and large clusters, there seems to be a slightly greater chance of significance when using a smaller cluster. However, both gave similar results, and using smaller clusters also means that the measurement is more sensitive to faulty channels. That concludes research question 3) for ERP components, by giving a recommendation for latency windows and scalp regions for all components. We also provided more information to decide on the size of cluster, although we cannot give a clear answer of what is best.

The Alpha frequency bands were all measured over the occipital scalp region, and Theta over the fronto-central region. The smaller clusters showed a significant effect between *no task* compared to both *easy task* and *hard task for Alpha bands*, which means that the small clusters again show greater promise to measure cognitive workload. However, as for the ERP components, the sensitivity of the method should be considered when choosing cluster size. Also, further investigation needs to be done to see if clusters gives comparable results to studies that used only one electrode to measure the frequency bands.

4) *Can a learning effect be observed during each condition by comparing the performance for each of the three blocks?*

We have been able to see some signs of a learning process within each condition, by comparing the number of lifts between the blocks. Accuracy in the oddball test did, however, not show any trend between the blocks.

Also, we would recommend a future study to counterbalance the order of the different conditions across subjects, to assure that fatigue does not affect the results.

7 Future work

Here we will provide some ideas for future work, to further investigate the concept of cognitive workload.

7.1 Experiment with anesthesia

We have, by this study, provided a starting point for performing a similar study using anesthesia to mimic the loss of sensory feedback that many prosthetic users experience. That way, it might be possible to show the cognitive benefits of cognitive workload, and thereby facilitate the development of this technique. Aside from using the results and conclusions of this work, there is also a possibility to further investigate and analyze the data. Since gathering data for EEG analysis is very time consuming, this could be a way to investigate other methods without having to conduct more measurements. However, for future studies the order of the conditions should be counterbalanced across subjects, to minimize the effects of fatigue.

7.2 Possible improvements of the cube

As discussed above there were some technical problems with the force sensitive cube during the measurements and there are several things that could be improved. To begin with the counting of how many times the cube is broken, together with the lighting of the red LED bar has room for improvement. As it is now, the cube counts every time the force on one of the sensors rises above a set limit. However, this could be for a so short time that the subject does not have the possibility to see the LED bar light up. This could be solved by updating the algorithm for when the cube counts a break so that it is based on a larger amount of values from the sensors or such that it does not count more than once if the cube is broken several times within a short time range. Also, the LED bar can be set to light up for a fixed amount of time every time the cube is broken. The design of the cube can also be improved in terms of size and weight distribution. Space in the cube can be gained by developing a circuit card for the electronics and using a smaller battery. The weight distribution of the cube can be improved by using plane weights put in the bottom of the cube, instead of the box used now which only spreads over approximately half of the cube. A more even weight will keep the cube from rotating when the subjects lifts it and thereby make it easier to lift.

Another improvement worth working on is to install a load cell at the bottom of the cube. This could both be used to measure the vertical lifting force, as well as for counting the number of times the cube is lifted (which was counted manually in our study).

7.3 Examination of the usefulness of EEG to evaluate learning processes

It is possible that the EEG data could be used to evaluate differences between the different blocks as well. This could make it possible to further investigate the process of for example learning to use a prosthesis. However, this would lower the number of epochs for each measurement, since each block only contains about a third of the epochs used for each condition. Using a combination of all the assessments made in this work might make it possible to still deduct information from less data, according to a suggestion by Brouwer et al. 2012 [40]. This would be one step further compared to the compiled ERP measurement calculated in this work. Enabling this kind of analysis could lead to a better understanding of how the physical therapy should be executed to best help patients learn to use a prosthetic hand.

Bibliography

[1] Integrum, “e-OPRA.” [Online]. Available: https://integrum.se/opra-implant-system/e-opra/. [Accessed: 11-Feb-2020].

[2] E. L. CLAESSEN, “Bättre livskvalitet med protes som fästs direkt till lårbenet,” *SAHLGRENSKA AKADEMIN*, 2019. [Online]. Available: https://sahlgrenska.gu.se/forskning/aktuellt/nyhet//battre-livskvalitet-med-protes-som-fasts-direkt-till-larbenet.cid634566. [Accessed: 06-May-2020].

[3] I. AB, “Integrum’s history.” [Online]. Available: https://integrum.se/about-us/integrum-history/. [Accessed: 06-May-2020].

[4] M. G. KUBISTA, “Första fingerfärdiga och kännande handprotesen implanterad,” *SAHLGRENSKA AKADEMIN*, 2019. [Online]. Available: https://sahlgrenska.gu.se/forskning/aktuellt/nyhet/forsta-fingerfardiga-och-kannande-handprotesen-implanterad.cid1615603. [Accessed: 06-May-2020].

[5] Biomechatronics and Neurorehabilitation Laboratory, “Natural control of artificial limb through an osseointegrated implant,” 2017. [Online]. Available: http://www.chalmers.se/en/projects/Pages/Natural-control-of-artificial-limb.aspx. [Accessed: 11-Feb-2020].

[6] M. G. KUBISTA, “Första fingerfärdiga och kännande handprotesen implanterad,” *SAHLGRENSKA AKADEMIN*, 2019. .

[7] L. Berntsson, “Presentation: Evaluation of cognitive workload using EEG, Investigation of how sensory feedback improves function of osseo-neuromuscular upper limb prostheses,” 2019.

[8] M. OrtizCatalan, E. Mastinu, P. Sassu, O. Aszmann, and R. Brånemark, “Self-contained neuromusculoskeletal arm prostheses,” *N. Engl. J. Med.*, vol. 382, no. 18, pp. 1732–1738, 2020, doi: 10.1056/NEJMoa1917537.

[9] Etikprövningsmyndigheten, “Vad säger lagen?” [Online]. Available: https://etikprovningsmyndigheten.se/for-forskare/vad-sager-lagen/. [Accessed: 12-Feb-2020].

[10] Etikprövningsmyndigheten, “Etikprövning - så går det till.” [Online]. Available: https://etikprovningsmyndigheten.se/for-forskare/sa-gar-det-till/. [Accessed: 12-Feb-2020].

[11] S. Puma, N. Matton, P. V. Paubel, É. Raufaste, and R. El-Yagoubi, “Using theta and alpha band power to assess cognitive workload in multitasking environments,” *Int. J. Psychophysiol.*, vol. 123, no. October 2017, pp. 111–120, 2018, doi: 10.1016/j.ijpsycho.2017.10.004.

[12] W. Zhang *et al.*, “Cognitive workload in conventional direct control vs. pattern recognition control of an upper-limb prosthesis,” *2016 IEEE Int. Conf. Syst. Man, Cybern. SMC 2016 - Conf. Proc.*, pp. 2335–2340, 2017, doi: 10.1109/SMC.2016.7844587.

[13] O. Palinko, A. L. Kun, A. Shyrokov, and P. Heeman, “Estimating cognitive load using remote eye tracking in a driving simulator,” *Eye Track. Res. Appl. Symp.*, pp. 141–144, 2010, doi: 10.1145/1743666.1743701.

[14] K. J. Jaquess *et al.*, “Empirical evidence for the relationship between cognitive workload and attentional reserve,” *Int. J. Psychophysiol.*, vol. 121, no. September, pp. 46–55, 2017, doi: 10.1016/j.ijpsycho.2017.09.007.

[15] J. Gonzalez, H. Soma, M. Sekine, and W. Yu, “Psycho-physiological assessment of a prosthetic hand sensory feedback system based on an auditory display: A preliminary study,” *J. Neuroeng. Rehabil.*, vol. 9, no. 1, pp. 1–14, 2012, doi: 10.1186/1743-0003-9-33.

[16] K. Knaepen *et al.*, “Psychophysiological response to cognitive workload during symmetrical, asymmetrical and dual-task walking,” *Hum. Mov. Sci.*, vol. 40, pp. 248–263, 2015, doi: 10.1016/j.humov.2015.01.001.

[17] S. Deeny, C. Chicoine, L. Hargrove, T. Parrish, and A. Jayaraman, “A simple ERP method for quantitative analysis of cognitive workload in myoelectric prosthesis control and human-machine interaction,” *PLoS One*, vol. 9, no. 11, 2014, doi: 10.1371/journal.pone.0112091.

[18] M. W. Miller, J. C. Rietschel, C. G. McDonald, and B. D. Hatfield, “A novel approach to the physiological measurement of mental workload,” *Int. J. Psychophysiol.*, vol. 80, no. 1, pp. 75–78, 2011, doi: 10.1016/j.ijpsycho.2011.02.003.
[19] J. C. Rietschel *et al.*, “Psychophysiological support of increasing attentional reserve during the development of a motor skill,” *Biol. Psychol.*, vol. 103, pp. 349–356, 2014, doi: 10.1016/j.biopsycho.2014.10.008.
[20] S. G. Hart, “NASA-task load index (NASA-TLX); 20 years later,” *Proc. Hum. Factors Ergon. Soc.*, pp. 904–908, 2006, doi: 10.1177/154193120605000909.
[21] S. Mun, M. Whang, S. Park, and M. C. Park, “Effects of mental workload on involuntary attention: A somatosensory ERP study,” *Neuropsychologia*, vol. 106, no. February 2016, pp. 7–20, 2017, doi: 10.1016/j.neuropsychologia.2017.08.021.
[22] I. M. Shuggi, P. A. Shewokis, J. W. Herrmann, and R. J. Gentili, “Changes in motor performance and mental workload during practice of reaching movements: a team dynamics perspective,” *Exp. Brain Res.*, vol. 236, no. 2, pp. 433–451, 2018, doi: 10.1007/s00221-017-5136-8.
[23] L. Berntsson, “Evaluation of cognitive workload using EEG Investigation of how sensory feedback improves function of osseo-neuromuscular upper limb prostheses,” 2019.
[24] E. P. Shaw *et al.*, “Measurement of attentional reserve and mental effort for cognitive workload assessment under various task demands during dual-task walking,” *Biol. Psychol.*, vol. 134, no. December 2016, pp. 39–51, 2018, doi: 10.1016/j.biopsycho.2018.01.009.
[25] B. H. Kantowitz, “Attention and Mental workload,” pp. 456–459, 1998.
[26] R. A. Magill and D. I. Andersson, “Attention as a Limited Capacity Resource,” in *Motor Learning and Control, Concepts and Applications*, 8th ed., McGraw-Hill, 2007, pp. 194–220.
[27] M. W. Miller, J. C. Rietschel, C. G. McDonald, and B. D. Hatfield, “A novel approach to the physiological measurement of mental workload,” *Int. J. Psychophysiol.*, vol. 80, no. 1, pp. 75–78, 2011, doi: 10.1016/j.ijpsycho.2011.02.003.
[28] W. K. Y. So, S. W. H. Wong, J. N. Mak, and R. H. M. Chan, “An evaluation of mental workload with frontal EEG,” *PLoS One*, vol. 12, no. 4, pp. 1–18, 2017, doi: 10.1371/journal.pone.0174949.
[29] E. P. Núñez Castellar, J. N. Antons, D. Marinazzo, and J. Van Looy, “Mapping attention during gameplay: Assessment of behavioral and ERP markers in an auditory oddball task,” *Psychophysiology*, vol. 56, no. 7, pp. 1–13, 2019, doi: 10.1111/psyp.13347.
[30] L. Sörnmo and P. Laguna, “The Electroencephalogram- A Brief Background,” in *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, Elsevier Science & Technology, 2005, pp. 25–54.
[31] S. J. Luck, *An Introduction to the Event-Related Potential Technique*. MIT Press, 2014.
[32] J. Cionek, “Wet, dry, active and passive electrodes. What are they, and what to choose?,” *Brain Support*, 2020. [Online]. Available: https://www.brainlatam.com/blog/wet-dry-active-and-passive-electrodes.-what-are-they-and-what-to-choose-413. [Accessed: 14-May-2020].
[33] G. H.Klem, H. O. Lüders, H. H. Jasper, and C. Elger, “The ten-twenty electrode system of the International Federation,” in *Recommendations for the Practice of Clinical Neurophysiology: Guidelines of the International Federation of Clinical Physiology (EEG Suppl. 52)*, Elsevier Science B.V, 1999, pp. 3–7.
[34] C. Wickens, A. Kramer, L. Vanasse, and E. Donchin, “Performance of Concurrent Tasks : A Psychophysiological Analysis of the Reciprocity of Information-Processing Resources Author (s): Christopher Wickens , Arthur Kramer , Linda Vanasse and Emanuel Donchin Reviewed work (s): Source : Science , New Serie,” *Science (80-.).*, vol. 221, no. 4615, pp. 1080–1082, 1983.
[35] E. S. Kappenman and S. J. Luck, Eds., *The Oxford Handbook of Event-related Potenial*. Oxford University Press, Inc., 2011.
[36] A. F. Kramer, L. J. Trejo, and D. Humphrey, “Assessment of mental workload with task-irrelevant auditory probes,” *Biol. Psychol.*, vol. 40, no. 1–2, pp. 83–100, 1995, doi: 10.1016/0301-0511(95)05108-2.

[37] B. Z. Allison and J. Polich, “Workload assessment of computer gaming using a single-stimulus event-related potential paradigm,” *Biol. Psychol.*, vol. 77, no. 3, pp. 277–283, 2008, doi: 10.1016/j.biopsycho.2007.10.014.
[38] F. B. Dyke *et al.*, “The efficacy of auditory probes in indexing cognitive workload is dependent on stimulus complexity,” *Int. J. Psychophysiol.*, vol. 95, no. 1, pp. 56–62, 2015, doi: 10.1016/j.ijpsycho.2014.12.008.
[39] P. Ullsperger, G. Freude, and U. Erdmann, “Auditory probe sensitivity to mental workload changes - An event-related potential study,” *Int. J. Psychophysiol.*, vol. 40, no. 3, pp. 201–209, 2001, doi: 10.1016/S0167-8760(00)00188-4.
[40] A. M. Brouwer, M. A. Hogervorst, J. B. F. Van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, “Estimating workload using EEG spectral power and ERPs in the n-back task,” *J. Neural Eng.*, vol. 9, no. 4, 2012, doi: 10.1088/1741-2560/9/4/045008.
[41] J. B. Isreal, C. D. Wickens, G. L. Chesney, and E. Donchin, “Event-Related Brain Potential As an Index of Display-Monitoring Workload.,” *Hum. Factors*, vol. 22, no. 2, pp. 211–224, 1980, doi: 10.1177/001872088002200210.
[42] J. Song, D. Zhuang, G. Song, and X. Wanyan, “Pilot mental workload measurement and evaluation under dual task,” *Proc. - 2011 4th Int. Conf. Biomed. Eng. Informatics, BMEI 2011*, vol. 2, no. 2010, pp. 809–812, 2011, doi: 10.1109/BMEI.2011.6098376.
[43] S. Debener, S. Makeig, A. Delorme, and A. K. Engel, “What is novel in the novelty oddball paradigm? Functional significance of the novelty P3 event-related potential as revealed by independent component analysis,” *Cogn. Brain Res.*, vol. 22, no. 3, pp. 309–321, 2005, doi: 10.1016/j.cogbrainres.2004.09.006.
[44] A. L. Pruziner *et al.*, “Biomechanical and neurocognitive performance outcomes of walking with transtibial limb loss while challenged by a concurrent task,” *Exp. Brain Res.*, vol. 237, no. 2, pp. 477–491, 2019, doi: 10.1007/s00221-018-5419-8.
[45] R. J. Gentili *et al.*, “Combined assessment of attentional reserve and cognitive-motor effort under various levels of challenge with a dry EEG system,” *Psychophysiology*, vol. 55, no. 6, pp. 1–17, 2017, doi: 10.1111/psyp.13059.
[46] J. C. Rietschel, M. W. Miller, R. J. Gentili, R. N. Goodman, C. G. McDonald, and B. D. Hatfield, “Cerebral-cortical networking and activation increase as a function of cognitive-motor task difficulty,” *Biol. Psychol.*, vol. 90, no. 2, pp. 127–133, 2012, doi: 10.1016/j.biopsycho.2012.02.022.
[47] G. G. Knyazev, “Motivation, emotion, and their inhibitory control mirrored in brain oscillations,” *Neurosci. Biobehav. Rev.*, vol. 31, no. 3, pp. 377–395, 2007, doi: 10.1016/j.neubiorev.2006.10.004.
[48] J. N. Mak and R. H. M. Chan, “Evaluation of Mental Workload in Visual- - Motor Task : Spectral Analysis of Single- - Channel Frontal EEG,” *IECON 2013 - 39th Annu. Conf. IEEE Ind. Electron. Soc.*, pp. 8426–8430, 2013, doi: 10.1109/IECON.2013.6700546.
[49] W. O. Tatum, “Normal EEG,” in *Handbook of EEG Interpretation*, Demos Medical Publishing, 2014, pp. 1–56.
[50] W. K. Y. So, S. W. H. Wong, J. N. Mak, and R. H. M. Chan, “An evaluation of mental workload with frontal EEG,” *PLoS One*, vol. 12, no. 4, pp. 1–18, 2017, doi: 10.1371/journal.pone.0174949.
[51] D. L. Schomer and F. L. da Silva, Eds., *Niedermeyer’s Electroencephalography : Basic Principles, Clinical Applications, and Related Fields*, 6th ed. Wolters Kluwer Health, 2010.
[52] J. N. Mak, R. H. M. Chan, and S. W. H. Wong, “Spectral modulation of frontal EEG activities during motor skill acquisition: Task familiarity monitoring using single-channel EEG,” *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, no. 1962, pp. 5638–5641, 2013, doi: 10.1109/EMBC.2013.6610829.
[53] J. Polich and A. Kok, “Cognitive and biological determinants of P300 : an integrative review,” vol. 41, pp. 103–146, 1995.
[54] F. Clemente, M. D’Alonzo, M. Controzzi, B. B. Edin, and C. Cipriani, “Non-Invasive, Temporally Discrete Feedback of Object Contact and Release Improves Grasp Control of

Closed-Loop Myoelectric Transradial Prostheses," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 12, pp. 1314–1322, 2016, doi: 10.1109/TNSRE.2015.2500586.
[55] P. E. Castellar Núñez, J.-N. Antons, and D. Marianazzo, "Mapping attention during gameplay : Assessment of behavioral and ERP markers in an auditory oddball task," no. October 2018, pp. 1–13, 2019, doi: 10.1111/psyp.13347.
[56] M. Fabiani, V. A. Kazmerski, Y. M. Cycowicz, and D. Friedman, "Naming norms for brief environmental sounds: Effects of age and dementia," *Psychophysiology*, vol. 33, pp. 462–475, 1996.
[57] S. G. Hart and L. E. Staveland, *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*. Elsevier Science Publishers B.V., 1988.
[58] L. J. Trejo, R. Kochavi, K. Kubitz, L. D. Montgomery, R. Rosipal, and B. Matthews, "Measures and Models for Predicting Cognitive Fatigue," vol. 5797, pp. 105–115, 2005, doi: 10.1117/12.604286.
[59] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, 2004, doi: 10.1016/j.jneumeth.2003.10.009.
[60] J. Lopez-Calderon and S. J. Luck, "ERPLAB: An open-source toolbox for the analysis of event-related potentials.," *Front. Hum. Neurosci.*, no. 8, 2014.
[61] S. Aliakbaryhosseinabadi, E. N. Kamavuako, N. Jiang, D. Farina, and N. Mrachacz-Kersting, "Classification of EEG signals to identify variations in attention during motor task execution," *J. Neurosci. Methods*, vol. 284, pp. 27–34, 2017, doi: 10.1016/j.jneumeth.2017.04.008.
[62] S. Windmann and H. Hill, "Dissociating electrophysiological correlates of subjective, objective, and correct memory in investigating the emotion-induced recognition bias," *Conscious. Cogn.*, vol. 29, pp. 199–211, 2014, doi: 10.1016/j.concog.2014.08.010.
[63] B. L. E. E. Drisdelle and S. Ebrina, "Dealing with ocular artifacts on lateralized ERPs in studies of visual-spatial attention and memory : ICA correction versus epoch rejection," vol. 54, pp. 83–99, 2017, doi: 10.1111/psyp.12675.
[64] S. Kraftnät, "The control room." [Online]. Available: https://www.svk.se/en/national-grid/the-control-room/. [Accessed: 10-Jun-2020].
[65] A. Hyvärinen and E. Oja, "Independent component analysis: A tutorial," *Notes Int. Jt. Conf. Neural Networks (IJCNN'99), Washingt. DC*, vol. 1, pp. 1–30, 1999, doi: 10.1093/mnras/stv2744.
[66] M. X. Cohen, "EEG Artifacts: Their Detection, Influence, and Removal," in *Analyzing Neural Time Series Data*, The MIT Press, 2014.
[67] L. Pion-Tonachini, "ICLabel Tutorial: EEG Independent Component Labeling." [Online]. Available: https://labeling.ucsd.edu/tutorial/labels. [Accessed: 06-May-2020].
[68] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, 2011, doi: 10.1111/j.1469-8986.2010.01061.x.
[69] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, "ICLabel: An automated electroencephalographic independent component classifier, dataset, and website," *Neuroimage*, vol. 198, pp. 181–197, 2019, doi: 10.1016/j.neuroimage.2019.05.026.
[70] J. Lopez-Calderon and S. J. Luck, "ERPLAB: An open-source toolbox for the analysis of event-related potentials," *Front. Hum. Neurosci.*, vol. 8, no. 1 APR, pp. 1–14, 2014, doi: 10.3389/fnhum.2014.00213.
[71] T. C. Handy, "Basic principles of ERP quantification," in *Event- related Potentials: A Methods Handbook*, MIT Press, 2005, pp. 33–56.
[72] K. J. Jaquess *et al.*, "Changes in Mental Workload and Motor Performance Throughout Multiple Practice Sessions Under Various Levels of Task Difficulty," *Neuroscience*, vol. 393, pp. 305–318, 2018, doi: 10.1016/j.neuroscience.2018.09.019.
[73] A. Holm, K. Lukander, J. Korpela, M. Sallinen, and K. M. I. Müller, "Estimating brain load from the EEG," *ScientificWorldJournal.*, vol. 9, pp. 639–651, 2009, doi: 10.1100/tsw.2009.83.

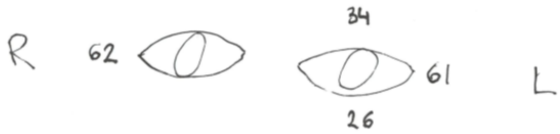
Appendix A: Experiment protocol

Before participants arrives

- Put the battery on charge
- Set up EEG equipment
- Plug in g.TRIGbox to g.GIamp, laptop and sound
- Connect the Krios
- Open audio script. Set search path. Run it to make sure it is working correctly.
- Set up the grasping task: cube, weights and barrier
- Make sure there are enough copies of the self-report, and a pen
- Prepare two copies of the two different *Informed consent* (for the variation study and for resting state). One for the participant to keep and one with the signature for the lab
- Prepare *Photo agreement* to obtain permission to take pictures and videos of the participant
- Prepare Waterloo form
- Open the picture of the cross on the laptop and make sure that it is on **soundless**
- Open g.Recorder software. Admin mode. Make sure you are recording from all the 128 electrodes at 2400 Hz. No filters.
- Prep syringes and paper towels
- Take out shampoo and a clean towel
- Reset cube and count-button
- Untangle the EOG electrodes and put adhesive labels on them

Prepare the participant

- Ask the patient to read and sign the informed consent. Answer possible questions
- Ask the participant to sign the photo agreement
- Ground yourself and the participant before starting
- Fit the participant with cap
    - Put on the cap. Measure the distances between nasion and inion and the two preauricular points with your fingers to make sure that Cz is correctly placed in the center
    - Insert gel into electrodes 1-128 (26, 34, 61 and 62 excluded) and check impedances. They should all be green or yellow
    - Clip 63 and 64 at ear lobes with some gel. Check impedances.
    - Add eye electrodes, insert gel and check impedances.



    - Take a **screenshot** of the impedance page and save as “YYYYMMDD Name before”
    - Take pictures of the participant fitted with the cap from different angles. Channel numbers should be visible

- Do electrode digitization with Krios.
    - Make sure electrode cables are not covering the reflective labels of the electrodes
    - You should scan a total of 133 electrodes
    - Probe the anatomical landmarks (nasion, left and right auricular points)
    - Probe additional points at known location to make the labelling easier
        - Cz
        - T8
        - T7
        - FPz
        - (POz)
    - Export the document on the desktop without labels (for backup reasons, in case you delete something during the labelling)
    - Label the probed points
    - Apply template. Check that the labels seems to be ok
    - Save as “Name Labelled”
- Ask the participant to close the eyes in order to see alpha, chew in order to see EMG and move the eyes around to see EOG. Also make sure the channels are recording independent signals.

Before the tasks

- Make sure the participant is sitting comfortably, that their phone is on soundless and that the patient can reach everything
- Secure the sound cables to the table with tape
- Instruct the participants of the tasks
    - Explain that the tasks are divided into 3 trials, and each trial consists of 3 blocks
        - Each block is ~4 minutes
        - At least one minute break between each block
        - A sound will be played at the beginning and end of each block
        - ~5 min break between each trial
        - Lastly: Resting state
            - 7 minutes closed eyes
            - 7 minutes open eyes
    - Inform the participant of the audio stimuli
        - Play one of each type of stimuli (more if requested)
        - Ask the participant to adjust the audio
        - Instruct the participant to quietly count the high pitch tones and that the answer will be collected after each block
    - Explain the grasping task
        - Move the cube as many times as possible without breaking or dropping it
        - Touch the middle between every lift, to make sure you let go off the cube completely
        - Let them try to break the egg a few times to get a sense of its sensitivity
        - If the cube is broken, put it down, touch the middle and try again
        - Emphasise that this is the main task
    - Trials
        - Trial 1: No grasping task, only oddball task
        - Trial 2: Grasping task, easy + oddball task
        - Trial 3: Grasping task, difficult + oddball task
    - Inform the participant that they will be asked to fill out a self-assessment questionnaire after each trial
        - Show the questionnaire and explain the questions and how it is to be filled out
    - Check that the stimuli is both heard by the patient and seen in the EEG-data

    - Put a post-it on the door that there is an EEG measurement

Without task

- Ground yourself and the participant again
- Place the laptop with the image of the cross in front of the participant
- Ask the patient to focus on the cross and count the high pitch tones
- Start the EEG recording. Name the file with patient name (avoid ÅÄÖ) and “no task”
- Ask the participant to begin the first trial and start the audio script. Name the file with date, patient name and “no task”
- At the end of each block, ask how many high pitch tones they counted and tell them that there will be one minute break. Stop the EEG recording.
- When 1 minutes has passed, ask the participant if he or she is ready to start the next block
- Start the EEG and continue the audio script
- When all blocks are finished, stop the EEG recording
- Ask the participant to fill out the questionnaire
- Control that the impedances are still ok

With task

- *The same procedure is used for both conditions (easy and hard)*
- Ground yourself and the participant again
- Start the camera
- Remind the participant of the task and let him or her try the sensitivity of the cube
- Start EEG recording. Name the file with patient name and “easy task”/”hard task”
- Ask the participant to begin and start the audio script. Name the file with date, patient name and “easy task”/”hard task”
- At the end of each block, ask how many high pitch tones they counted and tell them that there will be one minute break. Stop the EEG recording
- When 1 minutes has passed, ask the participant if he or she is ready to start the next block
- Start the EEG and continue the audio script
- When all blocks are finished, stop the EEG recording and the camera
- Ask the participant to fill out the questionnaire
- Control that the impedance is still ok

Resting state

- Make sure the participant is sitting comfortably and that their phone is on soundless
- Ground yourself and the participant again
- First part, closed eyes
    - 7 minutes
    - Ask the patient to sit with closed eyes and move as little as possible
    - Set the time and name the file
    - Turn out the light
    - Start the measurement
    - Leave the room and put a post-it on the door that there is an EEG measurement
- Second part, open eyes
    - 7 minutes
    - Place the laptop with the image of the cross in front of the participant
    - Ask the participant to sit still and focus on the cross
    - Start the measurement
    - Leave the room and put a post-it on the door that there is an EEG measurement

After the final trial

- Take a screenshot of the impedance page
- Move everything to the correct folder
- Remove the cap and the electrodes and clean them
- Bring the towel and shampoo and show the participant to the shower

- *Take at least one picture during a fake trial asking the participant to pretend to do the task. It is important not to do this during the trial in order to avoid making noise that might generate ERP similar to the ones generated by the auditory probes.*

Total time:

Prepare participant for EEG:	75 min
Explaining the task:	10 min
3 trials á 3x5 min:	45 min
Breaks+questionnaire:	20 min
Resting state:	15 min
Buffert:	15 min

→ 3 h with participant

Setting up:	*30 min*
Cleaning:	*30 min*

→ 4 hours total time

Appendix B: Earlier versions of the force sensitive cube

Version 1

The first version of the cube can be seen in Figure 24 and consisted of the following parts

- Red LED lamp
- Resistor of 230 Ω
- Force sensor Interlink FSR 406
- Resistor of 10 kΩ
- 3D-printed cube 55x55x55 mm
- Arduino UNO
- 9V battery
- Coupling board Luxorpart 45x34 mm
- Cables

The weight of this cube was 181 g including some filling of 45 g that kept the equipment inside the cube from moving. The force limit for breaking the cube was set to 8.1 N.

It was clear that an Arduino UNO is too big for this application and therefore we looked for a smaller one for later versions. The cube was also too small to fit the 9V battery inside. A simple red LED lamp was used for the visual feedback to the subject when the cube is broken.

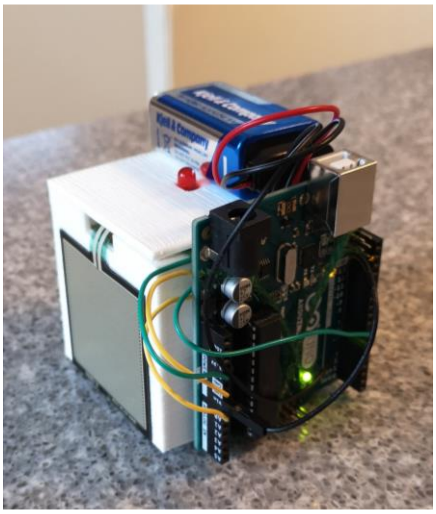


Figure 24. Version 1 of the force sensitive cube, including a red LED, one force sensor, an Arduino UNO and were driven by a 9V battery.

Version 2

The second version of the cube can be seen in Figure 25 and consisted of the following parts

- LED bar of 8 LEDs: NeoPixel 8 LEDs WS2812
- Resistor of 560 Ω
- 2x Force sensor Interlink FSR 406
- 2x Resistor of 10 kΩ
- 3D-printed cube 55x55x55 mm
- Arduino Micro
- 9V battery
- Coupling board Luxorpart 45x34 mm
- Cables

The weight of this cube was 164 g including some filling of 45 g. The force limit for breaking the cube was set to 6.6 N.

In this version of the cube the Arduino UNO from the previous version was replaced with an Arduino Micro, which could easily be fitted inside the cube. No new cube was 3D-printed for this version and therefore the battery still did not fit inside the cube. In the previous version of the cube it was a bit hard to recognize when the red LED was lit up and therefore it was replaced with a LED bar of 8 LEDs. This also gives the possibility to show a grading of how hard the cube is pressed for the subject all the time if wanted. The second version of the cube also had two force sensors instead of one to make it possible to measure both sides of the subjects' grip.

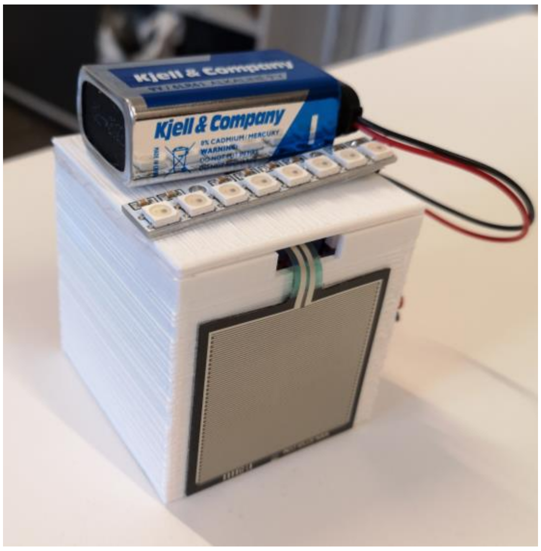


Figure 25. Version 2 of the force sensitive cube, including a LED bar with 8 LEDs, two force sensors, an Arduino Micro and were driven by a 9V battery.

Version 3

Picture of the third version of the cube is missing, but the parts that it consisted of is listed below

- LED-list of 8 LEDs NeoPixel 8 LEDs WS2812
- Resistor of 560 Ω
- 2x Force sensor Interlink FSR 406
- 2x Resistor of 10 kΩ
- 3D-printed cube 59x59x55 mm
- Arduino Micro
- 9V battery
- Coupling board Luxorpart 45x34 mm
- Cables
- Sandpaper, grit size 120
- Bluetooth transmission module Velleman HC05

The weight of this cube could be changed by adding weights inside it. The weight of the light cube, which corresponded to an easier level of difficulty, was 126 g. For the heavy cube the weight was 232 g. The force limit for breaking the cube was set to 4.1 N.

To fit all equipment inside the cube, a bigger version was 3D-printed. Sandpaper was put over the force sensors to make the friction higher such that the cube was easier to grip and lift. Also, a Bluetooth transmission module was included to make it possible to send data from the cube to a computer. This was used to display how many times the cube was broken.

Calibration of the cube

When the values from the force sensors were read by the Arduino, values between 0 and 1023 was given. To be able to convert these values into more understandable values in unit Newton, a rough calibration was done by adding weights in the range zero to 1147 g to the force sensors. The resulting values from the Arduino was plotted against the weight, which can be seen in Figure 26. The graph was seen to be approximately logarithmic and therefore a logarithmic curve fit was done, which also can be seen in Figure 26. This resulted in the curve fit with equation

$$y = -47.4 + 145.5 \cdot ln(x)$$

were y being the value from the Arduino and x being the weight. By calculating the weight x corresponding to the force limit 925 (Arduino units) and multiplying by the gravitation 9.82 m/s^2 we got the force limit 7.8 N.

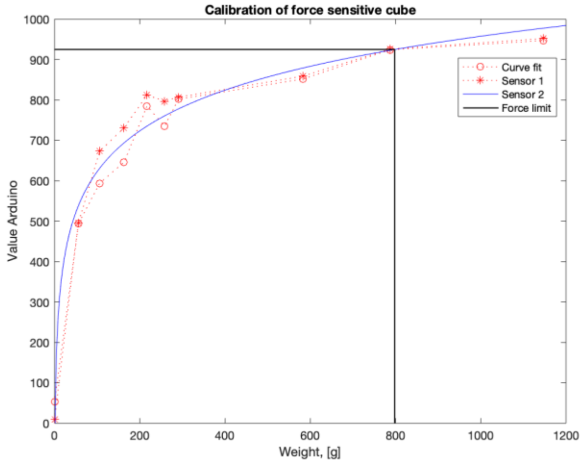


Figure 26. Calibration of the force sensitive cube. A logarithmic curve fit was done to be able to calculate the force limit 925 (Arduino units) in Newton. The limit was calculated to 7.8 N.

Appendix C: Informed consent

Informed Consent

Title: Comparison of cognitive workload between individuals and between different difficulties of the same task

You will hereby receive a request to participate in the research study, "Comparison of cognitive workload between individuals and between different difficulties of the same task", led by Associate Professor Max Ortiz Catalan, at Biomechatronic and Neurorehabilitation Laboratory (BNL) at Chalmers University of Technology.

Purpose of the study
The purpose of the study is to test a method to investigate the cognitive workload of the brain at the execution of a simple task where an object is to be lifted and moved a short distance. The object must not be pressed with too much force and the difficulty of the task will be varied by changing the weight of the object.

Participant Selection
You are invited to participate in this research study because you are healthy without motorical difficulties with your dominant hand, you have normal or corrected to normal vision and functioning hearing.

Voluntary participation
Your participation in this study is entirely voluntary. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you do not want to answer. There is no penalty if you withdraw from the study (and you will not lose

any benefits to which you are otherwise entitled).You should read the information below and ask questions about anything you do not understand, before deciding whether or not to participate.

Procedures
If you decide to volunteer for this study, we will ask you to participate in a research session at BNL lab at Chalmers University of Technology in Gothenburg. The session will last about 3 hours including pauses between the activities.

First, you will be asked to fill out a form, that inquire the extent of you having a dominant hand or foot, respectively. You will fill out this yourself, but the scientist will be available for answering questions if anything is unclear. After that you will be asked to participate in an experimental session.

You will sit down in a chair and be prepared with EEG equipment where electrodes will be placed on your head using a cap. After that a connection will be made between each electrode and your scalp by using a water-soluble gel. You will also get electrodes attached to your earlobes and next to your eyes. The exact position of all the electrodes will be scanned and we will ask you to blink, close your eyes and chew to control the equipment.

The trial will consist of three trials with each three blocks. Each block is about four minutes. Between each block you will get at least a minute rest and between each trial you get a five-minute break. In the first trial you will be asked to count a certain kind of sound as you focus your eyes on a picture of a plus sign on a screen. In the following two trials we will ask you to lift a force sensitive cube back and forth over a low barrier *at the same time* that you are counting the sound signals. If you press the cube to hard a diode will light up to indicate that the cube has been "broken". Your goal is to lift the cube as many times as possible without "breaking" it or dropping it. In one of the trials the cube will be light weight and that corresponds to the easy level, and in the next trial it will be heavier to increase the difficulty. After each trial you will be given five minutes to rest when you shall also say how many signals you heard and fill out a self-estimation of your cognitive effort during the trial.

Potential risks and discomfort
This study will not pose any risks to you as a study participant. Furthermore, we expect that if any discomfort or discomfort would arise after all, they will be of an extremely marginal nature. If, after all, the discomfort is perceived as a problem, you are free to finish the activity or your participation in the study at any time. Only standardized, electrically insulated, bioelectric signal registration apparatus will be used for EEG registration. We will include breaks between activities to minimize the risk of mental and physical fatigue.

You will get water soluble gel in your hair, that should be easy to wash away with water, but you will need to shower after the session.

Potential benefits for study participants/or society in general
Participation in the study will not give you any direct benefits to you as an individual. However, the results from this study will be used in a study that will investigate how sensory feedback affects the cognitive workload in similar trials as in this study. That way this study is a step towards further investigate how the sense of touch affects the performance of simple tasks. That result could therefore be a part of pushing the development of better feedback for hand prostheses.

Compensation for participation
Through your participation in the study, you will receive 1 movie theatre ticket as a reward.

Privacy
This section describes how BNL uses the personal data and the collected experimental data.

BNL at Chalmers University of Technology processes information in accordance with the General Data Protection Regulation (GDPR). Written consent must be obtained in order to collect and process personal data. Information in this form will be processed safely and will only be used for research purposes (see purpose of the study).

We collect the following personal information: Name, date of birth, photos, EEG data and the results of the form and self-estimations mentioned above.

Researchers Max Ortiz Catalan and Eva Lendaro will use the information collected in this way we consider best suited for publication or education. Information used for public email will not identify you as an individual.

How we store your data
BNL stores your data securely on the self-owned "network attached storage" (NAS) hard drive located at Chalmers University of Technology. Only laboratory members have access to this storage unit. Your data will be stored under a pseudonym. Only one person at BNL has access to the list where the pseudonym connects you to your name. This list is saved offline, separate from the collected data. We will save your personal data for a maximum of 10 years.

What are your data security rights?
You have the right to request copies of your personal data as well as the right to request that your personal data be transferred to another organization. Furthermore, you can request bnl to modify information that you believe is incorrect or incomplete. You also have the right to oppose and limit the processing of your personal data. Finally, you have the right to request BNL to delete your personal data.

If you would like to exercise any of these rights, please contact maxo@chalmers.se

You also have the right to contact the Swedish Data Protection Authority for complaints about how BNL handles your personal data.

Share results
Nothing you tell us today will be shared with someone outside bnl's research group, and nothing will be linked to you with your name. Some research results may be shared with another research group, but it will be pseudonymized data, which means you cannot be identified from the data.

Right to refuse or terminate participation
You are not obliged to participate in this research study. Whether you choose to participate in the study or not, will not affect you in any other respect. You are free to terminate your participation in the study at any time.

Identification of investigators
If you have any questions or concerns about the research, please contact Forskningskontakt
Max Ortiz Catalan
+46708461065
maxo@chalmers.se

Certificate of consent

Project:____

Date:____

Certificate of study participants:
I have read the above information, or have received it read out to me. I have had the opportunity to ask questions about it and the questions I have asked and have been answered to my satisfaction. I give my consent to voluntarily participate in this study.

Name of Participant____

Signature of Participant ____

Statement by the researcher/person taking consent
I have accurately read out the information sheet to the potential participant, and to the best of my ability made sure that the participant understands the study purpose and procedure.

I confirm that the participant was given an opportunity to ask questions about the study, and all the questions asked by the participant have been answered correctly and to the best of my ability. I confirm that the individual has not been coerced into giving consent, and the consent has been given freely and voluntarily.

Name of Researcher/person taking the consent____________________________

Signature of Researcher /person taking the consent____________________________

Annex to consent to participate in research

Notification/Consent for the collection and use of study data

This research will collect data about you that can identify you, refer to as Personal Data. The General Data Protection Regulation (GDPR) requires researchers to provide this Notification to you when we collect and use Personal Data from people in a State belonging to the European Union or the European Economic Area (EEA).

Study title
Comparison of cognitive workload between individuals and between different difficulties of the same task

Purpose of the study
The purpose of the study is to test a method to investigate the cognitive workload and compare this between different individuals and different levels of difficulty.

Personal data
The research team will collect and use the following type of personal data for this research:

- Contact information
- Name
- Date of birth
- Information about dominant hand and rot
- Photographs

Photographs / video files / audio files
Photographs, video files and audio files are also counted as personal data. This data will be protected and processed according to the same GDPR requirements used for the rest of the collected study data.

Potential risks
We do not foresee any risks regarding the safety of your personal information. All your personal data will be pseudonymized and will be stored in a protected file to which only the lead researcher has access.

Potential benefits for study participants/or society in general
The personal data is collected to understand the state and foundations of this research and to perform the purpose described above in the consent document.

Privacy
Any information obtained in connection with this study and that may be identified with you will be kept confidential and will only be displayed with your permission or as required by law. The confidentiality of your data will be ensured with a password-protected computer stored in Biomechatronics and Neurorehabilitation Laboratory (BNL) at Chalmers University of Technology. In addition, the file will be password protected and only the lead researcher will have access to it. To minimize the identification of your personal data, you will be named with a pseudonym which only the lead researcher can identify using the password-protected file. Data and consent documents will be stored for five years from the completion of the data collection and then shredded or completely deleted.

This research will keep your personal data for 10 years from the end of the research.

Access to Personal Data
The following categories of individuals may receive personal data collected or created about you:

- Members of the research group at Biomechatronics and Neurorehabilitation laboratory (BNL) at Chalmers University of Technology so that they can properly carry out the research
- Data controller, Dr. Max Ortiz Catalan, who is reviewing the study and analyzing the data.

The research team strives to protect the confidentiality of your personal data. Additional information about our protection of your personal data l is included in the consent document.

Right to your Personal Data
GDPR gives you rights regarding your personal data including the right to:

- Access, correct, or delete your Personal Data. On the other hand, the research team may need to retain personal data if necessary to carry out the purpose of the research
- Limiting the type of activities the research group can do with your Personal Data.

- Object to using your Personal Data for specific types of activities.
- Withdraw your consent to use your Personal Data for the purpose, described in the consent document and in this document (Please understand that you can withdraw your consent to use new Personal Data but Personal Data already collected will continue to be used as described in the Informed consent document and in this Notification).

If you want to know how your personal data is used, or if you feel that we have used your personal data in a way that violates the agreement or applicable legislation, please contact the lead researcher. If you have complaints about how Biomechatronics and Neurorehabiliation Laboratory at Chalmers University of Technology processes your personal data, you have the right to contact the Swedish Data Protection Authority.

Identification of investigators
Biomechatronics and Neurorehabilitation Laboratory (BNL) at Chalmers University of Technology, commissioned by Dr. Max Ortiz Catalan, is responsible for the use of your Personal Data for this research

You can contact Dr. Max Ortiz Catalan by phone +467 08461065 or by email maxo@chalmers.se If you have any:

- Questions about this Notification
- Complaints about the use of your Personal Data
- If you have a request regarding the rights listed above.

Signature of research participant or legal representative

I understand the procedures described above. I also give consent for the use of my Personal Data for the purposes outlined in this notice; for my Personal Data to be transferred overseas pursuant to the terms, conditions and limits specified at Section 43 of Legislative Decree n. 196/2003 as well as under the provisions of article 49 of the EU GDPR.

My questions have been answered to my satisfaction, and I agree to participate in this study. I am over the age of 18 years and have been given a copy of this form.

Name of Participant:____

Signature of Participant or Legal Representative:____________________________

Signature Investigator

In my judgement the participant is voluntarily and knowingly giving informed consent and possesses the legal capacity to give informed consent to participate in this research study.

Name of investigator:____

Signature of investigator:____

Appendix D: Artifact management details

Table 5. The components that were removed from each subject and condition following ICA analysis. The minimum is one component and the maximum 22. The mean number of removed components is 4.7.

SUBJECT	NO TASK	EASY TASK	HARD TASK
1	1	1,2,3,16	1, 2
2	1	1, 2, 3	1, 2, 12, 53
3	1, 2, 3, 4, 7	1, 2, 51	1, 2, 3, 11, 16
4	1, 4	1, 3	1, 2
5	1, 6, 10	1, 2	1
6	1	1, 2	1, 3
7	1, 3, 5, 9	1, 2	1, 4, 6, 7, 8, 9, 11, 12, 15, 22, 39
8	1	1, 3	1, 2
9	1, 4, 6, 22	1, 2, 3, 9, 10, 11, 12, 14, 15, 16	1, 2, 3, 5, 8, 9, 11, 13, 16, 17, 19, 22, 23, 29, 30, 33, 34, 36, 38, 39, 91
10	1, 5	2, 4, 6, 7, 8, 12, 13, 15, 16, 18, 20, 22, 23, 24, 25, 26, 29, 40, 41, 44, 53, 55, 120	1, 4, 5, 6, 8, 9, 11, 12, 16, 17, 19, 23, 29, 45, 46

Table 6. The number of epochs that were accepted after artifact rejection. In parenthesis is the percentage of how many of the total number of epochs that were accepted. Note that subject 9 was removed from further analysis because less than 50 % of the epochs were accepted. The number of accepted epochs remaining for analysis varied between 37 and 72 (mean 60.7) while the percentage of accepted trials ranged between 59.7 and 100 % (mean: 92.5 %)

SUBJECT	NO TASK	EASY TASK	HARD TASK
1	61 (98.4 %)	60 (95.2 %)	67 (97.1 %)
2	58 (96.7 %)	57 (87.7 %)	63 (92.6 %)
3	72 (100 %)	64 (94.1 %)	62 (93.9 %)
4	65 (97 %)	57 (83.8 %)	61 (96.8 %)
5	60 (98.4 %)	61 (95.3 %)	64 (98.5 %)
6	64 (98.5 %)	56 (87.5 %)	59 (86.8 %)
7	54 (81.8 %)	37 (59.7 %)	60 (93.8 %)
8	64 (97 %)	61 (93.8 %)	68 (95.8 %)
9	45 (67.2 %)	9 (14.8 %)	0 (0 %)
10	64 (98.6 %)	60 (87 %)	61 (91%)

Appendix E: Two-way ANOVA analysis

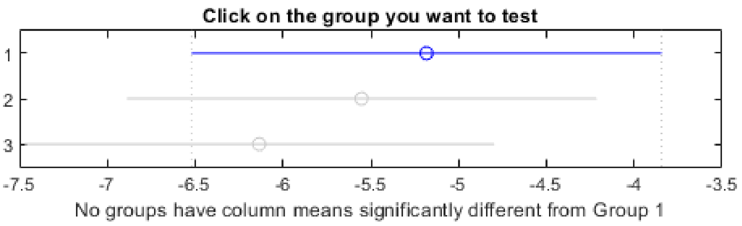
ANOVA table

Pairwise comparisons

N1 (S)

ANOVA Table

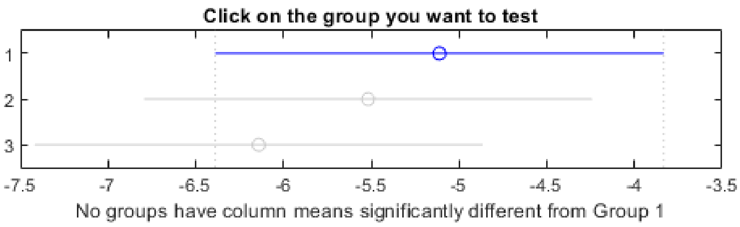
Source	SS	df	MS	F	Prob>F
Columns	4.151	2	2.0754	0.43	0.6581
Rows	169.494	8	21.1867	4.38	0.0058
Error	77.31	16	4.8319		
Total	250.955	26			



N1 (L)

ANOVA Table

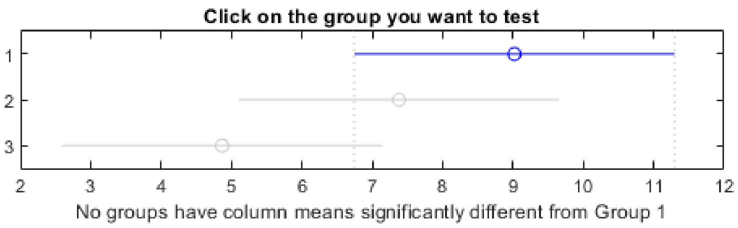
Source	SS	df	MS	F	Prob>F
Columns	4.846	2	2.4228	0.55	0.5877
Rows	144.117	8	18.0147	4.09	0.008
Error	70.536	16	4.4085		
Total	219.498	26			



P2 (S)

ANOVA Table

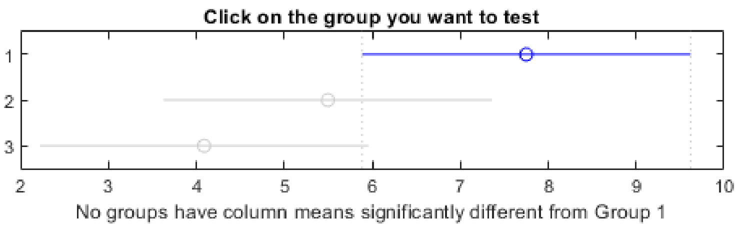
Source	SS	df	MS	F	Prob>F
Columns	78.77	2	39.3849	2.82	0.0895
Rows	243.97	8	30.4962	2.18	0.0877
Error	223.653	16	13.9783		
Total	546.393	26			



P2 (L)

ANOVA Table

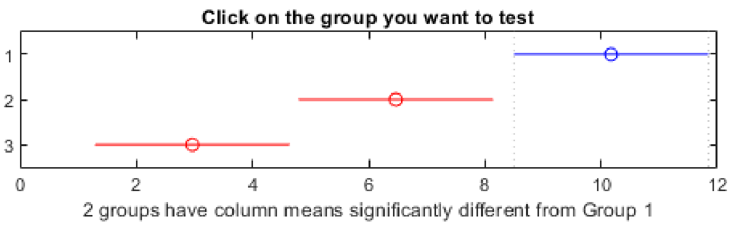
Source	SS	df	MS	F	Prob>F
Columns	61.462	2	30.731	3.26	0.065
Rows	230.524	8	28.8155	3.06	0.0273
Error	150.872	16	9.4295		
Total	442.858	26			



P3 (S)

ANOVA Table

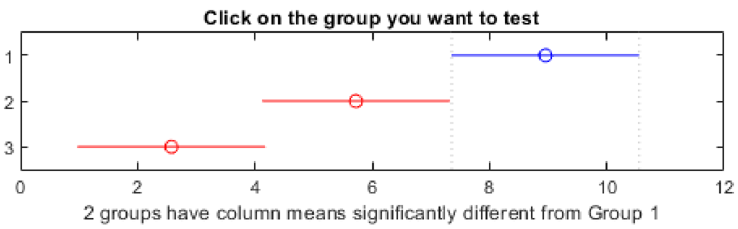
Source	SS	df	MS	F	Prob>F
Columns	233.898	2	116.949	15.44	0.0002
Rows	277.02	8	34.628	4.57	0.0047
Error	121.195	16	7.575		
Total	632.114	26			



P3 (L)

ANOVA Table

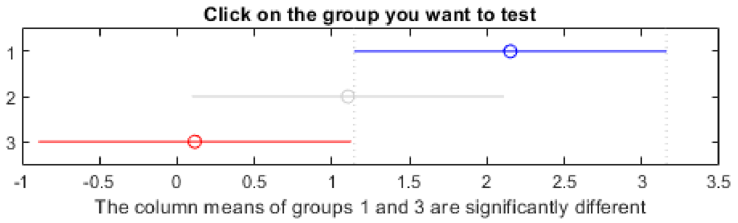
Source	SS	df	MS	F	Prob>F
Columns	183.117	2	91.5586	13.22	0.0004
Rows	193.092	8	24.1365	3.48	0.0161
Error	110.83	16	6.9269		
Total	487.039	26			



LPP
(S)

ANOVA Table

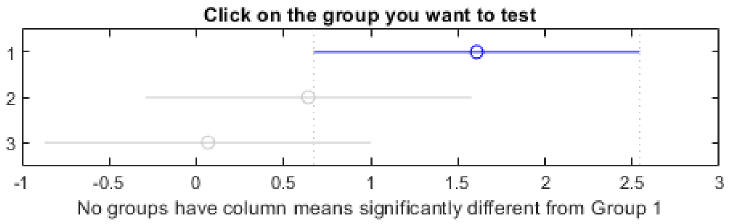
Source	SS	df	MS	F	Prob>F
Columns	18.674	2	9.33675	3.4	0.0589
Rows	76.004	8	9.50049	3.46	0.0166
Error	43.962	16	2.74763		
Total	138.639	26			



LPP
(S)

ANOVA Table

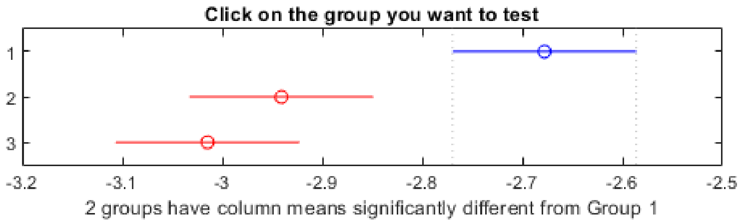
Source	SS	df	MS	F	Prob>F
Columns	10.919	2	5.4594	2.31	0.1315
Rows	80.39	8	10.0487	4.25	0.0067
Error	37.827	16	2.3642		
Total	129.135	26			



Low-
alpha
(S)

ANOVA Table

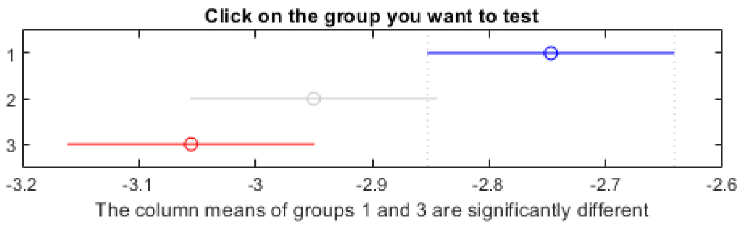
Source	SS	df	MS	F	Prob>F
Columns	0.56409	2	0.28205	12.37	0.0006
Rows	2.58281	8	0.32285	14.16	0
Error	0.36475	16	0.0228		
Total	3.51165	26			



Low-
alpha
(L)

ANOVA Table

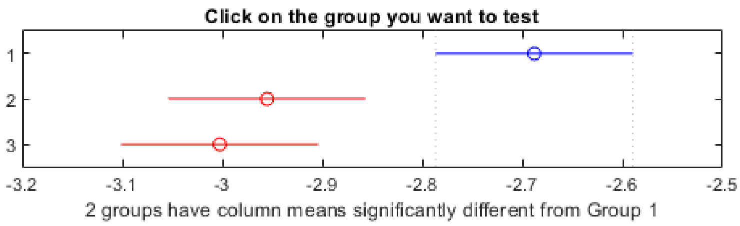
Source	SS	df	MS	F	Prob>F
Columns	0.44251	2	0.22126	7.3	0.0056
Rows	3.09657	8	0.38707	12.77	0
Error	0.48501	16	0.03031		
Total	4.02409	26			



High-
alpha
(S)

ANOVA Table

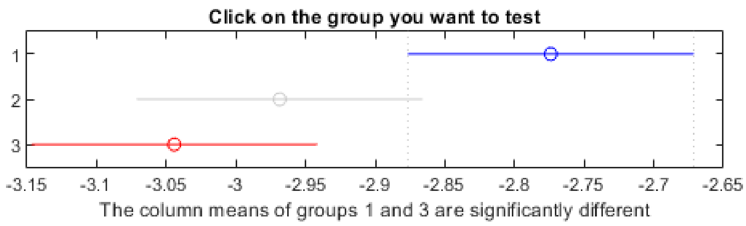
Source	SS	df	MS	F	Prob>F
Columns	0.5174	2	0.2587	9.86	0.0016
Rows	2.3713	8	0.29641	11.3	0
Error	0.41981	16	0.02624		
Total	3.30851	26			



High-
alpha
(L)

ANOVA Table

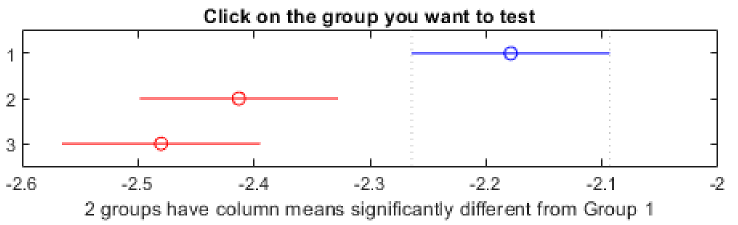
Source	SS	df	MS	F	Prob>F
Columns	0.35013	2	0.17507	6.15	0.0104
Rows	2.36655	8	0.29582	10.39	0
Error	0.45535	16	0.02846		
Total	3.17204	26			



Alpha
(S)

ANOVA Table

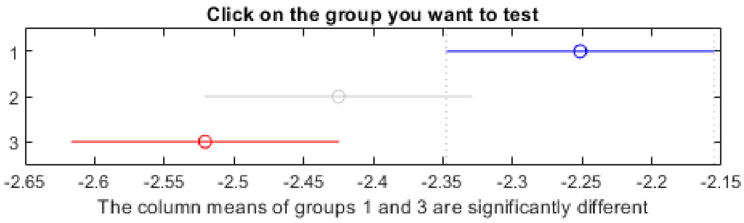
Source	SS	df	MS	F	Prob>F
Columns	0.45085	2	0.22543	11.45	0.0008
Rows	2.03481	8	0.25435	12.91	0
Error	0.31511	16	0.01969		
Total	2.80078	26			



Alpha (L)

ANOVA Table

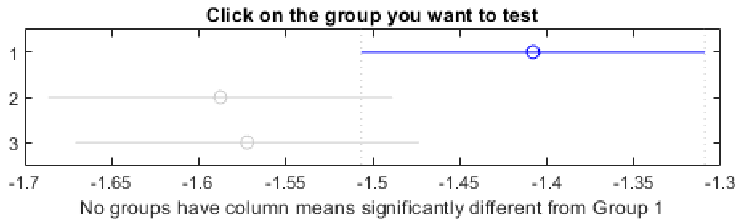
Source	SS	df	MS	F	Prob>F
Columns	0.33547	2	0.16774	6.72	0.0076
Rows	2.37423	8	0.29678	11.88	0
Error	0.39956	16	0.02497		
Total	3.10927	26			



Theta (S)

ANOVA Table

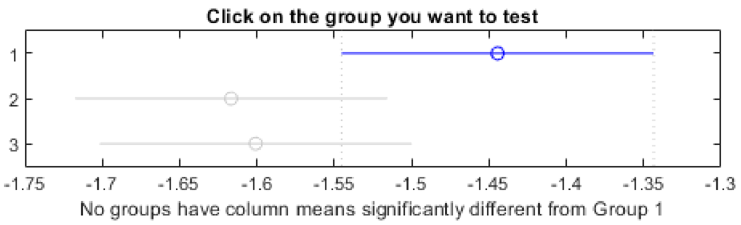
Source	SS	df	MS	F	Prob>F
Columns	0.1786	2	0.0893	3.38	0.0595
Rows	0.52616	8	0.06577	2.49	0.0571
Error	0.42217	16	0.02639		
Total	1.12693	26			



Theta (L)

ANOVA Table

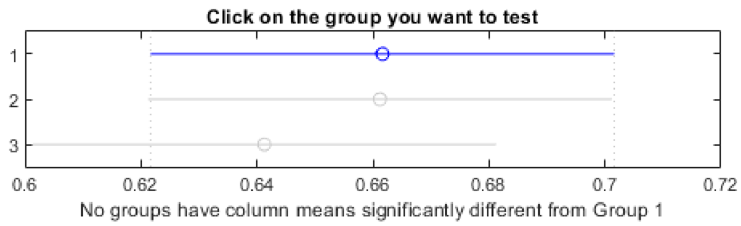
Source	SS	df	MS	F	Prob>F
Columns	0.16326	2	0.08163	2.96	0.0803
Rows	0.48923	8	0.06115	2.22	0.083
Error	0.44058	16	0.02754		
Total	1.09307	26			



Theta/ alpha (S)

ANOVA Table

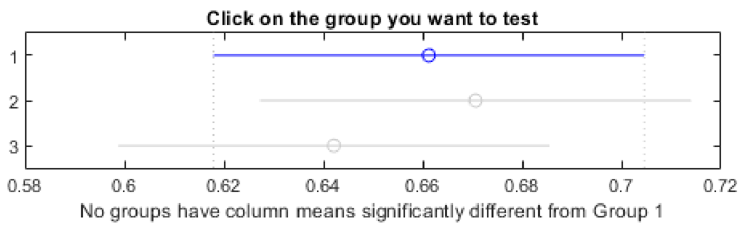
Source	SS	df	MS	F	Prob>F
Columns	0.00244	2	0.00122	0.28	0.7577
Rows	0.18528	8	0.02316	5.36	0.0021
Error	0.0691	16	0.00432		
Total	0.25682	26			



Theta/ alpha (L)

ANOVA Table

Source	SS	df	MS	F	Prob>F
Columns	0.0038	2	0.0019	0.37	0.6941
Rows	0.19238	8	0.02405	4.73	0.004
Error	0.0813	16	0.00508		
Total	0.27748	26			



Appendix F: Electrode clusters

Table 7. Description of the construction of each of the small (five electrodes) and large (seventeen electrodes) electrode clusters chosen for measuring each of the components N1, P2, P3 and LPP.

CLUSTER NAME	SCALP SECTION	ELECTRODES
N1 (S)	Central, right	119, C2, C4, 131, FC2
N1 (L)	Central	FC3, FC1, FCz, FC2, FC4, 115, 116, 117, 118, 119, 120, C3, C1, Cz, C2, C4
P2 (S)	Frontal/Parietal	FCz, 105, 106, 117, 118
P2 (L)	Frontal/Parietal	F1, Fz, F2, 104, 105, 106, 107, FC1, FCz, FC2, 116, 117, 118, 119, C1, Cz, C2
P3 (S)	Frontal/Parietal	Fz, 105, 106, FCz, 118
P3 (L)	Frontal/Partietal	172, 95, 96, 173, F1, Fz, F2, 104, 105, 106, 107, FC1, FCz, FC2, 117, 118, Cz
LPP (S)	Occipital	Pz, P1, 154, 152, 153
LPP (L)	Occipital	150, 151, 152, 153, 154, 155, 156, 157, PO3, POz, PO4, PO7, 162, 163, 164, 165, PO8

Table 8. Description of the construction of each of the small (five electrodes) and large (seventeen or sixteen electrodes) electrode clusters chosen for measuring the frequency bands Theta and Alpha (Broadband, Low-, and High-).

CLUSTER NAME	SCALP SECTION	ELECTRODES
THETA (S)	Frontal/Parietal	105,106, 117, 118, FCz
THETA (L)	Frontal/Parietal	104,105,106, 107, 116, 117, 118, 119, FCz, FC1, FC2, Fz, F1, F2, Cz, C1, C2
BROADBAND, LOW AND HIGH ALPHA, (S)	Occipital, right	154, 155, 156, P2, PO4
BROADBAND, LOW AND HIGH ALPHA (L)	Occipital	151, 152, 153, 154, 155, 156, 162, 163, 164, 165, Pz, P1, P2, POz, PO3, PO4

CHALMERS
UNIVERSITY OF TECHNOLOGY