CHALMERS
UNIVERSITY OF TECHNOLOGY

Pupillary Measures as Indicators of Cognitively Versus Automatically Controlled Processes

Master's Thesis in Biomedical Engineering and Systems, Control and Mechatronics

Sofia Granberg
Malin Wallhede

Department of Mechanics and Maritime Sciences
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2018

Master's Thesis 2018:33

Pupillary Measures as Indicators of Cognitively Versus Automatically Controlled Processes

SOFIA GRANBERG
MALIN WALLHEDE

Department of Mechanics and Maritime Sciences
Division of Vehicle Safety
Chalmers University of Technology
Gothenburg, Sweden 2018

Pupillary Measures as Indicators of Cognitively Versus Automatically Controlled Processes
SOFIA GRANBERG
MALIN WALLHEDE

Supervisor: Emma Nilsson, Volvo Cars (SAFER)
Supervisor: Per Lindén, Volvo Cars
Supervisor: Fredrik Granum, Volvo Cars
Supervisor: Pinar Boyraz Baykas, Division of Vehicle Safety, Chalmers
Examiner: Robert Thomson, Division of Vehicle Safety, Chalmers

Master's Thesis 2018:33
Department of Mechanics and Maritime Sciences
Division of Vehicle Safety
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Close up image of a gazing eye.

Gothenburg, Sweden 2018

Pupillary Measures as Indicators of Cognitively Versus
Automatically Controlled Processes
SOFIA GRANBERG, MALIN WALLHEDE
Department of Mechanics and Maritime Science
Chalmers University of Technology

Abstract

The Cognitive Control Hypothesis is a hypothesis that states how car drivers are affected by cognitive load. It says that *"Cognitive load selectively impairs driving subtasks that rely on cognitive control but leaves automatic performance unaffected"*. To test if this hypothesis is true, a method for measuring both cognitively demanding and automatized tasks is needed, but also a method for distinguishing between these two conditions. This thesis was conducted at the Vehicle Safety Division at Chalmers Univeristy of Technology together with Volvo Cars, with the aim to investigate if pupillary measures can act as indicators of cognitive load. The aim was also to investigate what eye measures are the most reliable and/or preferable for assessing cognitive load. The research questions were what eye/pupillary measures that can be used in order to measure cognitive activity and if these pupillary measures can be indicators of whether a task is cognitively demanding or automatically performed. The aim was also to develop an experimental method where the data of pupillary measures was to be collected with an eye tracking camera. The project did not investigate other physiological measures and no naturalistic driving was performed.

The main experiment included a method where participants performed a motor piano task consisting of different key pressing combinations, performed at two sessions. The task also had a sequence where math questions were induced in order to vary the mental workload. The motor task was expected to (to some extent) be automatized at the second session. The most important finding was a 10Hz peak in the frequency domain, mostly visible during the sequences where the participants were asked math questions. The results also showed that the energy content within 3.75-15Hz was higher for sequences with math questions and thus indicating a higher mental workload. It was also observed that the pupil behavior varied a lot, both within and among the individuals. No significant pattern was observed regarding pupil dilations within and among individuals when they were exposed to math questions.

Keywords: Cognitive control, Automatized tasks, Eye tracking, Pupillary measures, Motor piano task, Cognitive Control Hypothesis, Cognitive Load Theory

Acknowledgements

This report was performed during the first six months of 2018 by Sofia Granberg and Malin Wallhede, students from the Master Programs Biomedical Engineering and Systems, Control and Mechatronics. The study was conducted at the Vehicle Safety institution at Chalmers University of Technology, where the examiner was Robert Thomson.

First, we would like to thank all our supervisors, Emma Nilsson (Chalmers/Volvo Cars), Per Lindén (Volvo Cars) and Fredrik Granum (Volvo Cars), for your guidance and encouragement. A special thanks to our Chalmers supervisor, Pinar Boyraz Baykas, for all her enthusiastic feedback and guidance.

We would like to thank Volvo Cars, for giving us the opportunity to perform our master thesis at the company. We would also like to thank our colleagues and the test persons who contributed to our study.

Lastly, we would like to thank friends and family for all their love and support during our time at Chalmers.

Sofia Granberg and Malin Wallhede, Gothenburg, June 2018

Contents

1| Introduction

The introductory chapter of this report presents the project background and its objective. The aim is then presented and the chapter ends up with the limitations of the project.

1.1 Background

As the level of automation for vehicles increases, the drivers' role will include more supervisory control, where measuring the mental workload might be an important parameter for road safety [1]. When developing driver support systems, it is important to have knowledge and comprehension about the driver's behaviour. It is also important to understand the limitations of the driver's state, for example level of concentration [1], distraction, sleepiness and stress [2][3]. A driver state that is currently debated is cognitive distraction, which happens when the driver is mentally engaged in something other than driving and thus affects the driving performance [4]. In order to learn more about how these factors affect car drivers, reliable measures of cognitive load in addition to understanding the interplay between there measures and the cognitive load are helpful. To understand and explain how drivers are affected by cognitive load, the Cognitive Control Hypothesis can be used [4], stated as:

"Cognitive load selectively impairs driving subtasks that rely on cognitive control but leaves automatic performance unaffected" [4].

It suggests that cognitive load should only affect performance in tasks that requires attention and control by the subject, and not tasks that are performed automatically. The Cognitive Control Hypothesis uses Guided Activation Theory (GAT) when defining the difference between automatized and controlled tasks from a neurological perspective. According to GAT, when a perception is mapped within the brain it can be viewed as a neural pathway. These pathways will gradually strengthen with a repeated exposure to the perception and exposing the neural pathway [4]. Automatically performed tasks have strong neural pathways. The term automatically performed refers within this project to processes or tasks that are performed by a human who does not have to pay attention or use his/her cognitive resources load to solve the task while performing it. Today there is no well known method to measure cognitive load, and therefore no possibility of determining whether a task requires cognitive control or is performed automatically [5]. This causes difficulties when using the Cognitive Control Hypothesis [6].

Physiological changes related to cognitive load have been identified in physiological parameters, for example in skin conductance, dilation of pupil and EEG signals [2]. It has also been observed in other eye activities and endocrine responses [1]. The changes in pupillary measures have proven to be sensitive to mental workload in earlier work [1], and will in this thesis be further analyzed to investigate whether they can act as indicators of cognitively controlled or automatically performed tasks [5]. Other eye measures, possible to collect with an eye tracking equipment, will also be gathered during the data collections. Examples of these are blinks and gazing.

1.2 Aim

The aim with this thesis work is to analyze if pupillary measures can act as indicators of cognitive load and what eye measures that are the most reliable and/or preferable. An initial hypothesis is that the pupillary diameter, which is collected with an eye tracking camera, could be an indicator for cognitive load. Auxiliary pupillary measurements are also investigated.

The aim is also to develop an experimental method, where data is collected with an eye tracking camera, and construct an algorithm which can possibly correlate the measures to the cognitive load.

This project has the objective of finding the answers to the following research questions:

1. What eye/pupillary measurements, possible to collect with an eye tracking equipment, can be used to measure cognitive activity?
2. Can these eye/pupillary measurements be indicators of whether a task is cognitively demanding or automatically performed?

1.3 Limitations

This study is only covering pupillary/eye measures as indicators of cognitive load and is not investigating other measurements, such as EEG or galvanic skin potential. From the data collections, only the subjective and physiological measures are analyzed. Other measures that were collected, for example blinks, gazing and performance measures, will not be able to be investigated during the project due to a limited amount of time. The project is mainly focusing on tests during experiments, carried out with an eye tracking camera, and not on real naturalistic driving.

2| Theory

This chapter presents information from a literature study, starting with a summary of the anatomy of the eye and the visual perception process. This is followed by how the human processes information from the senses and surroundings, including memory models and different information processes. The next part of the chapter defines cognitive load and states its correlation to human physiological parameters, followed by how to measure it. A summary of tasks that can be used to induce cognitive load is also presented, followed by information of using eye measurements as indicators of cognitive load. The chapter is then finalized with facts about human motoric learning.

2.1 Human Visual Perception and Physiology of the Eye

Human perception uses information from the senses, gathered from the world and its surroundings, in order to make sense of it [7]. When focusing on the visual sensory system, it collects information about properties and locations of objects and the world via the eyes, so the human can interact with the surroundings.

The eye, which can be seen in Fig. 2.1, is sensitive to and registers energy reflected or emitted by objects in the surrounding, traveling in the form of electromagnetic energy light waves [7]. The energy from the surrounding visual scenes enters the eyeball via a black ”hole”, the pupil, which is located in the middle of the colored part of the eye, the iris [8]. To regulate the amount of energy and light that passes through the pupil, the iris consists of two different types of smooth muscles, which together create movements that can regulate the diameter of the ”hole” [8]. These are circular and radial smooth muscles, which can be seen in Fig. 2.2 as the the white circles around the pupil and white radial lines respectively. These muscles are controlled by the autonomous nervous system (ANS), and therefore not voluntarily controlled. When the eye adjusts to bright light, see Fig. 2.2a, the pupil constricts, meaning that it decreases in size as the circular smooth muscles of the iris contracts. When the eye is stimulated by dim light, see Fig. 2.2c, the pupil dilates due to the contractions/relaxations of the radial smooth muscles. It has been found that the pupil also has a behaviour called ”pupillary unrest”, meaning that it will oscillate despite constant lightning condition. This happens with a frequency of about 0.5Hz and the dilation is around ±0.5mm [9].

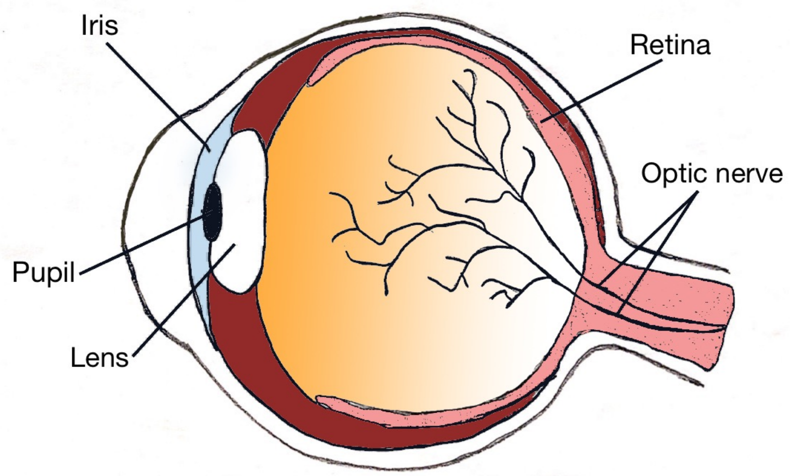
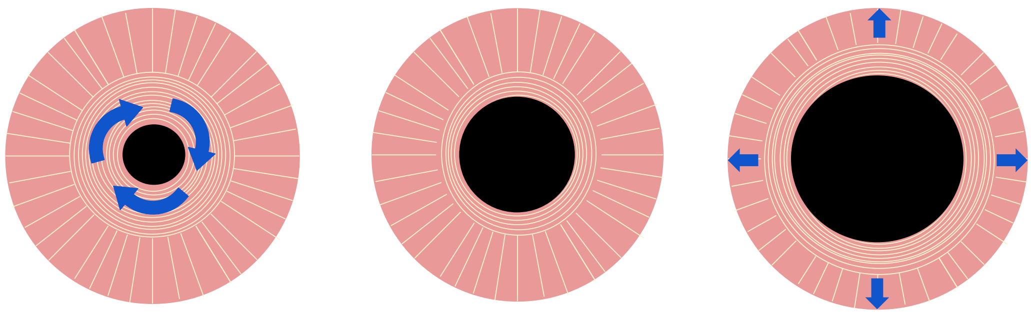


Figure 2.1: Image of the eye



(a) Constricted pupil **(b)** Normal pupil **(c)** Dilated pupil

Figure 2.2: The different behaviours of the pupil, and thus the musculature of the iris.

Pupil oscillations have been studied on nonhuman primates by conducting an experiment based on a visual task [10]. The pupil size was collected in order to detect pupil dilations and its frequency. The result demonstrated that the mental workload correlated with the frequency from the rapid pupil dilations. They identified that the frequency of pupil dilation peaked right after the visual task was presented, and dropped until the response was completed. The highest frequency was identified to be 9.1±0.8 Hz.

When the light passes trough the pupil it goes trough the lens, which is a transparent structure that focuses light beams onto the retina [8]. The retina is located in the back of the eye, see Fig. 2.1, covering two thirds of the eyeball and consists of a layer of light sensitive receptors. These receptors are called photoreceptors and convert the intensity of light into electrochemical signals. These signals are then transferred to the brain via the optic nerve, with a first stop at the thalamus. The information is ”sorted” at the thalamus and then sent further to the visual cortex in the brain, where it is processed and interpreted.

2.2 Information Processing of the Human

A large amount of information is collected by the human sensors, and processed by the human brain every day. Some information will be kept and stored to be used at later occasions, while other will not be registered or paid attention to, and thus not remembered. To help the human sort among the information that is presented, the brain uses different systems and processes. These are both used to interpret incoming information, but also to manipulate and produce outcoming data, for example conducting long-term memories or replying to a math question. This section will summarize different processes of the human brain including storing, information processing and manipulation.

2.2.1 The Memory

The memory of the human refers to processes that allow to record/encode, store and then later retrieve/access the stored experiences and information [11]. These characteristics are important factors for having the possibility of generating for example automatized processes. To describe how the memory operates and cooperates with the information processing of the human, the *Three-stage model* developed by Atkinson and Shiffrin can be applied [7]. A simplified image of the model can be seen in Fig. 2.3.

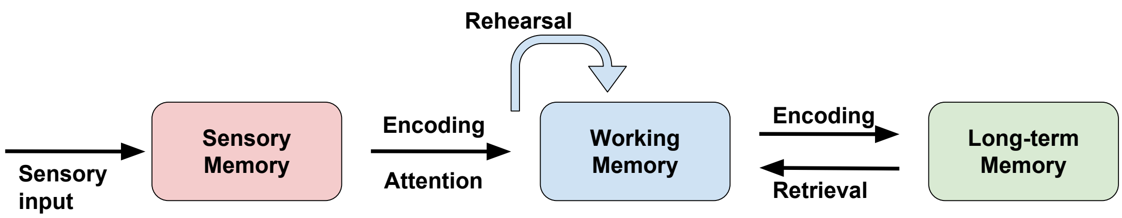


Figure 2.3: The three-stage model of the memory, consisting of sensory memory, working memory and long-term memory.

The three-stage model states that the memory has three major components: *the sensory memory*, *the working memory* and *the long-term memory.* The sensory memory can briefly store incoming sensory information, for example from the visual perception system [12]. Most information in the sensory memory fades away, but only information that is considered crucial or selectively paid attention to will be sent further/be encoded to the working memory. This information is retained until an opportunity to reuse it occurs [12]. The stored information often requires not only to be stored, but also to be cognitively operated by the mind, for example by manipulating or transferring it. This capacity, the short-term mental storage and cognitive operations, is defined as the *working memory* [7]. How the working memory operates can, according to *The Baddeley-Hitch Model*, be described with four components, as shown in Fig. 2.4.

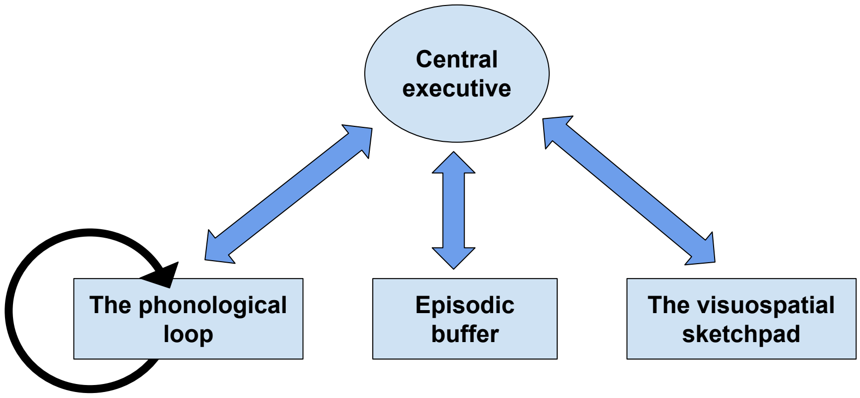


Figure 2.4: The four components of the working memory, according to the most recent model of Baddeley [7]. The phonological loop and the visuospatial sketchpad stores representations of sounds and visual/spatial information respective. The episodic buffer is a space where information from the long-term memory and the working memory can be integrated, manipulated and made available for the conscious awareness of the human. These three storages are controlled by the central executive, which plans and sorts the sequence of actions that need to be performed.

There are two short-term storages intended for storing different types of information. One of them works as a memory buffer for verbal information, the *phonological loop*, and the other for visuospatial information, the *visuospatial sketchpad* [7]. The phonological loop refers to the capability of remembering for example numbers and words, by silently repeating the combination with the humans' own voice in the mind. The combination can then be ”rehearsed” by repeatedly hearing the voice. The visuospatial sketchpad is the capability of entering a mental image, such as a room or memory, and develop, inspect and navigate trough that image. This function is giving the human the capability of visualizing a place in space without having to physically look at or being in it. These two memory buffers interact with each other via a control system named the *central executive*, which is the third component in the model [7]. This system governs what information should be kept or removed from the short-term storage, enabling an effective workplace for cognitive activities. It is thus controlling and allocating the attention of the mind, by planning what operations to be executed and performing the ”work” of the working memory.

The fourth part of the working memory, the *episodic buffer*, is also a space for temporary storage. In the episodic buffer, information from the long-term memory and the working memory can be integrated, manipulated but also made available for the conscious awareness of the human. In summary, an example of how the parts of the working memory operates, mental arithmetic numbers can be used. When one hears or reads the question ”How much is 35 plus 14”, the working memory initially starts to work. The phonological loop maintains the auditory codes for the numbers 35 and 14, and mental images of the numbers might be presented by the visuospatial sketchpad. To be able to perform the addition, knowledge from the long-term memory has to be retrieved and stored in the episodic buffer, where it is integrated with the auditory and visual information. The central executive has not stored any information, but planned and controlled the sequence by which order the actions were performed in.

The last part of the three-stage model of the memory, seen in Fig. 2.3, is the long term memory [7]. This stage is responsible for the long-lasting retention of information, experiences, and skills. Theoretically, the long term memory capacity is unlimited and rather more limited by accessability of the memories. There are two general types of long term memories, the *declarative* and the *procedural* memory. The declarative memories contain factual knowledge and includes the knowledge one can declare and tell other people. The seconds part, the *procedural memory*, contains the non-declarative memories of a human, such as actions and skills. This part is responsible for knowing and remembering how to perform things in particular situations, like riding a bike or tying shoelaces.

2.2.2 Automatic Versus Controlled Processes

From a cognitive psychology viewpoint, the mental life and information processing of a human is built up by conscious and unconscious parts working in harmony [12]. There are many activities or tasks that require *controlled (explicit) processing*, which indicates conscious use of attention and effort to fulfill a task. In comparison, there are processing activities that can be performed without any conscious effort and awareness, called *automatic (implicit) processing.* The transition from explicit to implicit processes can be seen as a graded phenomenon [4]. The automatic and controlled processors can also be referred to as System 1 and System 2 operations [13]. System 1 handles thoughts and actions that are processed automatically, whereas System 2 operates when something requires attention. System 2 tasks often involves the working memory [13]. In this report, the terms *automatic processes* and *System 1 operations* can be used interchangeably, but the term automatic will primarily be used.

Automatic Processes - System 1

An automatic process is performed with little or zero attention or effort [14]. It is robust to stress and immune to cognitive load [4]. System 1 tasks are able to work in parallel, even in situations when high workload is present [14]. All types of tasks and behaviours have a various tendency to become automatic. Dedication, knowledge of the procedure of the task and memory load are crucial factors when tasks are processed by System 1. Many cognitive processes, such as arithmetic operations or conducting conversations, can be performed automatically if they are simple [15]. This could be, for example, recognizing familiar symbols from an alphabet, recognizing small number of objects or conducting a simple conversation. Since

System 1 tasks are performed with little or no attention, the processes are performed automatically, and the ”solutions” or approaches often lacks contemplation. That reduces the chance of finding new way of thoughts and procedures on habited patterns of behaviour. For example, one does not usually question or change a learned pattern, as the way one walks or tie shoelaces, until getting aware of issues with the behaviour.

Controlled Processes - System 2

A controlled process requires attention and effort from the subject [14]. A System 2 task is associated with concentration, making choices and decisions about what to think and/or do [13], [15]. A controlled task is limited by the capacity of the working memory in the brain, and such tasks are affected by stress [14]. The capacity limit depends on the available resources in the working memory, which varies between individuals but is never infinite.

Using System 1 or System 2

The complexity of a task decides if System 1 or System 2 should be used [13]. Calculating $2+2$ is a System 1 activity (in general for grown ups) whereas remembering a phone number requires attention, and is thus a System 2 activity. Driving on an empty road is a System 1 task for an experienced driver, whereas parking in a narrow space is a System 2 operation. System 1 thoughts are often involuntary, one cannot resist thinking of 4 when $2+2$ is displayed. Interaction between the systems occurs continuously throughout the day. In general, every feeling and action originates in System 1 but is taken care of by System 2 when things get too complex for System 1. System 1 intakes impressions, feelings and intuitions, which then are suggested/presented for System 2. If System 2 endorses/accepts these suggestions, the impulses from System 1 will turn into voluntary actions. System 2 is also activated to control feelings, meaning one’s self-control.

2.3 Cognitive Load and its Correlation to Human Physiological Parameters

Cognitive load, also known as mental (work)load, is in this project defined as:

The mental effort in the limited working memory capacity,
that is required to fulfill a task.

When applying cognitive load, it implies that the person enters System 2 - the cognitive processes [13]. The cognitive load is correlated to a subject’s perception of a certain task and psychological conditions, such as mental stress and fatigue [16]. The cognitive load is also affected by cognitive distraction, defined as the withdrawal of attention from the primary task. This type of distraction has been observed during tasks with both high and low cognitive load.
When trying to understand the cognitive load and its correlation to learning, the

Cognitive Load Theory (CLT) can be applied. The Cognitive Load Theory is a framework that is developed based on results from empirical studies regarding learning and instructions [17]. An assumption from this theory is that cognitive load can be divided into three parts, see Fig. 2.5.

- **Intrinsic:** The complexity of a learning task and amount of capacity from the working memory that is dedicated to it. A complex task corresponds to a higher intrinsic load.
- **Extraneous:** The level of extraneous load depends on the instructions and possible disturbances. Clear instructions and no disturbances indicates a low extraneous load and vice versa.
- **Germane:** Very similar to intrinsic load, but related to the long-term memory. It comprises the level of resources devoted to acquiring and reach automation/System 1 activities of schematas [18]. A schema is within psychology seen as a sensory, motoric or cognitive structure that forms the humans' way of perceiving sensory inputs, both conscious and unconscious [7]. Germane load indicates how much capacity from the working memory that is dedicated to the learning [19].

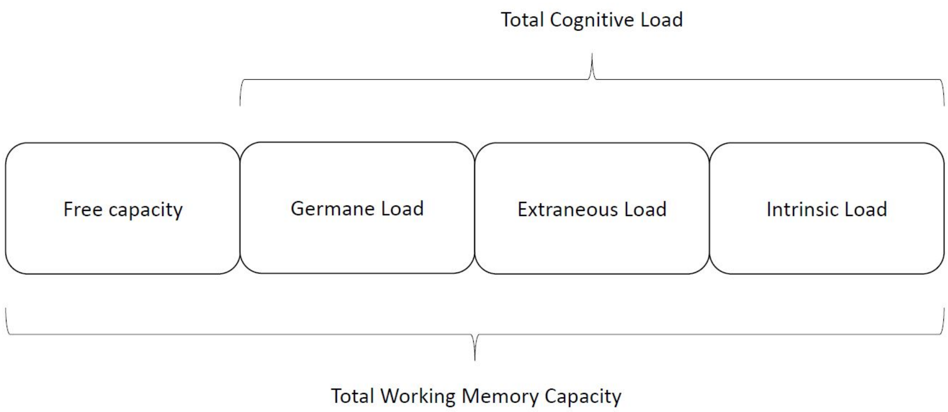


Figure 2.5: Model of cognitive load, describing the mental workload and its three components, which together sum up into the total use of cognitive load. The free capacity show the remaining available capacity, when subtraction of the Total cognitive load from the Total Working Memory Capacity.

Variations in the mental workload is, besides psychological conditions and distractions, also characterized by involuntary changes in physiological measures caused by the autonomous nervous system [2]. These are associated with the invested mental effort in dealing with a task. The physiological changes due to increased cognitive demand can for example be obtained in skin conductance, heart rate, dilation of pupil and EEG signals. It can also be obtained in other eye activity, muscular activity, head motion, body movement and posture, and endocrine response [1].

2.3.1 Measuring Cognitive Load

The involuntary changes in physiological parameters arising from an increase in mental workload are desirable to measure, since they might give momentary measures of the temporary mental workload. When measuring mental workload in experimental studies, there are three different categories often used in research: *performance*, *subjective*, and *physiological measures* [1], [2].

Performance measures are objective and measurable, and used to evaluate how well the test object is performing a task [1]. The results give a clear indication if the test object fulfills a task, i.e manages to remember all digits in a memory task. A common method to assign mental workload and measure performance is by using the dual-task method, where a primary task is performed and then a secondary task is introduced to evaluate the test objects remaining mental capacity, under the assumption that the tasks use the same kind of load to the working memory [1]. From a performance aspect, it is important to evaluate the possibility of adaption and learning of resource strategies, since all individuals have a varying learning ability.

The subjective measurements capture the test object's own assessment of mental workload across the overall-task, with administered questionnaires after the experiments [1]. These measurements focus on the individual experience, are easy to implement and widely used for many learning contexts [17]. Subjective measurements are however criticized because of the difficulty and complexity to distinguish a valid universal subjective rating. This approach is measured subsequently and not continuously during the task and the method therefore lacks information about cognitive load during the task. It does however give a good indicator about the intrinsic cognitive load that is based on element interaction. Examples of an assessment tool is the NASA Task Load Index (NASA-TLX). NASA-TLX captures the subjective mental workload of each participants individual experience, and includes six dimensions: mental, physical and temporal demand and performance, effort and frustration [17]. The participant should rate each dimension on a Likert scale, where the span goes from ”very low” to ”very high” [19]. However, NASA-TLX lacks the ability to separate between mentally and visually challenging tasks [20].

The physiological measures are recordings of the test objects physiological state and variations during an experiment. Measuring these parameters often require specialized equipment and technical expertise. Many of these physiological measures can be obtained noninvasive, which is an advantageous factor for making the test objects feel natural when performing the experiment [2]. Changes in these measurements arise from workload but also from other factors, such as physical fatigue and/or stress [1]. Environmental variables, such as illuminance and temperature, can also affect the physiological measures in an undesired way. All these factors are important to have in mind when designing an experimental study.

2.3.2 Tasks Used to Induce Cognitive Load

Inducing and varying a person's mental workload, to try to imitate both automatized and cognitively controlled behaviours, can be done in multiple ways. Some tasks that have been used in studies similar to this thesis project will be presented in this chapter.

Single- and Dual-task

The performance of one or several tasks cannot be greater than 100 %. A single-task performance occurs when all of the capacity is invested in one activity [11]. A dual-task method consists of two tasks that are performed concurrently while measuring performance on both of them [17]. These simultaneously performed tasks can be both cognitive or automatic [11]. The tasks will compete for resources in the working memory, and will thus be allocated between these. When performing several tasks, only one of the them is defined as the primary one and the other one is thus defined as the secondary task. Dual-tasks that rely on cognitive control will impair the performance on either one or both, unless they are data-limited. A data-limited task has a limit where perfect performance is possible to achieve, i.e remembering a three digit number, where one cannot perform greater than remembering all three digits. Secondary tasks are often used to assess mental workload. Adding load to the working memory to invest in a new task will lead to a decrement in the concurrent one, unless the concurrent data is limited or automated and there is capacity available in the working memory. It has been found that humans are able to time-share one automatic and one cognitive (resource demanding) task with perfect efficiency if they will have the chance to practice. The practice leads to the ability to allocate their resources away from the automatic task and only focus on the cognitive one. A dual-task method usually consist of a learning task as the primary one and the secondary is containing a cue [17]. A cue is a stimulus which induces a certain behavior, both conscious and unconsciously [7]. The performance of the secondary task is often measured as the time it takes for the participant to response to a cue, i.e response time, and the cue is often visual or auditory [17].

"The rhythm method" is a version of a dual-task method conducted by a research team [17]. The rhythm method differs from a regular dual-task methods in the sense that the secondary task does not involve any external cues. Because of this, no sensory interference between the two tasks is present. The rhythm method is validated from one study where the primary assignment was working with multimedia-learning program and the secondary task was tapping their foot. The rhythm of the secondary task was introduced and practiced before the experiment. The performance was analyzed by observing the deviation within each participant's individual tapping pattern. A sequence with little deviation, i.e good tapping precision, indicated that less cognitive load was used on the primary task.

Stroop Task

The Stroop Task is a verbal task where the participant is told to read coloured words out loud [21]. The task can be designed in multiple ways, a common one is to have two conditions: one neutral and one with conflict. In the neutral condition (congruent), the coloured word matches the word that is read out loud whereas in the conflict condition (in-congruent), words are written in an incongruous colour. This conflict causes delay and disruption.

A prior study has shown a correlation between pupil diameter and Stroop task such that the diameter increases during the in-congruent condition, in comparison to reading the same words as uncolored (black) [22]. The delay that the Stroop task induced showed that the responses came faster for the color congruent words and were delayed for the color in-congruent words.

2.3.3 Eye Measurements as Indicators of Cognitive Load

The use of eye measures have been found reliable regarding measuring cognitive load [2]. As the quality of cameras are increasing rapidly, and the fact that it is a non-invasive tool, the use of eye tracking cameras for assessing mental workload are growing in popularity.

Within the category of eye measurements, there are several different measures of eye activities that have been confirmed as sufficient indicators of cognitive workload [1]. These are blinks, fixations and pupillometry.

Blink Rate and Duration

Among blink-related parameters, there are several that can be be used to indicate mental workload. These are for example blink rate and blink duration.

Blink rate is a possible indicator of whether a task is visually or mentally challenging [20]. A demanding task, meaning high mental workload, with no visual input generally causes an increase in blink rate, whereas a demanding task with visual input decreases the rate.

When analyzing blink duration, it has been shown that duration decreases with increased mental workload and task demand, both for visual and mental tasks [1]. A conducted study analyzed blink duration by comparing a single-task and dual-task, where the dual-task was considered as the more mentally demanding. The result showed that there were blink duration inhibitions, meaning shorter blink durations, for the dual-task compared to the single-task. The behaviour is thought to occur to avoid loss of visual information. The blinks are postponed until sufficient amount of information is obtained.

Gazing/Fixations

Fixation characteristics of the eye is also correlated to mental workload [1]. These characteristics include number of fixations and fixation durations, but also number of saccades, saccadic duration, saccadic amplitude and gaze distribution. Saccades are defined as the quick movements of the eyes when one shift vision between two fixations [7]. It has been seen that an increase in mental workload generally causes an increase in fixation duration. An increase in workload also narrows the gaze variability, called gaze concentration.

Pupillometry

Using pupillometry, meaning changes in the pupil size, as an indicator of mental workload is considered to be a significant measure [1]. The pupil size is sensitive to cognitive activities such as mental arithmetic, short-term memory load and logical problem solving [11]. A change in mental workload implies variations in the pupil diameter, where an increased workload induces a larger pupil size. The pupil size has been found to increase in proportion to the difficulty of a task [13]. When one is exposed to a demanding task, the diameter increases and remains dilated until an answer is reached. After reaching the answer, the pupil directly constricts again. The increase in pupil size during mentally demanding tasks is often shaped as an inverted V.

The pupil size changes mainly because of its sensitivity to different lightning conditions, mentioned in section 2.1. The pupil size increases as brightness decreases. The size is also sensitive to other factors such as memory load, medications, fatigue associated with cognitive tasks, stress and pain [2], [15], [23]. Therefore, an important factor when using pupillary measures is to isolate and account for these external parameters [1]. When defining fluctuations of the pupil size, it has been found that variations up to 0.5mm might be reactions due to cognitive load, while responses to illumination can vary between 2-8mm [1], [24]. The reaction to light is thus dominant over the cognitive pupillary component [23]. Also, while reading a long text the pupil diameter has been observed to decrease as an effect of fatigue [15]. Since the pupil diameter also has been observed to be sensitive to stress, it is important to make test persons feel ”natural” when measuring their eyes. Enabling some time for mental adjustment before starting a test with eye-measuring equipment is advantageous, and will most likely provide a more significant result as the test objects are not focusing on the unnatural feeling of the equipment. The pupil is also affected by accommodation reflexes. These reflexes appear as a result of changing focus from one object to another that is nearer or more distant. This can be prevented from occurring in experimental settings, by ensuring that the focal distance to visual areas of interest is kept constant.

One study regarding pupil diameter identified a pupil dilation of 50 % from its original size during the first five seconds of a memory test with digits [13]. The test persons were given a sequence of numbers short enough to be able keep in the working memory. The workload thus increased and a dilation in the pupil was observed [13]. These dilations were observed to be greater for harder tasks compared to easier ones. The pupil diameter then decreased to its original size as soon as the task was completed, i.e the working-term memory became unloaded. Another study conducted by the same research team introduced more digits to their participants than what is possible to keep in the working memory. The results showed that the pupil stopped dilating when the task became too hard, or in some cases even shrank. This indicates that the participants gave up when the task was too hard.

Task Evoked Pupillary Response, TEPR

Task evoked pupillary response (TEPR) is defined as small and involuntary changes in pupil size due to a task [25], often a System 2 task [1]. This measure can be used when estimating workload [20]. TEPR occurs a short moment after the onset when performing a task and diminish rapidly when the task is completed [25]. Example of such mental tasks are both working- (short) and long-term memory access, sentence comprehension, mental calculation, and auditory and visual perception [1]. Many pupillometry studies uses TEPR-averaging methods, meaning that multiple TEPRs are recorded and the data averaged, to obtain a reliable response of each individual's mental workload.

2.4 Human Motor Learning

When humans' perform tasks, for example in experiments, the factor of learning often has to be considered. As for this thesis project, where motor tasks primarily will be performed, the aspect of human motor learning is of interest. To gain an automatized behaviour, one has to learn and rehearse it.

Motor learning is defined as improvement through practice of a motor task [26]. The aim is to rapidly produce a sequence of movements as accurate as possible with limited effort and/or attention [27]. Motor learning consists of three stages that are processed in the following order: adaption, skills and role of explicit cognitive processes [11]. Adaption is the act of changing to a situation. It can either be physiological, i.e the eye adjusting to varying lightning conditions, or a habit that improves a function in the environment. Adaption for motor learning can be defined as *"a form or learning characterized by gradual improvement in performance in response to altered conditions"* [26]. One adapts to a task via trial-by-trial for motor-to-sensory mapping. Repeated practice, explicit instructions, trial-and-error discoveries and detecting regularity are typical for motor sequence learning [27].

Training and consequential repetition leads to a more automated operation which consequently requires less conscious activity [11]. A motor task that is performed automatically can be described as a *"highly over-learned skill that does not depend on guidance from visual feedback"* [11]. Consistent motor tasks with frequent behavior are more prone to become automatically performed whereas tasks with several variables and infrequent behavior tend to rely on cognitive control and thus processed by System 2 [4]. Motor learning cannot be adapted to all situations, i.e when they are irregular and/or where perturbations are present [26].

3| Method

This chapter starts with technical information and specifications connected to the methodology of the project, which included several sessions of collecting data with an eye tracking camera. These data collections are based on processing methods which are presented in section 3.1, and their approach in section 3.2.

3.1 Technical Information and Specifications

This section presents the eye tracking equipment and the theoretical background of the data processing methods.

3.1.1 Eye Tracker Glasses

Head-mounted eye tracking glasses equipped with infrared cameras were used to collect data [28]. Eye tracking cameras can also be remote, i.e stationary, which then do not interfere with the test object. The head-mounted glasses generally capture more details and give a higher precision than the stationary ones [20].

The equipment consisted of glasses from Sensomotoric Instruments (SMI), seen in Fig. 3.1 [28]. The company has provided eye tracking solutions since 1991 to fields within research and professional solutions. Their glasses can be combined with other instruments such as Electroencephalography (EEG) which records brain activity. The SMI glasses were used with their associated software: $BeGaze^{TM}$ 3.6.40 and $iView$ 2.7. The technology behind SMI glasses is based on distinguishing between facial features, for instance the eye, pupil, and corneal reflection. Eye movements, gaze direction and pupil size are some of the features that the SMI eye tracker calculates. The pupil size was analyzed in this project, which SMI presents in a vector with the resolution being a hundredths of a millimeter. The sampling rate was chosen to 60 samples per seconds, i.e 60Hz. The eye tracker categorizes the pupillometry measures for each data sample into a category which either is ”Visual Intake”, ”Saccade” or ”Blink”. This project excluded all samples identified as ”Saccade” or ”Blink”. The remaining samples were analyzed further. From the data files, it was observed that the pupil size during ”Visual Intake” was often equal in size before and after the saccades and/or blinks were excluded. Thus, this filtering does not create any abrupt changes in the signal. The glasses also record sound and a video, sampled with 24 frames per second, with snapshots of the eye. These recordings were also analyzed for deriving the measures.

Prior to eye tracking, calibration is necessary. This procedure measures geometric characteristics of the eye and adjusts the distance between the cameras to the current user. A preferable procedure to calibrate the SMI glasses is with a three-point calibration. The participant then attaches the glasses to fit tightly to the head and is asked to look at three points, one at a time, while keeping their head still. Two

Figure 3.1: Image of the eye tracking glasses used for data collection.

different three-point calibration images were used in this project, see Appendix A.5 and Appendix A.2.

The quality of the eye tracking for each user will depend both on physiological and external factors. Some factors tend to impair the tracking but do not necessarily occur in the experiments of this project. Factors that can complicate the tracking are for example:

- If the test person is wearing mascara
- If the test person has dark pigment spots on the iris
- If the test person is wearing contact lenses
- Shape of eyelids

3.1.2 Baseline Measurement

A baseline works as a reference point or initial value, that is measured and later used for comparisons when changes in measurements or behavior have occurred [29]. Within this project, the baseline was calculated as the mean of a data sequence. The sequence was based on samples when the participant was non-loaded, i.e performing a simple task related to the "real" task. The baseline was calculated individually for all the experiments, and also calculated for each task in the tests to acquire as high accuracy as possible. Outlier samples were then able to be identified from this, according to:

$$\text{lower threshold} < \text{valid data sample} < \text{upper threshold} \tag{3.1}$$

All samples outside this interval was considered as noise/outliers and excluded from the data that were further analyzed.

3.1.3 Data Processing of Eye Measures

The raw data from pupillary measures often contain various levels of noise. To be able to see trends and patterns, and draw conclusions, the data must first be processed. These processes involve both denoising, but also averaging for pattern searching and different transformations. This section presents theory of such processes.

Index of Cognitive Load, ICA

Index of Cognitive Load (ICA) is a signal processing technique where different reflex properties are used to separate effects of illumination and mental workload in the pupil size [1]. The method identifies reactions in pupil dilation that originates from cognitive activity [17]. The activity is identified from wavelet analysis and is then calculated automatically by a software from EyeWorks™ EyeTracking Inc. ICA is a valid method for data that is collected with head-mounted eye tracking cameras since they have a high precision [20]. Illumination is identified automatically and therefore excluded in the analysis. This feature strengthens the validity of ICA since illumination strongly affects the pupil size. Due to limited resources, this project did not have the opportunity to test this method.

Moving Averaging Filter

To discover trend components in data sets, filters can be used [30]. These filters work as functions that convert a data set or a time series into another, which with appropriate filter selections will show patterns in the original data set. One type of filter that shows trends is the Moving Average Filter (MAF). The MAF is a linear FIR (Finite-Impulse-Response) smoothing filter that with certain conditions can act as an ideal low-pass filter [31]. A smoothing technique can be applied since there is no need for online computation in this project, since all data is being processed afterwards. Smoothing uses data from the whole sequence whereas filtering only uses up to the point of interest.

The MAF representation is in discrete time defined according to Eq. 3.2:

$$\bar{x}(k) = \frac{1}{N} \sum_{i=0}^{N-1} x(k-i) \tag{3.2}$$

where $\bar{x}(k)$ is the output signal, $x(k)$ the current data sample, and N the number of elements in the window length of the MAF [31]. The window length should always be odd numbered, and the wider the window length is, the slower will the MAF transient response behaviour from the convolution be. With slower transient behaviour means a more averaged signal.

A problem with the MAF is that the convolution will not be possible to perform correctly in the beginning and end of the data set, which will induce data loss [30]. This is due to that the technique uses previous and subsequent samples, which are not available at the beginning and the end.

When implementing the MAF in Matlab, a centered symmetric filter is given by Eq. 3.3:

$$\hat{m_n} = \frac{1}{N}\sum_{j=-q}^{q} b_j \cdot y_{n+j}, \quad q < n < N - q \tag{3.3}$$

where y_n represents a data set, b_j represents weights that sum up to one and $N = 2q + 1$ being the length of the window (filter). By convolving the data set with a moving average filter, a filtered signal will become the output. An MAF filter smooths short-term fluctuations in a signal.
The implemented MAF in this project was defined according to Eq. 3.4:

$$y(n) = \frac{1}{5}x(n-2) + \frac{1}{5}x(n-1) + \frac{1}{5}x(n) + \frac{1}{5}x(n+1) + \frac{1}{5}x(n+2) \tag{3.4}$$

which represents a data averaging filter with respect to the current, two previous and two subsequent data samples. All samples are weighted equally ($\frac{1}{5}$).

Savitzky Golay Filtering

A Savitzky-Golay filter is an FIR filter, typically used to smooth a signal that contains a lot of noise and also often has a large frequency span [32]. The Savitzy-Golay filter tends to perform well under these circumstances compared to other standardized FIR averaging filters, since they have a tendency to filter out data that is a part of the high frequency content together with the noise. The Savitzky-Golay filter smooths the data by minimizing a local least-squares error polynomial approximation by fitting the polynomial to frame elements within a data set [32], [33].

A Savitzky-Golay filter is applied in Matlab with its built-in function *sgolayfilt* [32]. The input to this function is the data set x to be avereged, *order*, that determines the order of the fitting polynomial, *framelen*, which sets the frame length. The frame length must be odd numbered and longer than the order. This project had an order of 9 and frame length of 401.

Fourier Transform

All signals in the time domain can be described as waveforms by the sum of sinusoidals with different frequencies. The Fourier transform utilizes sine and cosine functions in order to identify the frequency content of a signal by first analyzing it in the time domain [34]. After finding the frequency domain characteristics of the signal, the time domain function is translated into the frequency domain and analyzed by using Fourier coefficients. These coefficients represent the contribution of every cosine and sine function for each frequency. Many operations can be performed in the frequency domain, such as distinguishing certain frequencies. The result from these frequency operations can then be transformed back to the time domain by using an inverse Fourier transform. The Fourier transform consists of multiplying a signal, *f(t)* with a window function g which is centered around 0. The Fourier coefficients are then computed from the product of $g \cdot f$ which gives an indicator about the frequency content of f when the time is approximately 0, t_0. This procedure is repeated for different time steps for the window function, i.e $g(t \pm t_0), g(2t \pm t_0)$.... These collection of Fourier coefficients can be presented as:

$$c_{mn}(f) = \int_{-\infty}^{\infty} dt e^{im\omega_0 t} g(t - nt_0) f(t) \tag{3.5}$$

The Fourier transform is not localized in space and the data is presented as oscillations in an infinite time [34]. A disadvantage with the Fourier transform is that it does a poor job in identifying abrupt changes in an efficient way.

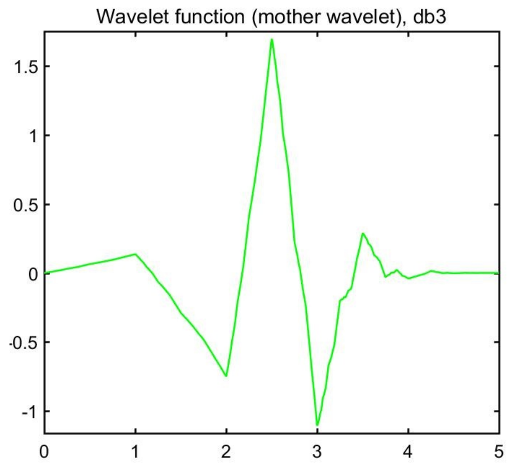
In comparison to using the Fourier transform, the power distribution of a signal can also be used to present the power spectral density or frequency content of a signal [35]. A power spectrum contains essential information for a raw data signal regarding its power distribution, which describes the power of the signal in the form of averaged frequencies over different spectra.

The built-in function named *periodogram* in Matlab was used in this project to study the frequency content. It returns a vector, where each data point in the vector represents the power spectral density (PSD) estimate corresponding to the same position of the input vector [35]. The output vector, showing the PSD estimates, has the unit of squared magnitude of the corresponding time series data, per frequency unit.

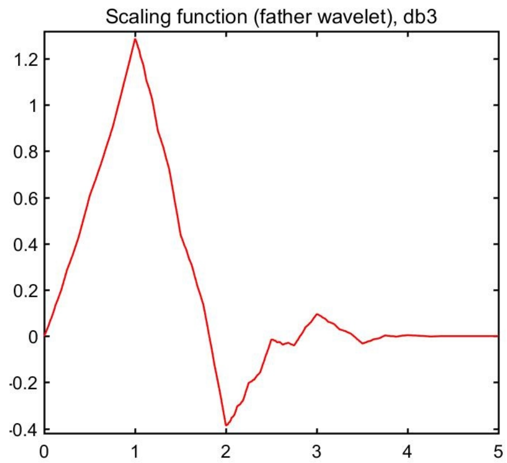
Wavelet Transform

When analyzing non-stationary signals, also called transient signals, the time- frequency tool Wavelets can be efficient to use [36]. In contrast to Fourier transform, which only localizes a signal in frequency domain, wavelet analysis localizes it in both frequency and time domain, and performs well when signals have discontinuities and sharp spikes [36]. It is therefore effective when extracting structures in the non-stationary signal that might be difficult to generate from other signal processing tools. These structures can for example be repeated patterns, trends in the raw data and/or discontinuities. A wavelet transform is a mathematical function that is convolving a pre-defined oscillating signal, also called wavelet function, with a raw data signal. The wavelet transform divides the resulting signal and presents it in different decomposition levels, which are representing features of the signal in different sub-groups. The sub-groups represent different scales and frequency span, depending on the features of the signal, and all decompositions will together sum up the original signal [34]. The wavelet function describes a wavelike oscillation with limited duration and a varying window length. These waves can be of different shapes and sizes but all have zero mean amplitude [36]. The idea behind wavelet transforms is to study and analyze the raw data signal with the wavelike component varying in resolution and scale, and distinguish where the varying wave is matching the raw data signal.

Wavelet transforms (WT) can either be continuous (CWT) or discrete (DWT). At CWT the raw data signal is convolved with the wavelet function in continuous time. The discrete wavelet transform is produced by taking the inner product of an original raw data signal, with the wavelet function at discrete number of data-points, which results in the decompositions. The wavelet function is the base for wavelet transform, often referred to as the *analyzing wavelet* or *mother wavelet* [34]. When defining the wavelet transform, the wavelet function first has to be determined. There are a lot of different wavelet families to consider. The different wavelet families all have different characteristics, meaning they describe different wavelike oscillations, but all wavelet functions within a family are derived based on the same mother wavelet. All the functions must satisfy the conditions for wavelets such as zero mean amplitude, continuity and limited duration [36] [37]. For this project, the DWT was applied, with the orthogonal and compactly supported wavelet from the Daubechies (db) family as wavelet function. Their sub-classes are classified by the level of iteration, number of coefficients and number of vanishing moments [34]. Daubechies wavelets are advantageous in representing polynominal behavior, and the db family order 3 wavelet function was determined to be used, since its behaviour and mathematical function matched well with the raw data signal of pupillary measures. The mother wavelet is displayed in Fig. 3.2a.



(a) The wavelet function, also called the mother wavelet, used in the project. Generated by Matlabs wavelet-display menu



(b) The scaling function, also called the father wavelet, used in this project. Generated by Matlabs wavelet-display menu

Figure 3.2: Image displaying the Daubeiches family order 3 wavelet function and scaling function. The plots are generated by Matlab Wavelet toolbox.

Wavelet transforms have two essential operations: dilation (scaling) and translation (shifting) [38]. The dilation factor is determined from a scaling function, also referred to as the *father wavelet*, which changes the signal in time, either stretches out or compresses it. The compressed wavelet (small scale factor) identifies abrupt changes in the signal and corresponds to high frequencies. A stretched wavelet (large scale factor) captures slowly varying changes and corresponds to low frequencies. The shifting either delays or advances the onset of the wavelet. The shifting is needed in order to align with the feature that one is looking for in the raw data signal [37]. The father wavelet used in the project is displayed in Fig. 3.2b.

By convolving the mother and father wavelet, a number of children wavelets will be produced, according to the number of decomposition levels chosen for the transform. These children wavelets will originate the decompositions, by convolutions one by one with the raw data signal, and be presented in levels showing features in different resolutions and frequency spans. The decomposition levels show both approximations of the signal, a, and details, d, at each level, see Fig. 3.3.

The resulting decompositions show at what frequency the energy of the signal accumulates, but also the localization in time of where different frequency behaviours occur [36]. To generate the different approximations of the signal, a1 to a5, details d1 to d5 are superimposed on the approximations. The first approximation, a5, contains the least details. By superimposing d5 on a5, a4 will be obtained Then, a3 will be obtained by superimposing of d4 on a4, and so on. Finally, the original signal s will be produced, from superimposing d1 on a1. The advantage with the multi-resolution representation of the signal in different sub-spaces, seen in Fig. 3.3, is that behaviours located in specific sub-spaces can be extracted and analyzed separately [36]. The frequency span in which the sub-spaces in the multi-resolution representations operate are related to the sampling frequency, f_s. For the equipment used in the project, which had a sampling frequency of 60Hz, this first sub-space corresponded to the span from half the sampling frequency, 30Hz, up to the sampling frequency, 60Hz. The frequency span for the second sub-space is defined as a fourth the sampling frequency up to half of the sampling frequency, and so on. That results in frequency spans for each sub-space according to Tab. 3.1.

When extracting decomposition levels in this project, a root-mean-square (RMS) envelope filter with a length of 50, was applied in order to smooth the wavelet analysis signals.

Decomposition level	Frequency span
d1	30Hz - 60Hz
d2	15Hz - 30Hz
d3	7.5Hz - 15Hz
d4	3.75Hz - 7.5Hz
d5	0Hz - 3.75Hz

Table 3.1: Frequency spans for each decomposition level of the multi-resolution representation of the wavelet analysis, correlated to the projects sampling frequency of 60Hz.

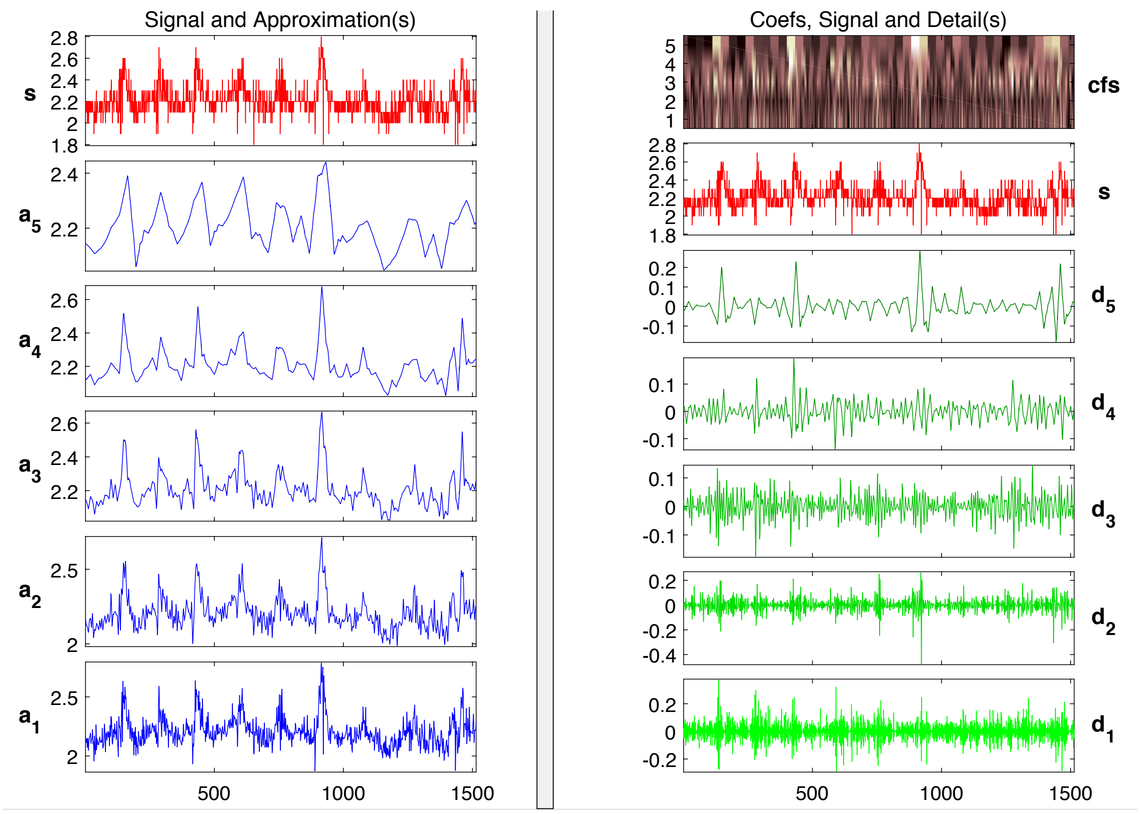


Figure 3.3: An example of how the resulting multi-resolution analysis from a wavelet transform is displayed. The red signal s represents the real signal, and below are five decomposition levels presented, where each level includes one approximation, a, of the signal and one detail signal, d. The real signal can be reproduced by superimposing the details on the approximations one by one.

3.2 Methodology

This section presents detailed descriptions of the three conducted data collections. The experimental set-up for the whole thesis, which includes both literature review, Pilot 1, Experimental set 1 and Pilot 2 can be viewed in Appendix A.1. The project had an iterative process, meaning that each data collection was concluded and findings were implemented in the method before performing the following data collection. Pilot 1 was a first trial with purpose to analyze pupil behavior and compare different tasks when measuring the eyes. Conclusions were drawn and led to Experimental set 1 where 13 new test persons participated. A last data collection, Pilot 2 was then performed with some adjustments from Experimental set 1.

3.2.1 Pilot 1

Pilot 1 consisted of four participants who attended one test session that lasted approximately 20 minutes. The participants were instructed to perform two unrelated tasks while wearing eye tracking glasses. The recruited participants for Pilot 1 were co-workers who were asked to contribute to the study. The session started with an introduction to the project and specifically the aim of Pilot 1. The participants were then asked to attach the glasses to their heads and adjust them to fit tightly. A three-point calibration was performed using an image with three black circles, see App. A.2, which the test person were told to focus on one by one. The participants were instructed to not touch or relocate the glasses during the test session, since that would change the calibrated distance and thus affect the data. The test session was divided into two parts, one connected to each sub-task. The two conducted tasks for Pilot 1 were Stroop task and a motor piano task, both presented on a computer screen. The schedule for Pilot 1 can be viewed in Fig. 3.4, including both introduction, task one and task two.

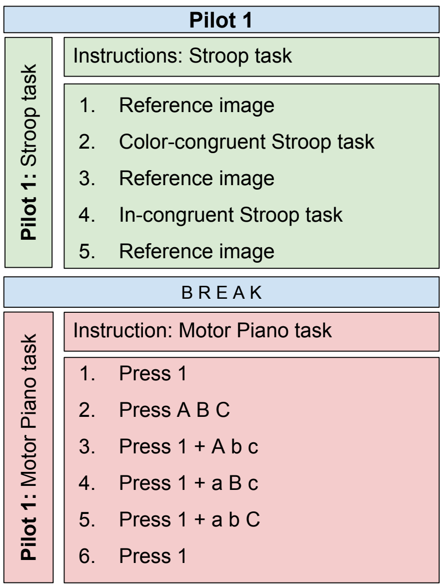


Figure 3.4: Pilot 1 test setup with details of the performed tasks.

Stroop Task

A Stroop task was performed, according to section 2.3.2. To ensure that the participants were not color blind and that they had the same perception of colors, a reference image showing congruent words with all colors included was used, see App. A.3. The recording started and ended with a reference image on the computer screen, see App. A.4, which purpose was to obtain a baseline (non-loaded) measure. The participants were informed that each time the reference image appeared, they should focus on the character to the left, and then systematically focus approximately one or two seconds on each character from left to right. The Stroop task started with the normal condition, meaning color-congruent words. The conflict condition, the in-congruent words were then presented directly after the normal condition, with a reference image in between. The sheets that were used can be viewed in App. A.5 and A.6.

Motor Piano task

The second task in Pilot 1 was a motor piano task. The participants were instructed to play a virtual piano using the computer keyboard [39]. The participants were told to look at the computer screen which was all black, so the pupils were unaffected by different lightning conditions. The left hand was instructed to only play one note, by pressing the key w, marked as 1. This sequence acted as a baseline task. The right hand had three possible keys to press; R, Y or I marked with yellow tape and a black character as A, B and C respectively. The participants were verbally instructed to play different combinations, according to Tab. 3.2. These combinations were performed repeatedly, until the next one was requested. The order of which the key pressing combinations were requested are presented in Tab. 3.4

Table 3.2: A summary of the key pressing combinations that were instructed verbally to the participants in Pilot 1. The possible combinations were 1, abc, 1A, 1B and 1C

Command	Instructions
1	Press only key 1 with left hand three times
abc	Press key a, then b, and lastly c with right hand
1A	Press key 1 with left hand and key A with right hand simultaneously, then only B with right hand and lastly only C with right hand
1B	Press only key A with right, then press key 1 with left hand and key B with right hand simultaneously, then lastly only C with right hand
1C	Press only key A with right hand, then press only key B with right hand, and lastly key 1 with left hand and key C with right hand simultaneously.

This task is according to the dual-task theory inducing two tasks, where the primary task was to play just one key with left hand. The other hand was added as a secondary task, with the instruction to alternate synchronized key pressings between left and right hand.

Isolation of Parameters

As described in section 2.3.3, the pupil size is sensitive to many parameters. To isolate as many as possible of these, the experimental design was adapted with the parameters in mind. These are presented in the list below.

- All participants were sitting in the same room with equal illuminance.
- The baseline was calculated from two data sequences, one before and one after each task was performed
- The test was designed such that the participants were relatively still in their position of the body and head, since the measurements might be affected by artifacts from movements of the head etc.
- The keys that were used in the combinations were marked in order to minimize the extraneous load.
- During the motor task where the online piano was used, no auditory feedback was given from the computer. This was to prevent the test persons from hearing the notes, and thus load the phonological loop in the working memory. There is though a possibility that test persons hear the key pressing sounds in their head, which then generates auditory feedback from the process. This aspect was not further analyzed.

Signal Processing

After the data collection for all four participants were finished, several signal processing operations were performed. The operations were mainly focusing on denoising and filtering the data such that general trends could be found. The performed operations were:

- Including only *Visual Intake*, meaning excluding data samples categorized as *Saccades* and *Blinks*
- Setting *Valid data* according to Eq. 3.1 with thresholds generated from inspecting the results, to filter away obvious outliers for each test person based on their individual baseline.
- Moving Average Filter (centered symmetric filter with coefficients $b_j = \frac{1}{5}$)
- Savitsky-Golay filtering (order 9 and frame length 401)

Since some of the collected samples were excluded from the analysis, the signal is continuous in samples but not in time.

3.2.2 Experimental Set 1

Experimental set 1 is considered to be the main test of the project. After concluding the results and findings in Pilot 1, see section 4.1 and 4.2, the test method was modified.

A motor task was chosen as the primary task because of the lack of any verbal response. The working memory did thus not get extra load from "digging” for answers to reply. Experimental set 1 included two occasions for each participant, session 1 and session 2, with approximately one week in between. The same schedule was followed during both sessions. Experimental set 1 consisted of a motor task, divided into different sub-tasks, where the test persons’ cognitive load varied. The idea with one week in between the sessions was to allow the test persons to rehearse the motor task at home. With enough amount of rehearsal, the expectation was that the motor task would be (partly) automatized during session 2, and thus require less mental capacity of the working memory. This decrease would hypothetically be possible to see in the pupillary measures. During session 1, where both the test setup and the motor task were unfamiliar, the expectation was that the test persons would be more mentally loaded and the tasks would require more cognitive control. An image describing the set up schedule can be seen in Fig. 3.5. The main difference between session 1 and session 2 was that test session 1 started with mandatory instructions, both verbal and visual, while session 2 started with repeating the visual instruction if needed.

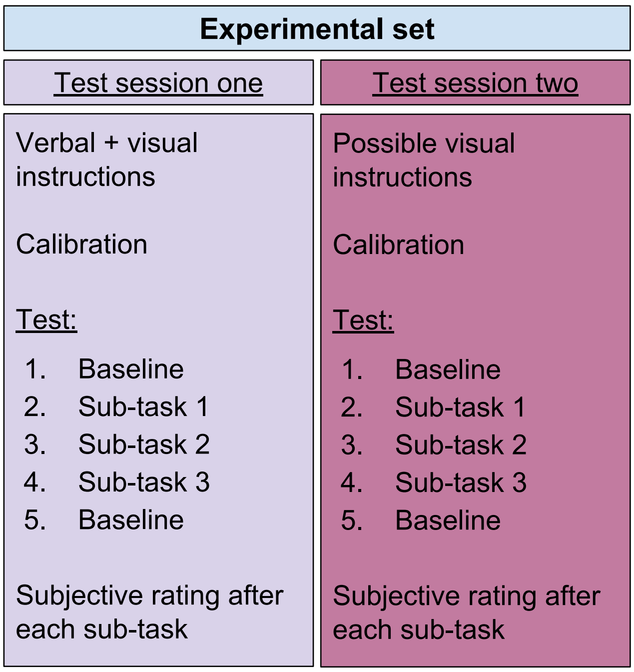


Figure 3.5: The Experimental set 1 test setup, with details of both session 1 and 2. The main difference was that test session 1 started with mandatory instructions, both verbal and visual ones.

The motor task was, similar to Pilot 1, a piano influenced key pressing task. The task did during Experimental set 1 contain more details. It was influenced by the Rhythm method, Dual-task and TERP, see section 2.3.2. The same verbal instructions for the key pressing combinations were used, see Tab. 3.2. In comparison to Pilot 1, Experimental set 1 included both verbal and visual instructions for the key pressing combinations, in order to minimize the extraneous load. The visual instructions are found in App. A.8. The task was introduced during session 1 and rehearsed before starting the eye tracking. After the session was finished, i.e performed both baseline tasks and the three sub-tasks, they were instructed to rehearse the motor task as much as possible during the week in between the sessions. The optimal was 5-10 minutes a day. They were also aware of the importance with their honesty when reporting how many minutes they had rehearsed in total, for the study to be able to draw the right conclusions.

The combinations *1* and *abc* were used as baseline tasks, since they only involved one hand. These baseline tasks were used to reach as close as possible to automatized motor behaviours, since it was found from the literature review that pushing buttons in a self chosen pace would be automatized. The baseline combination *1* was included in test tasks one and five in the experimental set 1 test set up, which also can be seen in Fig. 3.5. The combinations *1A*, *1B*, and *1C* represents the three sub-tasks in 3.5 and were included in the data analysis.

Six different test designs were created to ensure that the testing was not biased by the task sequence. These decided in which order the three motor tasks (1A, 1B and 1C) were performed. The test persons were assigned one test design which then were identical for both sessions. The designs are presented in Tab. 3.3.

Table 3.3: The six different test designs for the motor task that were used during the Experimental set 1 data collection

Design 1	**Design 2**	**Design 3**	**Design 4**	**Design 5**	**Design 6**
1A	1C	1B	1A	1B	1C
1B	1A	1C	1C	1A	1B
1C	1B	1A	1B	1C	1A

To be able to control and vary the mental workload during the sub-tasks, they consisted of one part with only key pressing and one with key pressing and mental arithmetic simultaneously, hereby referred to as math-part. The duration of each of these parts can be seen in Fig. 3.6. The mental arithmetic consisted of addition, where the numbers were relatively simple. The arithmetic used during the test sessions are presented in App. A.9. The arithmetic questions were asked continuously throughout the math-part and the number of math questions thus depended on how quickly the test persons responded with a solution. They got an average of five to six questions per sub-task. Each sub-task did also end with a subjective rating questionnaire, see App. A.10.

Test session: Experimental set 1

Baseline 1:
- 20 seconds motor task 1
- 20 seconds motor task 1 + math questions

Subjective rating

Sub-task 1:
- 30 seconds motor task 1/3
- 30 seconds motor task 1/3 + math questions

Subjective rating

Sub-task 2:
- 30 seconds motor task 2/3
- 30 seconds motor task 2/3 + math questions

Subjective rating

Sub-task 3:
- 30 seconds motor task 3/3
- 30 seconds motor task 3/3 + math questions

Subjective rating

Baseline 2:
- 20 seconds motor task 1
- 20 seconds motor task 1 + math questions

Subjective rating

Figure 3.6: Summary of the time schedule for the sub-tasks during a test session, both with and without mental arithmetic.

Recruiting the Test Persons

The participants were recruited by contacting randomized employees at Volvo Cars. In order to be able to participate in the study, specific criteria had to be fulfilled. The criteria on the test persons were:

- Understand and speak fluent Swedish
- An age between 25-55 years
- No glasses during the sessions. Contact lenses were accepted.
- No intake of prescription medicine
- Not suffer from any chronic disease
- Not intake any coffee two hours prior to the test
- Not wear mascara during the sessions

The importance with understanding and speaking fluent Swedish was to eliminate uncontrollable parameters such as not receiving the equivalent information and instructions in different languages. It was also for ensuring that all test persons had the same prerequisites when interpreting the sessions and instructions.

To ensure that the pupillary measures did not change due to other factors than cognitive load, the criteria for medications, chronic illness, and coffee (caffeine changes) were added.

A summary of the recruited test persons who participated in Experimental set 1 can be seen in Tab. 3.4.

Table 3.4: Summary of the recruited test persons for Experimental set 1, showing the number of test persons, the mean age, the span of age for the test persons and the gender distribution.

Test persons	Average age	Lower age	Upper age	Women	Men
13	42.5	39	55	23 %	77 %

Isolation of Parameters

Improvements from Pilot 1 generated isolation of the following parameters, in addition to the ones isolated in Pilot 1:

- To isolate accommodation reflexes and illuminance changes, the test persons were told to focus as much as possible at a white cross lying on the white table. The cross was placed 80cm from the edge of the table, straight in front of the seating position. A white keyboard was used with all keys, except for those used in the test, covered with white electrical tape, to ensure that the test object did not focus on the wrong keys.
- All tests were performed in a room with the same illuminance both before, during and after the test sessions. To control the illumination level the luminosity was measured with a portable light intensity measurement tool at the seating position in each test location.
- To not cause unnecessary stress, the image describing the five different key pressing commands, App. A.8, was lying on the table during the test sessions. The test persons were instructed that it was accepted to glance at the image if repetition of a command was necessary, or look at the keyboard if needing correct the position of their fingers. The focus should then return to the white cross as soon as possible.
- A test run was performed after the instructions, to ensure that the test person fully understood the instructions correctly and to try to eliminate stress or excitement prior to the test.

Measuring the Cognitive Load

As stated in section 2.3.1, there are three different categories from which one can measure cognitive load. These three were included in the experiment design and their implementations are presented one by one below.

Subjective Rating

A subjective self assessment questionnaire was used to capture the test persons own perception of each sub-task. The questionnaire was based on NASA-TLX, but with only one question adapted to the test subject and then translated into Swedish. This questionnaire can be seen in App. A.10. The test objects responded verbally with a number between 1-7 with their estimated mental workload, both for the sub-task with and without mental arithmetic. This subjective rating was performed after each baseline and sub-task.

Performance Rating

The keyboard was connected to a computer that ran a Matlab function which collected what keys were pressed and the time interval between every press, to identify individual tapping patterns and possible errors.

Physiological Measures

The physiological measures consisted of data from the eye tracker which included pupil diameter for both left and right eye, the category of which action the eye had in each data point (including Visual Intake, Saccade or Blink), snap shots of the eye and a recorded sound file.

Signal Processing

The filtering and signal processing operations performed on the Experimental set 1 data were:

- Including only *Visual Intake*, and thus excluding *Saccades* and *Blinks*.
- Setting *Valid data* according to Eq. 3.1 with thresholds generated from inspecting the results, to filter away obvious outliers for each test person based on their individual baseline.
- Using 1D discrete wavelets to denoise the signals. The inbuilt wavelet minmax function was applied, with soft threshold, no rescaling and 5 levels of decomposition. The MAF and SGF filtering operations were excluded since it was found that wavelets outperforms these.
- Analyzing the filtered data in time, both the pupillary responses to math questions, but also the general pupil behaviour over a longer time period for the part with no math questions.
- Conducting a t-test to compare the generalized pupillary changes for sequences in the sub-tasks when no math questions (and thus no disturbance). The level of significance was defined as p= .05
- Analyzing normalized data in the frequency domain for one of the sub-tasks.

Since some of the collected samples now are excluded from the analysis, the signal will be continuous in samples but not in time.

3.2.3 Pilot 2

Pilot 2 represents the last test session which included four participants. The experiments were conducted with the same approach and procedure as Experimental set 1, with two adjustments, as follows:

- The participants were strictly told to practice the motor combinations to make them more automatized during the second session.
- The math questions were recorded on beforehand and communicated via a speaker, to ensure that all participants got the same number of math numbers and at the same time instances. The math questions were presented every 8^{th} second and thus implied an equally consistent load for all participants during both session 1 and 2.

Test session: Pilot 2

Baseline 1:
- 20 seconds motor task 1
- 40 seconds motor task 1 + 5 math questions

Subjective rating

Subtask 1:
- 30 seconds motor task 1/3
- 48 seconds motor task 1/3 + 6 math questions

Subjective rating

Subtask 2:
- 30 seconds motor task 2/3
- 48 seconds motor task 2/3 + 6 math questions

Subjective rating

Subtask 3:
- 30 seconds motor task 3/3
- 48 seconds motor task 3/3 + 6 math questions

Subjective rating

Baseline 2:
- 20 seconds motor task 1
- 40 seconds motor task 1 + 5 math questions

Subjective rating

Figure 3.7: Summary of the sub-tasks for Pilot 2 test session. The summary includes the parts both with and without mental arithmetic.

The data from Pilot 2 was analyzed similar to Experimental set 1, except for one additional Wavelet analysis of decomposition details.

4| Results and Initial Conclusions - Pilot 1

This chapter covers the results from the first data collection, Pilot 1, which is included in the project for comprehension of pupillary data. This is followed by conclusions and hypotheses which the next data collection investigated in.

4.1 Results

Pilot 1 consisted of four participants, with a gender distribution of three females and one male. The results were primarily produced and analyzed to get a general comprehension of the characteristics of pupillary measures. The raw data signals were filtered using a moving averaging filter (MAF) and a Savitsky-Golay filter (SGF), according to implementations in section 3.1. The results for one of the participants are displayed in Fig. 4.1, where the session started with the Stroop task and ended with the motor piano task. The upper plot in the figure displays the raw data signal for the pupil diameter during the test session which lasted for 325 seconds, resulting from the sampling rate of 60Hz and approximately 20 000 samples. As mentioned in section 3.2, the raw data signal was filtered according to outliers and ”Visual Intake”, which made it continuous in samples but not in time. Therefore, the x axis presents samples. The middle plot displays the raw data signal filtered with an MAF, and the lower plot displays the resulting signal after filtering the MAF signal with an SGF.

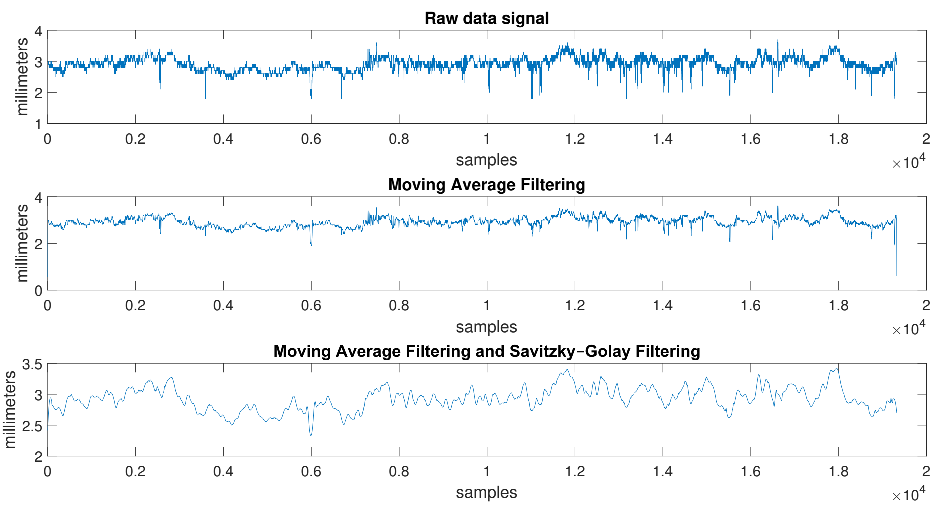


Figure 4.1: Three different signals for Pilot 1 for one of the participants. The top plot shows the raw data signal of the pupil diameter, the middle plot being the result from filtering the raw data with an MAF, and the lower plot the MAF signal filtered with an SGF.

It is noticeable that the pupillary measures have both small and rapidly changing fluctuations, and a slowly changing pattern over time.

By visual inspection of the Pilot 1 raw data, the thresholds values for outlier filtration were defined according to:

$$0.4 \cdot \text{baseline} < \text{valid data sample} < 1.75 \cdot \text{baseline} \tag{4.1}$$

4.2 Initial Conclusions

After analyzing the findings from Pilot 1, several conclusions were drawn. In general, the unfiltered pupillary data contained a lot of fluctuations, as can be seen in the upper image of Fig. 3.4. Several filtering passes were needed, and a definition of individual baselines were required for each test person, since that measure varies a lot among the individuals. When analyzing the raw data signal plot in Fig. 3.4, sharp and sudden changes up to 1mm are identified. These were considered as outliers, since the pupillary changes related to mental workload, according to theory, might be up to 0.5mm. The following conclusions were drawn from Pilot 1:

- For pupil diameter changes, factors such as illuminance matters. The most accurate results are most likely achieved when the test room is as light as possible, since it will be easier to identify changes in pupil size when its baseline is relatively small. The illuminance should be kept constant during the whole test.
- It is important and necessary to have consistent instructions, since the results will be more reliable when all test persons have the same prerequisites. Consistent instructions will minimize the extraneous load. All test persons should thereby be given the exact same instructions, both written and orally.
- Extracting and calculation of baseline should be over a period of time, possibly in both beginning and end of the data sequence. This is due to that the pupil diameter shrunk over time during the test sessions, which might have appeared from relaxation and overcoming of excitement and/or nervousness.
- Surrounding objects in the room might steal focus and thus affect the pupil behavior. To achieve as equivalent data as possible, all test persons should focus on the same spot/field during the sessions. The room should also be free from disturbing objects in the gaze field. Therefore, a computer screen should not be used, which might reflect objects and also have different colors and frames that are possible to focus on.
- The motor task in Pilot 1 might be possible to automatize, with the help of rehearsing. When rehearsed, and hopefully more automatized, an additional task can be used to disturb the test persons with the purpose to vary their mental workload.
- The initiated task should reduce the risk of rapid eye movements as much as possible, to decrease the amount of saccades in the data.

The results from Pilot 1 also gave rise to three hypotheses that Experimental set 1 was expected to investigate in, according to:

- **Hypothesis 1:** The pupil dilation will increase when adding a second task.

- **Hypothesis 2:** The performance, measuring what keys that were pressed and the time in between each pressing, will impair with increased cognitive load (meaning applying the second task). With an impaired performance means longer time in between the pressings and more errors by pressing the wrong buttons.

- **Hypothesis 3:** The combination 1B will be the most complex and difficult sub-task to perform, and thus automatize.

All of these conclusions and hypotheses were implemented in the experimental set-up as improvements for Experimental set 1 methodology.

5| Results and Initial Conclusions - Experimental Set 1

This chapter will first present the results from Experimental set 1, which is seen as the main data collection of the project. The result section starts with presenting the subjective self assessment rankings. An analysis regarding the math-parts are then presented by initiating triggers when the math questions are asked by the test leader and then answered by the test person. An analysis on the sequences without math is later presented with trendlines and the results are finalized with a frequency analysis which is conducted on both the sequences with and without math. The conclusions from the results are then stated as well as hypotheses for the last data collection.

5.1 Results

Experimental set 1 consisted of 13 participants for both session 1 and 2, with a gender distribution of three females and ten males. Additional information about the test persons are presented in Tab. 3.4. A summary of the assigned test design for each test person, time of rehearsal between session 1 and 2 and illuminance in the testing rooms are found in Tab. 5.1. The rehearsing is presented in minutes, gathered from the test persons own estimate of the amount of training they had performed between session 1 and session 2. Some of the participants stated that they had rehearsed the task by thinking of it and not rehearsing it physically. This is clarified as ”Thinking” in Tab. 5.1. The illuminance for each session is displayed as ”Lum” and measured in lux in the table.

Table 5.1: Description of participants, their test designs and illuminance levels for Experimental set 1. The number of minutes they spent on rehearsing between the sessions is also presented.

Test object	Test design	Lum. s1	Lum. s2	Rehearsal
P1	1	441 lux	620 lux	15 min
P2	2	441 lux	620 lux	0 min (thinking)
P3	3	441 lux	620 lux	20 min
P4	4	441 lux	620 lux	20-30 min
P5	5	441 lux	620 lux	5 min (thinking)
P6	6	441 lux	620 lux	0 min
P7	1	620 lux	560 lux	5 min
P8	2	620 lux	620 lux	0 min (thinking)
P9	3	620 lux	560 lux	3-4 min
P10	4	620 lux	620 lux	10-15 min
P11	5	620 lux	560 lux	3 min
P12	6	620 lux	620 lux	0 min
P13	1	620 lux	620 lux	10 min

After collecting the data and performing filtration according to the experimental method, see section 3.2.2, a varying number of valid data points were generated for each test person and session. These are presented in Tab. 5.2. The data from the right eye was chosen for all participants and sessions, since it generally included a higher amount of valid data points compared to the left eye. It was also desirable to measure the same eye for each individual both session 1 and 2. P11 was mistakenly sampled with 30Hz, and therefore has less data points.

In the data analysis, results for test persons P2, P7 and P10 are primarily presented in the result section in the report. The remaining participants are be presented in the appendix chapter.

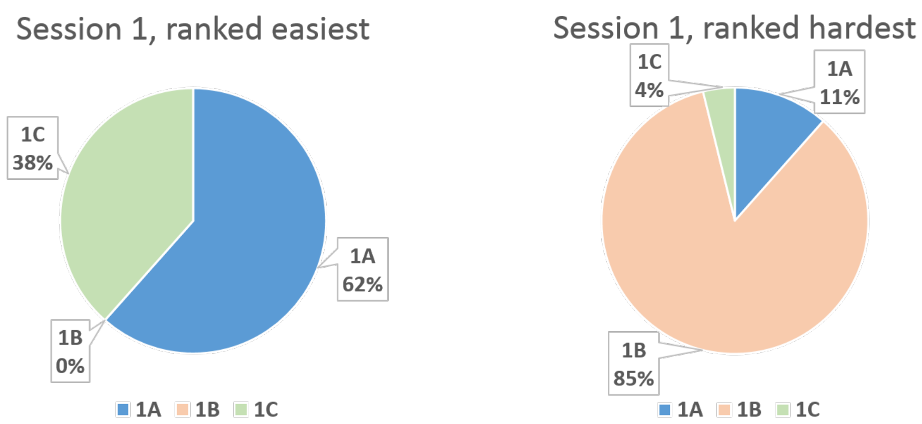
Table 5.2: Number of data samples for Experimental set 1, session 1 and session 2, both raw data and filtered data. The ratio, calculated as the filtered data points divided by the unfiltered data points, indicates how good the eye tracking equipment performed for each test person.

	Session 1			**Session 2**		
Test person	**Unfiltered**	**Filtered**	**Ratio**	**Unfiltered**	**Filtered**	**Ratio**
P1	23 341	12 826	55 %	21 541	8 837	41 %
P2	26 581	20 467	77 %	25 141	20 601	82 %
P3	25 021	14 800	59 %	23 341	14 908	64 %
P4	25 201	20 691	82 %	22 981	20 056	87 %
P5	27 601	14 956	54 %	24 721	15 058	61 %
P6	26 221	16 713	64 %	24 001	15 397	64 %
P7	22 441	17 079	76 %	20 581	16 100	78 %
P8	23 161	18 480	80 %	22 261	18 371	83 %
P9	22 381	17 817	80 %	21 481	18 414	86 %
P10	23 461	17 176	73 %	22 441	16 503	74 %
P11	13 501	8 369	62 %	22 441	16 491	73 %
P12	23 881	18 182	76 %	23 221	18 070	78 %
P13	27 601	22 324	81 %	23 761	18 890	80 %

A limit of 55% of valid data was defined as an acceptable ratio for the project. This limit was determined by visual inspection of the collected data in Experimental set 1. There were three test persons excluded from the data analysis. P1 was excluded due to technical issues with the equipment, which resulted in few data points and thus a low ratio of valid data. P5 was also excluded due to low ratio of data points, but also due to poor quality of the valid data. Lastly, P11 was excluded due to being sampled with 30Hz instead of 60Hz. The data analysis with data points included, meaning the physiological results, did therefore include results based on test persons P2, P3, P4, P6, P8, P9, P10, P12 and P13. For the subjective self assessment results, all thirteen participants were taken into account. Because of limited amount of time, the performance measurements were not analyzed.

5.1.1 Subjective Self Assessment Ranking

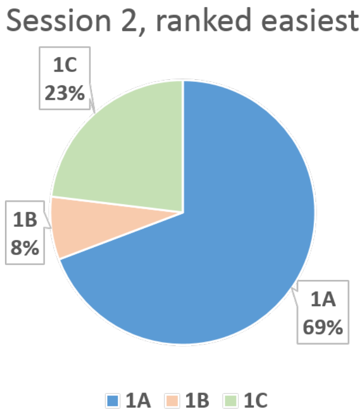
Results from the subjective ranking of the sub-tasks, generated by the self assessment questionnaire, are presented in App. B.1 and includes all participants. The rankings are presented both for the baseline tasks, meaning baseline 1 and 2 in Fig. 3.6, as well as the three sub-tasks, where the ratings between 1-7 that the test persons responded verbally with are presented. The subjective ratings of both sequences, meaning the part without math and the math-part, were summarized for each sub-task and session, resulting in a number in the column *Sum, mental workload*. These sums are in the column *Rank, mental workload* ranked in relation to each other and assigned to a number between one to five. The task(s) with rank one was/were thus interpreted as the easiest by the test person, and the task(s) with rank five as the hardest. If two sub-tasks got the same number, they were interpreted as equally hard or easy. The results from the subjective rankings were simplified and summarized according to which of the three sub-tasks, meaning 1A, 1B and 1C, that each test person ranked as easiest and hardest. These are presented for each session, where session 1 can be seen in Fig. 5.1 and session 2 in Fig. 5.2. Each color represents one of the sub-tasks, and the percentages how many out of the 13 test persons that ranked each task as either easiest or hardest.



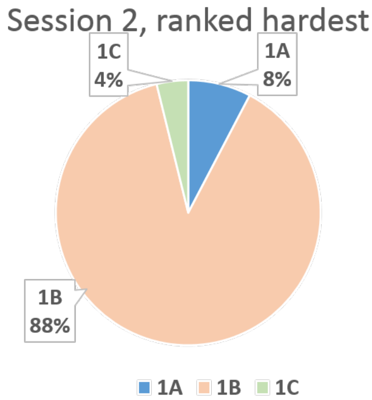
(a) Self assessment ranking of the easiest task for session 1, where 1A was ranked as easiest

(b) Self assessment ranking of the hardest task for session 1, where 1B was ranked as the hardest

Figure 5.1: Summary of the subjective self assessment rankings for session 1, including tasks 1A, 1B and 1C. Fig. 5.1a illustrates the distribution of tasks that were ranked as easiest, and Fig. 5.1b as hardest, where the percentages show how many of the 13 participants that ranked each task.



(a) Self assessment ranking of the easiest task for session 2



(b) Self assessment ranking of the hardest task during session 2

Figure 5.2: Summary of the subjective self assessment rankings for session 2, including tasks 1A, 1B and 1C. Fig. 5.1a illustrates the distribution of the tasks that were ranked as easiest, and Fig. 5.1b the distribution for the tasks ranked as hardest. The percentage represent how many test persons out of the 13 participants that ranked each task as the easiest or hardest.

One can observe that during both sessions of Experimental set 1, task 1A was ranked as easiest for the majority of the test persons, 62 % for session 1 and 69 % for session 2. It can also be observed task 1C was ranked as the easiest for 38 % for session 1 and 23 % for session 2. Task 1B is ranked as the hardest task for both sessions according to the subjective ratings. Further data processing were thereof focused on analyzing the results from sub-task 1A, since it was thought that that task had the highest probability of becoming automatized.

5.1.2 V-shapes and Triggers in Time

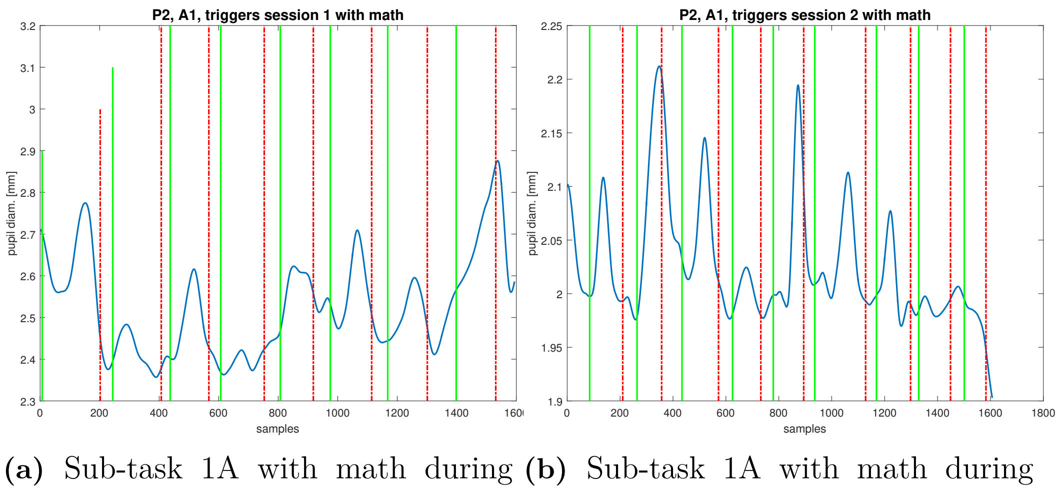
Pupil behavior related to the sequence with math were analyzed by applying triggers at the time stamp when the math questions were initiated, meaning when the test leader started to verbally ask the question, and when the questions were replied to, meaning when the test person had responded with an answer. These triggers are based on normalized data, y_{norm}, which is the values of the absolute pupillary changes. The normalized data was calculated as:

$$y_{norm}(i) = x(i) - \mu \tag{5.1}$$

where x represents the filtered data samples and μ being the mean of the pupil diameter for each sub-task with no math. The normalized data was calculated based on filtered samples and are therefore not containing any artifacts.

The trigger responses are presented for test persons P2, P7 and P10 for session 1 and 2 in Fig. 5.3a- 5.5b. Observe that the axis limits are differing between the individuals and sessions in this analysis, since all participants have their individual baseline and response. This analysis was only performed for sub-task 1A, and the resulting plots for remaining test persons can be found in App. B.6. In the figures, the blue fluctuating line represents the pupil diameter change during the math-part in sub-task 1A. The green vertical solid line represents the start of each math question and the red vertical dashed lines represent the end of each trigger/math question.

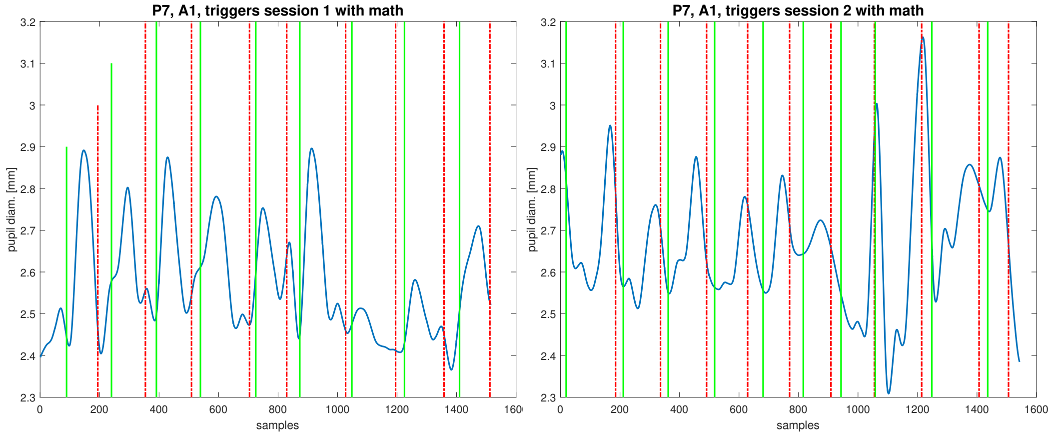
One can observe that an inverted v-shape, most likely related to an increase in mental workload as mentioned in section 2.3.3, is located relatively centered between the triggers. One can also identify other pupillary changes in between the triggers. These are though not that visible, since there were very little time in between the questions.



(a) Sub-task 1A with math during session 1 for test person P2. One can for this participant observe a distinct inverted v-shape between almost all green and red lines

(b) Sub-task 1A with math during session 2 for test person P2. Once again, one can observe distinct inverted v-shapes between most of the green and red lines

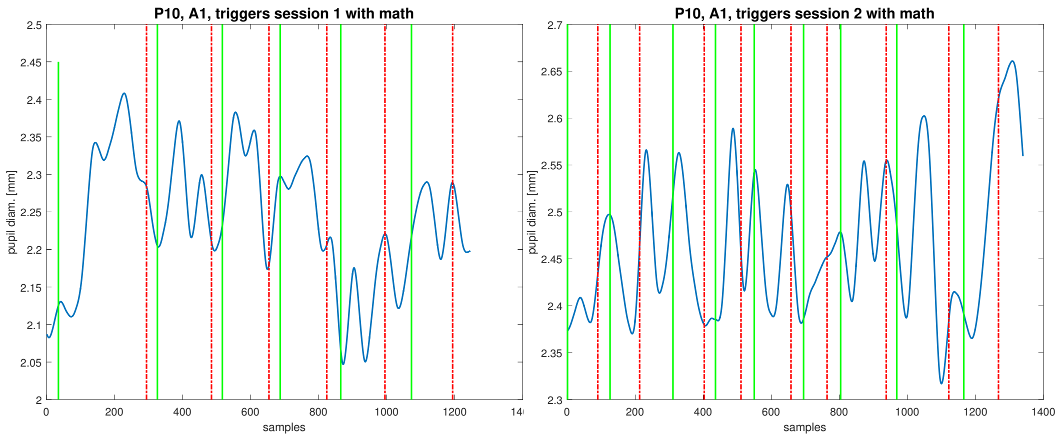
Figure 5.3: Images showing the filtered pupil diameter signal (blue line) during math-part for test person P2, plotted together with triggers for each start (green line) and end (red dashed line) of the math questions.



(a) Sub-task 1A, math-part, during session 1 for test person P7. The inverted v-shapes were for this session very distinct between the vertical trigger lines

(b) Sub-task 1A, math-part, during session 2 for test person P7. For session 2, distinct inverted v-shapes were centered in between most the vertical trigger lines

Figure 5.4: Images showing the filtered pupil diameter signal (blue line) during math-part for test person P7, plotted together with triggers for each start (green line) and end (red dashed line) of the math questions.

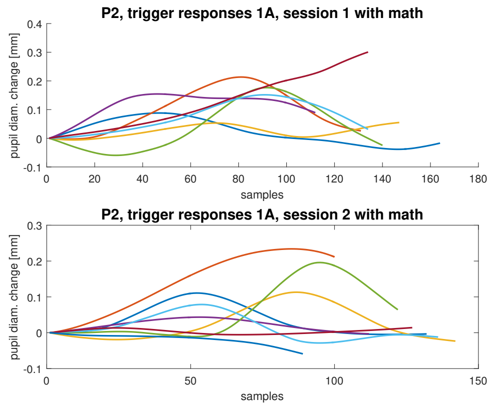


(a) Sub-task 1A, math-part, during session 1 for test person P10. Between the vertical trigger lines are general inverted v-shapes, sometimes including more than one peak.

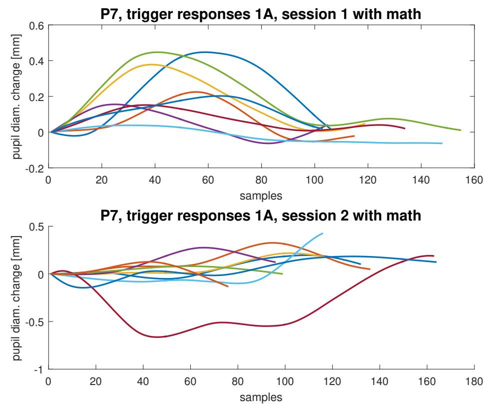
(b) Sub-task 1A, math-part, during session 2 for test person P10. During this session were in general only one peak present in inverted v-shapes, which were centered in between most of the trigger lines.

Figure 5.5: Images showing the filtered pupil diameter signal (blue line) during math-part for test person P10, plotted together with triggers for each start (green line) and end (red line) of the math questions.

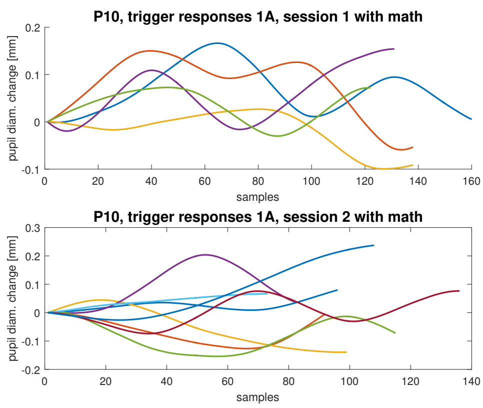
Analysis of the test persons pupillary responses to the math questions, one test person at a time, was also performed. The data points between the start and end of all triggers were then plotted at zero, to compare all responses to each other. These results can be seen for test person P2, P7 and P10 in Fig. 5.6a-5.6c, which display the normalized change in pupil diameter for the math questions, where one colored line represents one math question. Identical plots for the rest of the participants are found i App. B.6. Observe, once again, that both the x-axes and y-axes have different scales to show the variation with better resolution. This is due to that all responses were different in length, but also had different amplitude in pupil diameter change.



(a) Trigger responses for all math questions, P2, where one can observe a difference in pupil behavior for the questions.



(b) Trigger responses for all math questions, P7, where session 1 contain the most similar shape of the pupil responses for the questions.



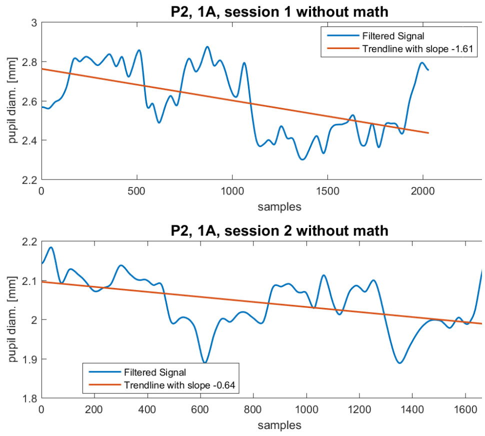
(c) Trigger responses for all math questions, P10, where most of the pupil responses differ from each other.

Figure 5.6: Trigger responses for session 1 and 2, where the coloured lines represent the normalized pupil diameter change related to one math question each, from the time stamp when it is asked until it is answered.

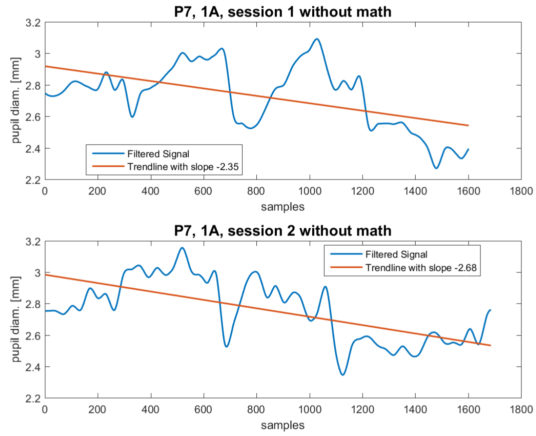
It can be seen from the plots that the trigger responses related to the math questions varies in scale and duration, both within and between individuals.

5.1.3 Trendlines

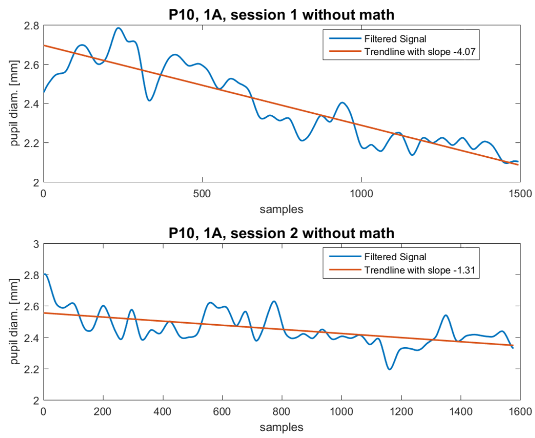
Sequences from the first part of each sub-task, the part without math as described in Fig. 3.6, are presented in Fig. 5.7a-5.7c for participant P2, P7 and P10, both for session 1 and 2. The blue fluctuating line displays the filtered pupil diameter and the red linear line the trendline of the signal. All resulting plots for sub-tasks 1A-1C are displayed in App. B.2- B.4. The axis limits differ between the plots since the individuals have different behaviors. The slope values presented in the figures show how the pupil diameter changes over 10 000 samples in time. If the slope value lies within $\pm 0.5mm$, the pupil diameter is in this project interpreted as stable over time.



(a) Trendline for P2, where one can observe a steeper downward slope for session 1 compared to session 2.



(b) Trendlines for P7, where one can observe a more steep, but almost equal, downward slope for session 2 compared to session 1.



(c) Trendline for P10, where one can observe a steeper downward slope for session 1 compared to session 2

Figure 5.7: Filtered pupil diameter over time (blue line) and its corresponding trendline for P2, P7 and P10 from Experimental set 1, sub-task 1A without math.

One can for most of the test persons observe a steeper downward slope of the trendline from the first session compared to the second one. A downward slope corresponds to a decreasing pupil size in general over time, for the part without math during sub-task 1A. A summarizing table presenting the slope of the trendlines for all participants, both during session 1 and 2 but also for all sub-tasks, are presented in Tab. 5.3. The slopes are steeper downwards for session 1 compared to session 2 for 9/10 participants for 1A, 3/10 for 1B and 9/10 for 1C. For the participants, whose pupil diameter contradicts this and thus have a steeper downward slope for session 2 compared to session 1, are marked as red in the table.

Since the expectation for session 2 was that the motor task should be automatized, one would want to observe trendline slope values closer to zero during session 2. That would according to the project indicate a more stable pupil behavior. For sub-task 1A, the trendline slope value is closer to zero during session 2 for 7/10 participants compared to session 1. For sub-task 1B, this ratio is 3/10, and 8/10 for 1C.

Table 5.3: Presentation of the trendline slopes from Experimental set 1, session 1 and 2 for all sub-tasks. The red colored numbers correspond to a task with steeper downward slope for session 2 compared to session 1 for that participant.

Participant	**Task 1A**		**Task 1B**		**Task 1C**	
Session:	1	2	1	2	1	2
P2	-1.61	-0.64	0.59	-1.94	-2.53	-0.75
P3	-2.07	-1.29	0.66	-4.00	-4.72	-4.02
P4	-9.19	-0.35	-4.44	0.46	-5.31	0.46
P6	-0.14	1.57	-0.14	1.28	-0.77	1.00
P7	-2.35	-2.68	-3.54	-3.83	-3.70	0.41
P8	-1.96	-0.69	-1.23	-1.59	-1.79	0.28
P9	-0.49	2.33	-0.13	-0.55	-1.34	-0.45
P10	-4.07	-1.31	-4.3	-1.43	-3.87	-0.88
P12	-1.61	-0.76	-1.72	-1.94	-1.69	-1.77
P13	-1.55	0.59	-0.18	-0.55	-1.99	-0.97

Fig. 5.8 illustrates the trendline slope values from Tab. 5.3 for sub-task 1A, 1B and 1C for session 1 and 2. A rising line represents a larger value of the trendline slope for session 2 compared to session 1, whereas a falling line represents a smaller slope value for session 2 than for session 1. The upper plot in Fig. 5.8 corresponds to sub-task 1A and one can observe that the slope of the trendlines (i.e the change in pupil size) are smaller during session 2 and thus a more constant pupil diameter over time. The middle plot represents sub-task 1B and it can be observed that the slopes did not follow any unitary pattern as for sub-task 1A. The behaviour for sub-task 1C is similar to 1A for 9/10 slopes.

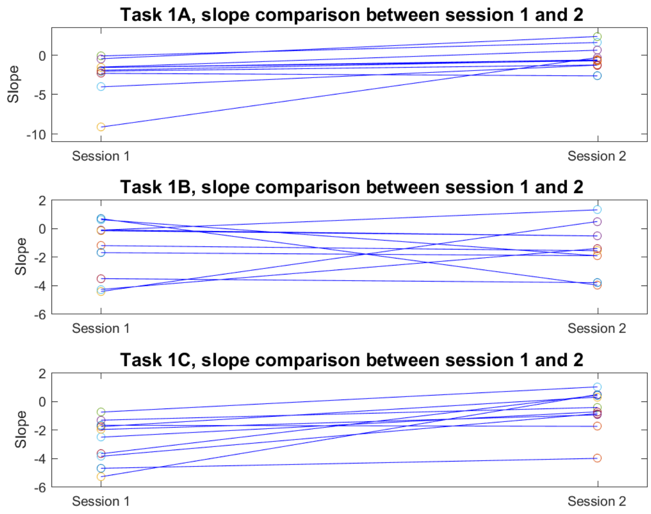


Figure 5.8: Calculated trendline values from session 1 compared to session 2, for sub-task 1A, 1B and 1C. A rising line in this figure represents a larger value of the slope for session 2 than for session 1, and thereof a pupil diameter that decreased less over session 2 than session 1. A falling line represents a smaller slope value for session 2 than for session 1, and thereof a pupil diameter that decreased more during session 2 than session 1.

A paired-samples t-test was conducted in order to compare the trendline slope values of the pupil diameters for session 1 and session 2 conditions. The t-test was performed for all three sub-tasks, meaning 1A, 1B and 1C.

For sub-task 1A, the results showed that there is a significant difference in the scores for session 1 (M= $-2.5\cdot10^{-4}$, SD=$2.58\cdot10^{-4}$) and session 2 (M=$-0.3\cdot10^{-4}$, SD=$1.47\cdot10^{-4}$) conditions; t(9)= -2.73, p=.023. These results suggest that there was a significant difference of the slope values between the two sessions for sub-task 1A.

For sub-task 1B, the results showed that there is no significant difference in the scores for session 1 (M=$-1.4\cdot10^{-4}$, SD=$1.98\cdot10^{-4}$) and session 2 (M=$-1.4\cdot10^{-4}$, SD=$1.68\cdot10^{-4}$) conditions; t(9)=-0.04, p=.967. These results suggest that there was not a significant difference of the slope values between the two sessions for sub-task 1B

For sub-task 1C, the results showed that there is a significant difference in the scores for session 1 (M=$-2.8\cdot10^{-4}$, SD=$1.53\cdot10^{-4}$) and session 2 (M=$-0.7\cdot10^{-4}$, SD=$1.44\cdot10^{-4}$) conditions; t(9)=-3.78, p = .004. These results suggest that there was a significant difference of the slope values between the two sessions for sub-task 1C.

5.1.4 Frequency analysis

It was observed that there are more rapid fluctuations in the data with math, when the participants are more cognitively loaded, both session 1 and 2. This normalized data is presented in Fig. 5.9 for sub-task 1A, where all participants' data are aligned after each other in a resulting data set. Because of this finding, the characteristics of the signals were studied in the frequency domain.

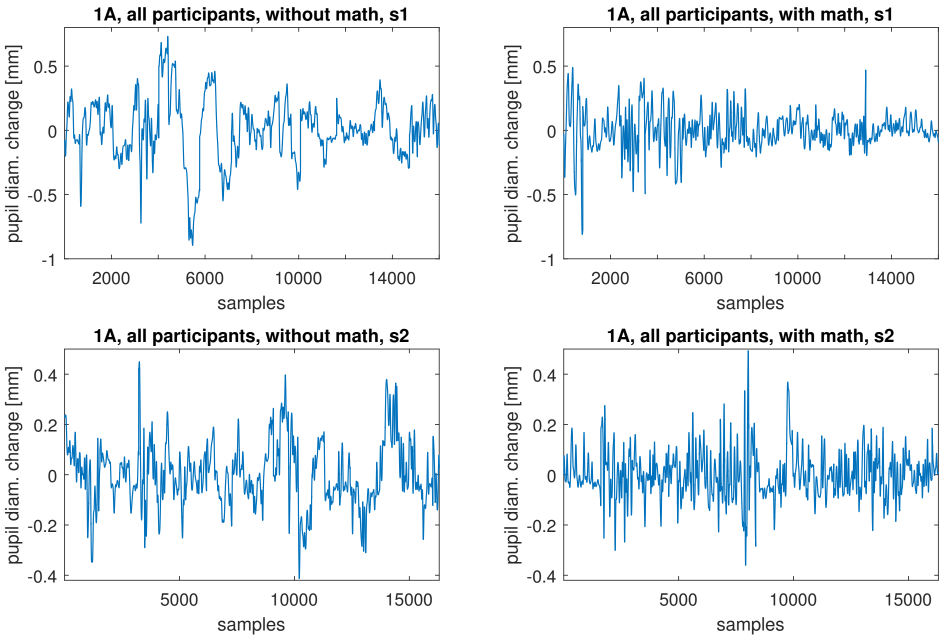


Figure 5.9: Normalized pupil diameter for sub-task 1A, including subsequently stacked data from all participants. The figure shows data from both session 1 (upper part of image) and 2 (lower part of image), where the parts without math are displayed to the left, and math-parts to the right. More rapid fluctuations are seen during math-part.

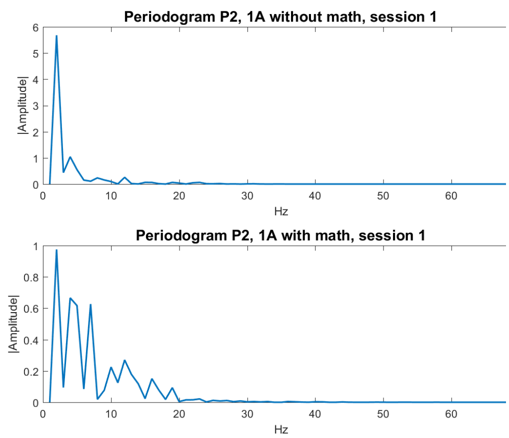
Periodograms were generated from the normalized data in order to analyze the frequency content, see Fig. 5.10-5.12. The periodograms for remaining participants are found in App. B.7. In this analysis, the x- and y-axes differ in the plots, to show fluctuations with better resolution. This was due to very different amplitude values, but also length of data sequences. One can observe that the frequency content is in general distributed over a larger frequency span for the math-part. A peak around 10Hz, for most of the participants, is also observable during math-part. It was also noticed that the first peak, containing low frequency content, in the periodograms generally had a larger amplitude, showing the power spectral density estimate, for the sequences without math, compared to the math-parts.

In order to find characteristics in the data and to actually define the distribution of the frequency content, it was divided in two components: low and high frequency. To distinguish between low frequency content (LFC) and high frequency content (HFC), the first peak and its corresponding frequency values on the x-axis up to the first local minimum point, was defined as LFC.

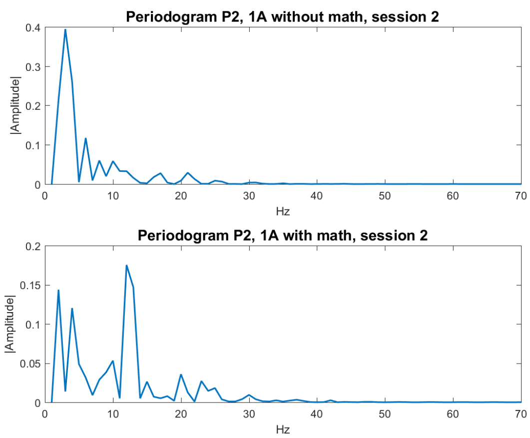
The amplitude of the first peak, for sub-task 1A and each participant, is presented in Tab. 5.4. It is observed that the sequence with math has a lower peak for 9/10 participants for session 1, and 7/10 for session 2.

Table 5.4: Amplitude of the first peak, showing the power spectral density estimate, from the periodograms for the participants during session 1 and 2. Numbers in red correspond to a larger amplitude value of the first peak during math-part compared to when no math questions are asked.

Participant	session 1		session 2	
	No math	Math	No math	Math
P2	5.68	0.98	0.39	0.14
P3	0.70	1.54	0.31	0.37
P4	71.87	1.10	0.42	0.07
P6	1.40	1.30	1.39	0.20
P7	4.42	0.63	8.25	0.47
P8	2.93	0.46	0.72	1.51
P9	0.19	0.61	7.05	0.13
P10	9.44	0.46	1.04	0.08
P12	2.87	0.53	0.73	0.19
P13	2.00	0.62	0.47	0.15

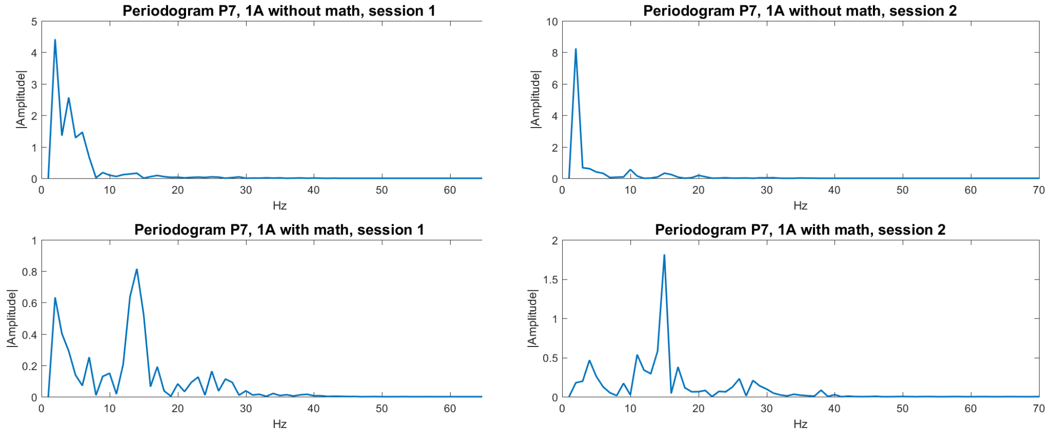


(a) Periodogram for P2, session 1.



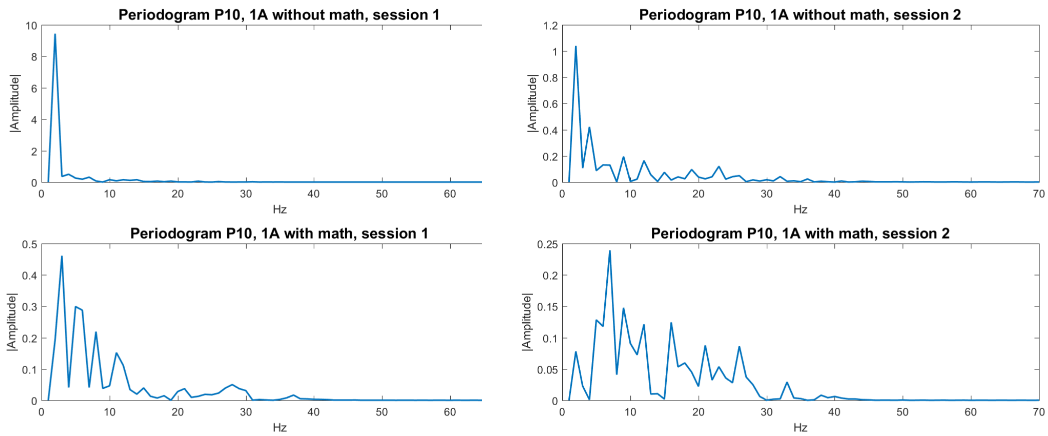
(b) Periodogram for P2, session 2.

Figure 5.10: Periodogram plots for test person P2, where the upper plot displays the frequency content for the sub-task without math and the lower one with math.



(a) Periodogram for P7, session 1. **(b)** Periodogram for P7, session 2.

Figure 5.11: Periodogram plots for test person P2, where the upper plot displays the frequency content for the sub-task without math and the lower one with math.



(a) Periodogram for P10, session 1. **(b)** Periodogram for P10, session 2.

Figure 5.12: Periodogram plots for test person P2, where the upper plot displays the frequency content for the sub-task without math and the lower one with math.

The high frequency content (HFC) in the periodogram was studied, and defined as:

$$\text{HFC} = \frac{\text{energy - first peak}}{\text{energy}} \tag{5.2}$$

where first peak is defined as the area under the first peak and up to the first local minimum and energy being the area under the whole signal. It can be observed that there is more high frequency content for math-part for 8/10 participants, both for session 1 and 2 compared with the sequence without math.

Table 5.5: A comparison of high frequency content (HFC) between sub-task 1A, with- and without math for the participants from session 1 and 2. There are in general more HFC in the sub-task with math compared to the sub-task without math.

	session 1		session 2	
Participant	**HFC no math**	**HFC math**	**HFC No math**	**HFC math**
P2	33.6 %	76.7%	36.0 %	87.4 %
P3	89.1 %	70.9 %	94.3 %	84.4 %
P4	8.6 %	84.2 %	64.7 %	90.6 %
P6	51.8 %	52.7 %	27.0 %	83.1 %
P7	56.3 %	72.1 %	17.3 %	81.7 %
P8	23.7 %	73.6 %	30.7 %	55.6 %
P9	76.8 %	69.1 %	18.3 %	90.6 %
P10	19.7 %	70.9 %	64.1 %	94.5 %
P12	35.6 %	51.8 %	72.0 %	59.8 %
P13	39.4 %	60.3 %	55.9 %	80.3 %

5.2 Initial Conclusions

The results from Experimental set 1 were not possible to analyze to the expected extent regarding conclusions about the motor task being possible to automatize, drawn from Pilot 1 results. The test persons did not spend enough time on rehearsing between the sessions, for the task to be experienced as more automatized during session 2, and the interpretation of the learning factor is thus vague. Therefore, it was decided that the third experiment, Pilot 2, was to be performed. Except for the absence of the learning factor, several other findings were concluded from the testing, as follows:

- Measuring the exact illuminance is not necessary as long as all participants sit in the same room (with no windows) during the sessions.
- There are specific behaviours that are observed for most of the participants during Experimental set 1. For example, the slope values of the trendlines are for session 2, in general, closer to zero than for session 1. The trendline slopes are in this project observed as learning rates, and that the motor task is more automatized during session 2. Since a trendline with slope value closer to zero evokes less pupil dilation over time, the task could be thought of as less cognitively controlled. In Pilot 2, the learning factor will be of greater importance and evaluated further.
- The participants for Pilot 2 will be given more clear instructions about the importance with rehearsing between the sessions. They will be requested to rehearse once a day.
- The t-tests that were performed on the slope values resulted in a p-value for sub-task 1B > 0.05. This was interpreted as that the slope values for sub-task 1B from session 1 and 2 were not significantly different, and that the task is too complex to automatize in this project with a week of training. The subjective self assessment rankings indicated the same results.
- The results from the trigger responses indicated that the pupil dilates when the participants are exposed to a math question. The dilations varies within individuals, but also among them. The difference among individuals will be further analyzed during Pilot 2.
- One can also observe dilations between the math questions. These might arise from factors not possible to measure/identify within the project, such as stress, fatigue and mind wandering. By reducing the number of math questions and instead having longer time intervals in between each question, these behaviours will be investigated further.
- The frequency analysis shows that the frequency content is distributed wider for the sessions with math, and the amplitude of the first peak for the sessions without math is in general higher. The high frequency ratio emphasized the same results, stating that for 8/10 test persons during both sessions, the energy content included more high frequency components for the part with math. This indicates that the difference in energy content might be a solution to distinguish between different mental workload levels.
- The math questions should be recorded on beforehand such that all partic-

ipants will be exposed to the exact same amount of distraction at the same time stamp. This will prevent the participants from generating different psychological responses due to different emphasizes of the questions by the test leader. The exact same timing of the questions will also facilitate the analyzing process.

Results from Experimental set 1 gave rise to new hypotheses which should be tested in Pilot 2. The hypotheses that arose from the results and conclusions were:

- **Hypothesis 1:** Task 1B might be too hard to automatize.

- **Hypothesis 2:** There is a peak around 10-17 Hz for most of the test persons, which appears during the math-parts. This peak might arise from an increased mental workload.

- **Hypothesis 3:** The pupil characteristics among the test persons might deviate less with a narrower age span and by recruiting participants who often use mental arithmetic in their daily life.

- **Hypothesis 4:** A more controlled selection of test persons, where they all will be loaded with the the same math questions at the exact same time stamp, will result in more similar pupillary responses.

- **Hypothesis 5:** Greater differences in the pupil behavior between session 1 and session 2 will be observable, if the participants will rehearse to a greater extent.

6| Results and Initial Conclusions - Pilot 2

The third data collection, Pilot 2, included some improvements from Experimental set 1 but were based on the same analyzing process. Its results are presented in this chapter. The math-parts were analyzed by v-shapes and trigger responses, and the parts without math were analyzed with trendlines. The frequency analysis included both parts, and the last part of the result chapter presents a wavelet analysis which was performed on one part with math and one without, for all participants. This is followed the conclusions from the results found in Pilot 2.

6.1 Results

Pilot 2 had, as stated in section 3.2, almost identical experimental set-up as Experimental set 1. It consisted of four participants with a gender distribution of three males and one female, with an average age of 24. Their test design and time of rehearsal are presented in Tab. 6.1. All participants had good quality (ratio above 55%, defined in section 5.1) on their data, seen in Tab. 6.2, and were thus included in the analysis. Because of limited amount of time, the performance measurements, measuring which keys that were pressed and the time interval in between two pressings, were not analyzed.

Table 6.1: Description of test persons participating and their test designs for Pilot 2. The number of minutes they spent on rehearsing between the sessions are also presented.

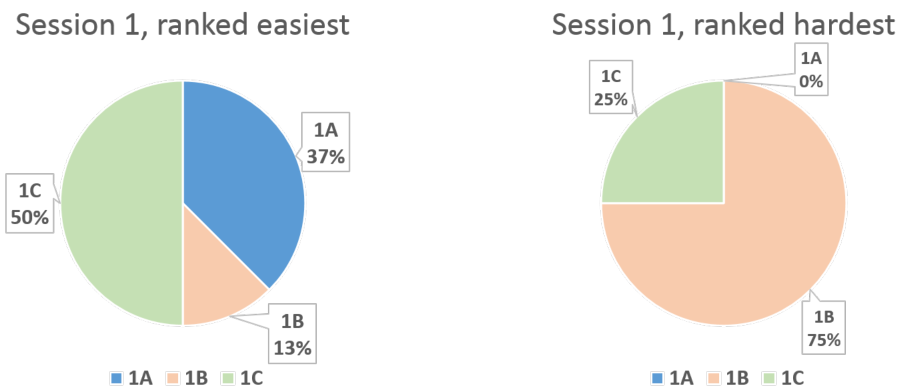
Test object	Test design	Rehearsal
P1	1	5 minutes, once a day
P2	2	2 minutes, once a day
P3	3	5 minutes, once a day
P4	4	5 minutes, once a day

Table 6.2: Data samples for Pilot 2 for raw data and filtered data. The ratio, calculated as the filtered data points divided by the unfiltered data points, indicates how good the eye tracking equipment performed for each test person.

	Session 1			Session 2		
Test person	Unfiltered	Filtered	Ratio	Unfiltered	Filtered	Ratio
P1	26 461	24 806	94 %	26 821	25 847	96 %
P2	27 421	25 090	92 %	25 081	23 884	95 %
P3	28 048	24 274	87 %	25 741	23 450	91 %
P4	27 299	24 547	90 %	25 981	22 566	87 %

6.1.1 Subjective Self Assessment Ratings

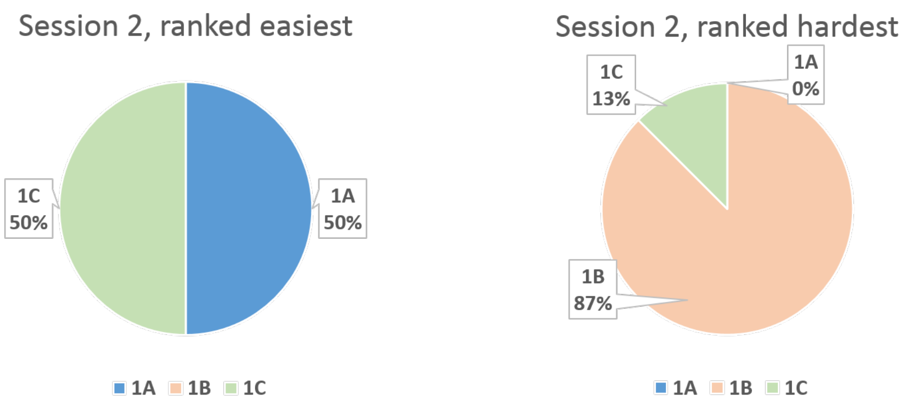
The subjective ratings for Pilot 2 were interpreted identically as in Experimental set 1, generated by the same questionnaire found in App. A.10. All verbally responded subjective rankings, the perceived mental workload from each participant, are presented in App. C.1. Simplified and summarized results for sub-tasks 1A, 1B and 1C, are presented in Fig. 6.1 for session 1 and Fig. 6.2 for session 2. The percentages display how many out of the four test persons that ranked each task as hardest or easiest.



(a) Rankings of the easiest task during session 1, where 1C was ranked as the easiest

(b) Rankings of the hardest task during session 1, where 1B was ranked as the hardest

Figure 6.1: Summary of the subjective self assessment rankings for Pilot 2 session 1, including ratings of tasks 1A, 1B and 1C. Fig. 6.1a illustrates how many out of the four participants that ranked the tasks as the easiest, and Fig. 6.1b the hardest.



(a) Rankings of the easiest task during session 2, where 1A and 1C were ranked as equally easy

(b) Rankings of the hardest task during session 2, where 1B was ranked as the hardest

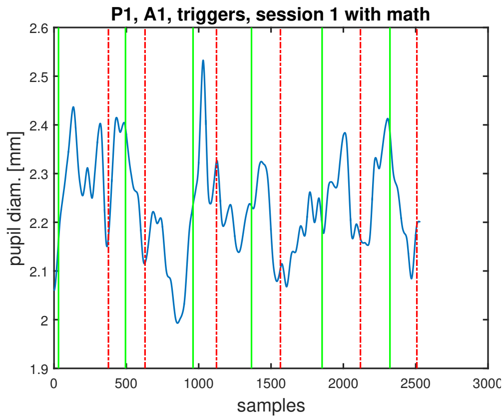
Figure 6.2: Summary of the subjective self assessment rankings for Pilot 2 session 2, including tasks 1A, 1B and 1C. Fig. 6.2a illustrates how many out of the four participants that ranked each as the easiest, and Fig. 6.2b as hardest.

The results indicated that task 1C was ranked as the easiest during the first session, with 50 % of the rankings. During session 2, 1C and 1A were ranked as equally easy with 50 % each. Task 1B was ranked as the hardest for both sessions, 75 % and 87% respectively, which were in line with previous ranking results.

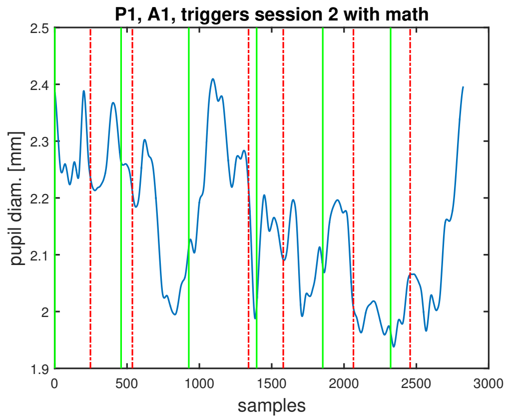
6.1.2 V-shapes and Triggers in Time

The pupil behaviour related to the sequence with math question during sub-task 1A was studied identically for Pilot 2 as for Experimental set 1, with v-shapes and triggers in time. Triggers for when the questions during math-part of sub-task 1A were initiated, and when they were responded to, were plotted together with the pupil diameter. The results are for all four participants presented in Fig. 6.3a-6.6b. The blue line corresponds to the pupil diameter change, the vertical green solid lines represent the start of the math questions and the vertical red dotted lines when the verbal reply to the questions was finalized. Observe that the y-axes of the trigger plots differ, due to that all test participants have different pupil size and behaviour. By having the y-axis optimized for each participant, the resolution of the fluctuations became more clear.

One can from the figures observe that, once again, an inverted v-shape is located relatively centered between the triggers, but not as distinct as for Experimental set 1. It is also noticeable that variations and fluctuations in pupil diameter do occur in between the math questions, when the test persons only perform the motor task, similar to Experimental set 1. These are however more visible during the results from Pilot 2, where the time in between two questions were longer. For test person P2, these variation in between the math questions are very present, with pupil diameter exceeding the parts with questions.

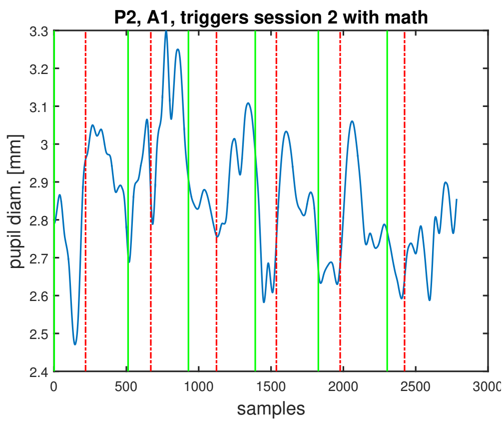


(a) Triggers for session 1. One can observe inverted v-shapes between most of the green and red triggers, but also an increasing pupil diameter prior to all start triggers, i.e. the green lines.

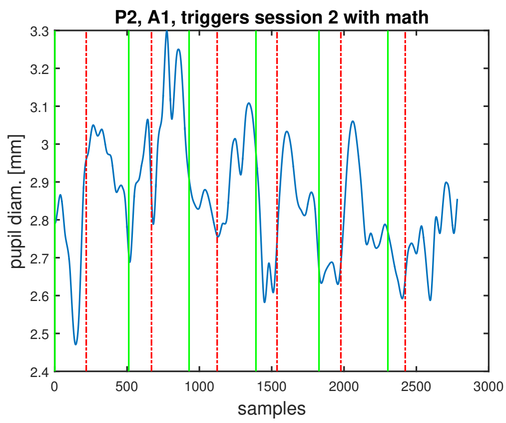


(b) Triggers for session 2. It is once again observable that inverted v-shapes are relatively centered between the triggers, with peaks/increases of diameter between the questions

Figure 6.3: Images showing the filtered pupil diameter signal (blue line) during math-part for test person P1, plotted together with triggers for each start (green line) and end (red dotted line) of the math questions.

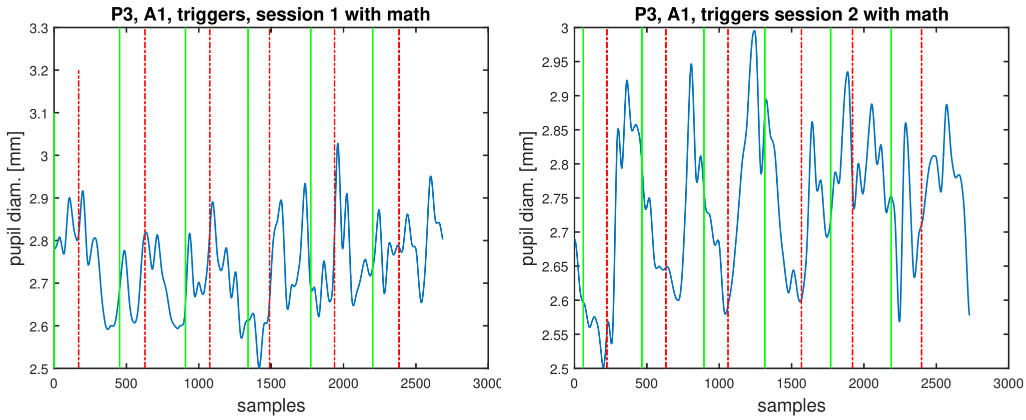


(a) Triggers for session 1. For this test person, the inverted v-shapes between the triggers are relatively centered and with small diameter, when comparing to the behaviour in between the math questions.



(b) Triggers for session 2. As for session 1, this test person has a lot of activity in between the math questions, with high diameter values. Inverted v-shapes are located between the triggers, but with small diameter values in comparison.

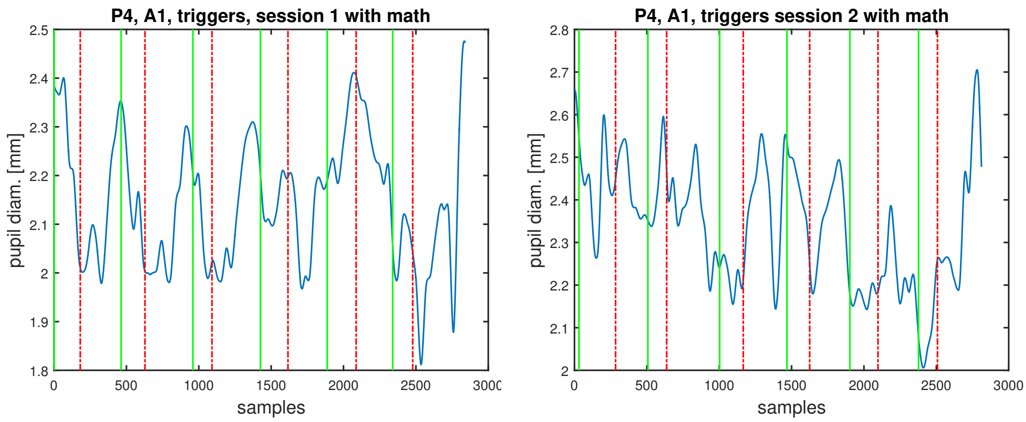
Figure 6.4: Images showing the filtered pupil diameter signal (blue line) during math-part for test person P2, plotted together with triggers for each start (green line) and end (red line) of the math questions.



(a) Triggers for session 1. One can observe inverted v-shapes located mostly between the triggers, but even more fluctuations in between the questions.

(b) Triggers for session 2. For this session, v-formed shapes with greater diameter values are located in between the math questions. There is also inverted v-shapes between most of the triggers.

Figure 6.5: Images showing the filtered pupil diameter signal (blue line) during math-part for test person P3, plotted together with triggers for each start (green line) and end (red dashed line) of the math questions.

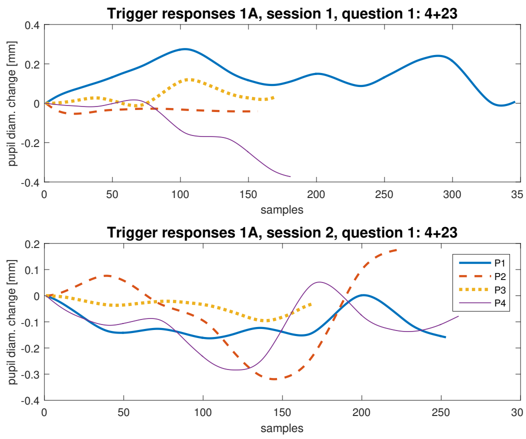


(a) Triggers for session 1. The inverted v-shapes are for this session not as distinct, with many pupil dilation's reoccurring in between the questions.

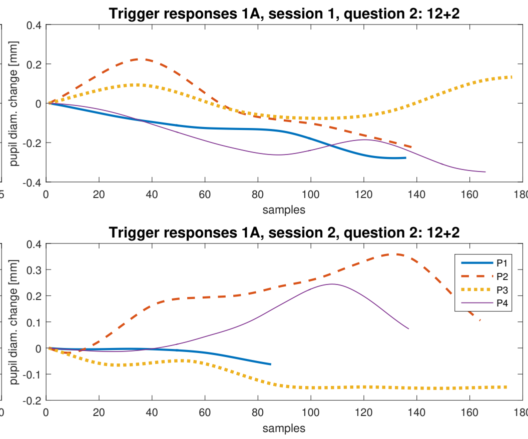
(b) Triggers for session 1. Once again, pupil diameter responses between the triggers are not very distinct, with responses even going in a non-inverted v-shape.

Figure 6.6: Images showing the filtered pupil diameter signal (blue line) during math-part for test person P4, plotted together with triggers for each start (green line) and end (red line) of the math questions.

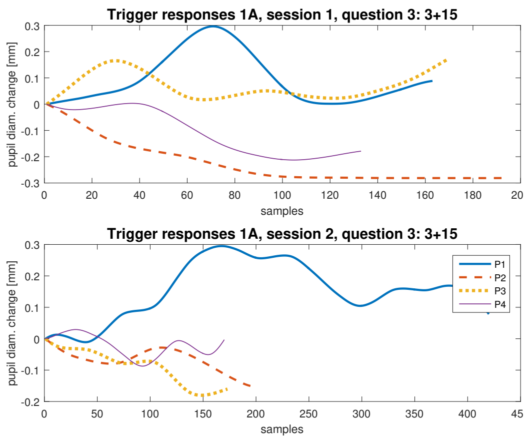
Findings in trigger responses from Experimental set 1 indicated that the pupillary behavior related to an increase in mental workload varies within an individual when loaded with math questions. The trigger responses are now compared among individuals for the same math question, see in Fig. 6.7a-6.9b, by plotting the line for pupil dilation for all participants when answering the same question in one figure. Each line corresponds to the size of the pupil dilation from the time stamp when the test persons are presented with the math question, until it is answered. It can be observed that the pupil dilation for P1-P4 varies between the individuals, even though they are presented with the same math question. One can also register, when comparing session 1 with session 2, that the same participant have distinctly different pupillary behaviour when performing the same mental arithmetic question.



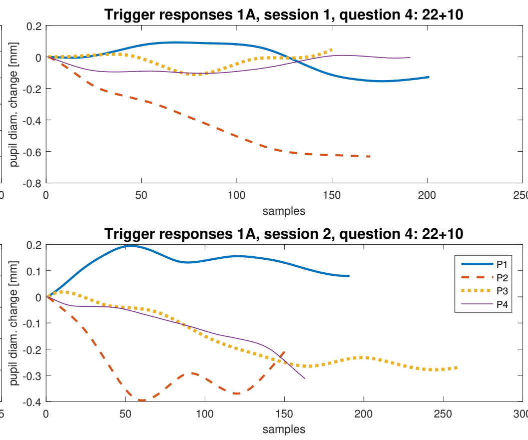
(a) Pupil dilation during sub-task 1A for P1-P4 when hearing and answering the question 23+4



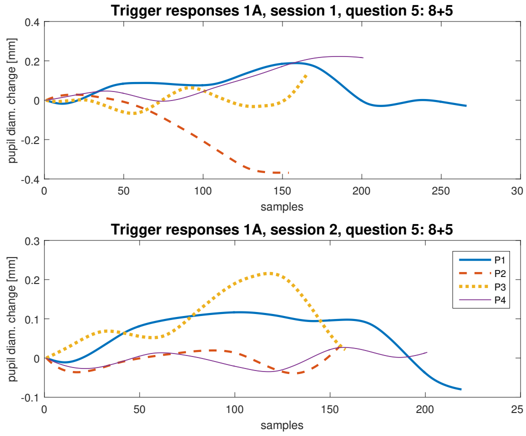
(b) Pupil dilation during sub-task 1A for P1-P4 when hearing and answering the question 12+2



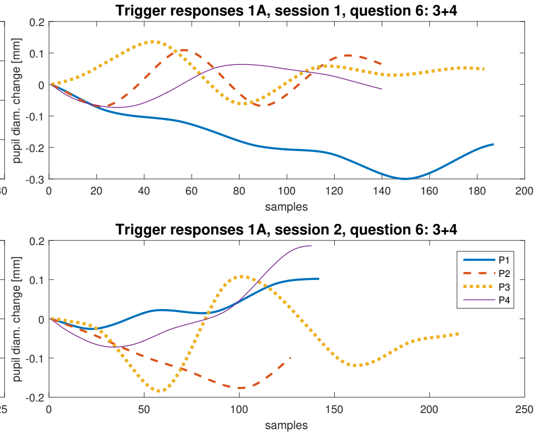
(a) Pupil dilation during sub-task 1A for P1-P4 when hearing and answering the question 3+15



(b) Pupil dilation during sub-task 1A for P1-P4 when hearing and answering the question 22+10



(a) Pupil dilation during sub-task 1A for P1-P4 when hearing and answering the question 8+5



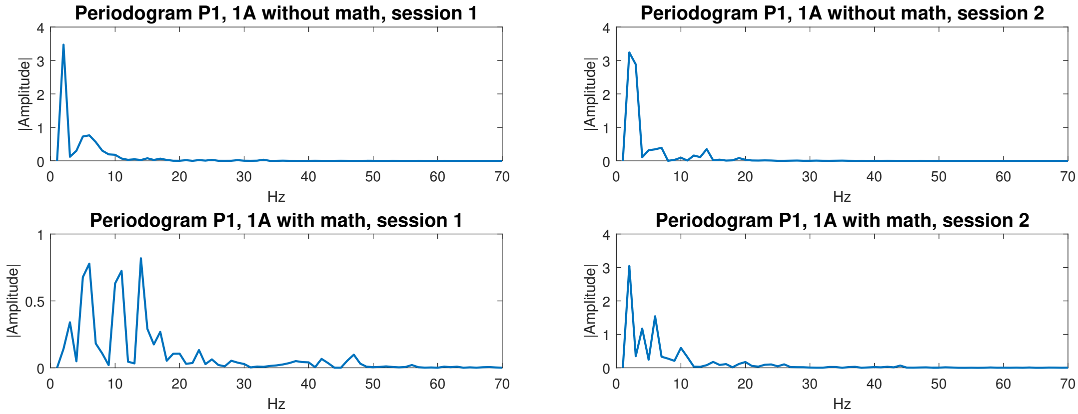
(b) Pupil dilation during sub-task 1A for P1-P4 when hearing and answering the question 3+4

6.1.3 Trendlines

Similar to Experimental set 1, the filtered data and its corresponding trendlines were analyzed in Pilot 2, with the expectation to find a more clear pattern since the participants now rehearsed to a greater extent. However, the results showed to be very similar to Experimental set 1. There were not enough number of participants to generalize the results, and no clear patterns were observed. Because of this, no relevant ways to interpret the data was found, and the trendline plots are thus only attached in App C.2- C.4.

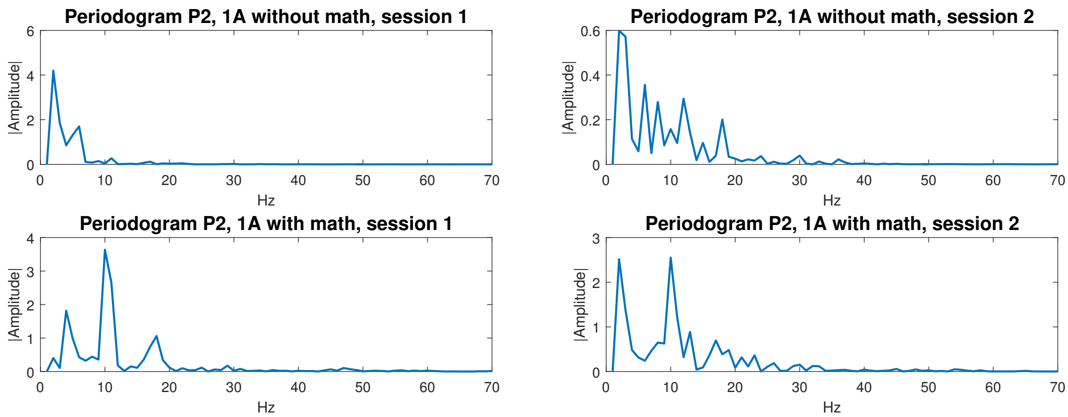
6.1.4 Frequency analysis

The frequency content for all the participants were studied in the same way as in Experimental set 1, where periodograms for both sessions during sub-task 1A were generated from normalized pupillary data, see Eq. 5.1. The periodograms can for all four participants be seen in Fig. 6.10 - 6.13a. From the periodograms for each participant and session, one can observe more HFC when the participants are exposed to math questions. A peak around 10 Hz is present for all of the participants during the math-part. There seems to be a larger frequency span for some of the participants during the math-part as well.



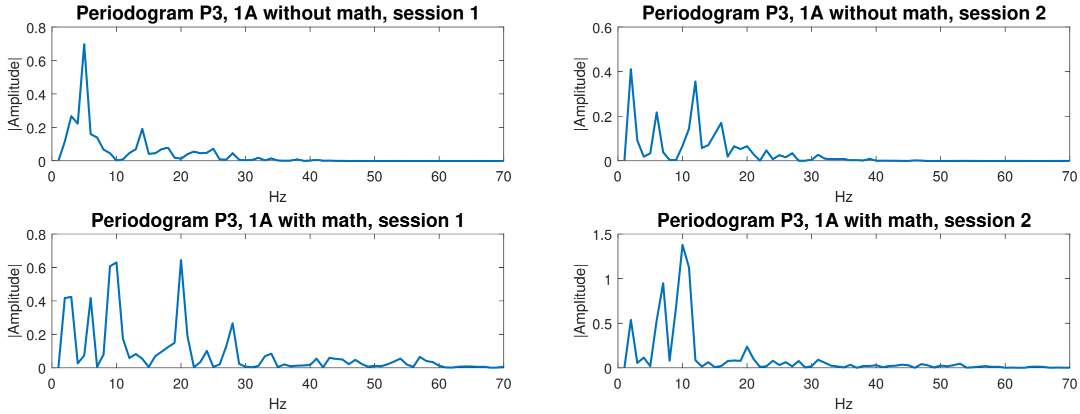
(a) Periodogram for P1, session 1. **(b)** Periodogram for P1 session 2.

Figure 6.10: Periodogram plots for Pilot 2, test person P1 during both sessions, where the upper plot displays the frequency content for the sub-task part without math and the lower one with math. More high frequency content and a peak around 10 Hz is observable during both sessions for the math-part



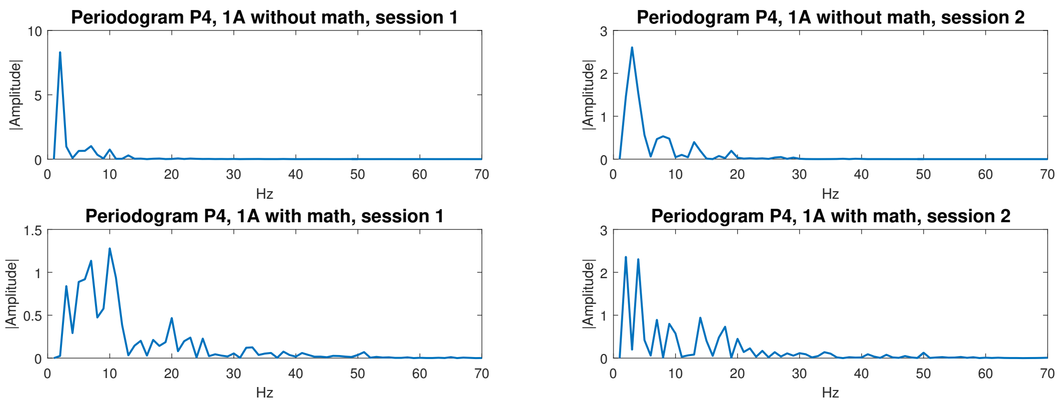
(a) Periodogram for P2 session 1. **(b)** Periodogram for P2 session 2.

Figure 6.11: Periodogram plots for Pilot 2, test person P2 during both sessions, where the upper plot displays the frequency content for the sub-task part without math and the lower one with math. More high frequency content and a peak around 10 Hz is observable during both sessions for the math-part



(a) Periodogram for P3 session 1. **(b)** Periodogram for P3 session 2.

Figure 6.12: Periodogram plots for Pilot 2, test person P3 during both sessions, where the upper plot displays the frequency content for the sub-task part without math and the lower one with math. More high frequency content and a peak around 10 Hz is observable during both sessions for the math-part



(a) Periodogram for P4 session 1. **(b)** Periodogram for P4 session 2.

Figure 6.13: Periodogram plots for Pilot 2, test person P4 during both sessions, where the upper plot displays the frequency content for the sub-task part without math and the lower one with math. More high frequency content and a peak around 10 Hz is observable during both sessions for the math-part

The amplitude value of the first peak defining the power spectral density estimate, which in this analysis determined as the low frequency content, were analyzed for the participants in Pilot 2 as well. These are presented in Tab. 6.3. A first peak with higher amplitude represents higher LFC. For 3/4 participants during session 1, and 2/4 for session 2, the first peak amplitude is higher when no math is present. For the rest, where the amplitude peak is higher during the math-part, the participants are marked with red.

Table 6.3: Amplitude values of the first peak from the periodograms. A high first peak value corresponds to low frequency content (LFC). The first-peak is lower for P1, P2 and P4 during session 1, and for P1 and P4 during session 2. For the participants and sub-tasks where the first peak amplitude are higher for session 2, the values are marked as red.

	session 1		session 2	
Participant:	**No math**	**Math**	**No math**	**Math**
P1	3.47	0.34	3.24	3.05
P2	4.19	0.40	0.60	2.52
P3	0.27	0.42	0.41	0.53
P4	8.30	0.84	2.61	2.36

The HFC for each participant, defined as in Experimental set 1 according to Eq. 5.2, were studied for Pilot 2 as well, and are presented in Tab. 6.4. By observing the values in the table, it is noticable that there is more HFC for all participants during math-part, both for session 1 and 2.

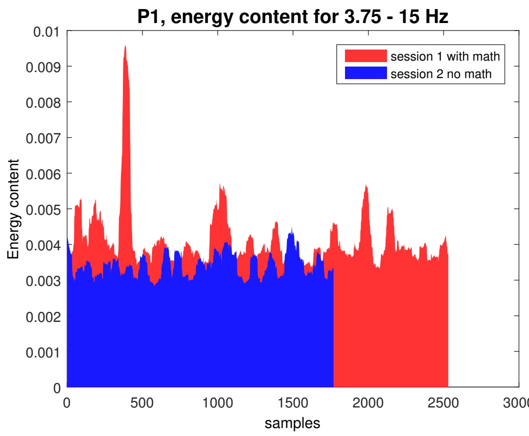
Table 6.4: Comparison of high frequency content (HFC) for all participants, all sub-tasks during session 1 and 2. There is more HFC for all the participants when they are exposed to math questions.

	session 1		session 2	
Participant	**HFC no math**	**HFC math**	**HFC No math**	**HFC math**
P1	50.2 %	92.1 %	25.4 %	65.8 %
P2	38.1 %	96.8 %	61.7 %	69.5 %
P3	77.6 %	85.3 %	77.1 %	91.9 %
P4	31.3 %	89.7 %	31.3 %	80.5 %

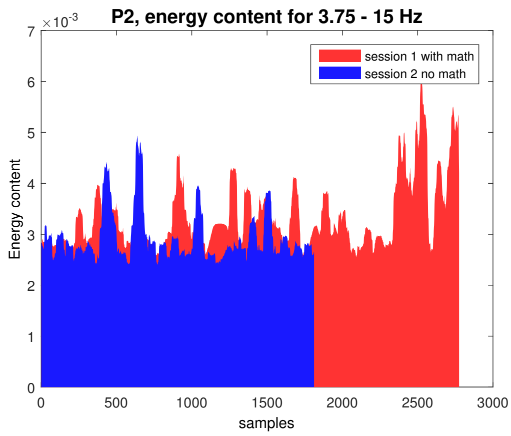
6.1.5 Wavelet analysis

A wavelet analysis gave additional results about the frequency content for all the participants, where the resulting multi-resolutional representations showed frequency content and details from the pupil diameter in different frequency spans. Details from decomposition level d3 and d4 (3.75-15Hz) were added and analyzed further, and are in this additional analysis defined as HFC. Levels corresponding to lower frequencies were defined as LFC, and frequencies above 15Hz as noise or artifacts. The energy content in the HFC span, meaning the RMS enveloped absolute value of the details, were studied and are illustrated in Fig. 6.14a-6.15b. Energy content from two different parts of the sessions were compared, sub-task 1A from session 1 with math and sub-task 1A from session 2 without math. These represented the hypothetically hardest and easiest sessions, concluded from the subjective self assessment and previous results. The sequence with math lasted for approximately 18 seconds longer compared to with no math, which explains why the blue region is shorter than the red one in the figures. It is also noticeable that the y-axes have different scales, to show the differences with as high resolution as possible and optimize according to each participants energy content.

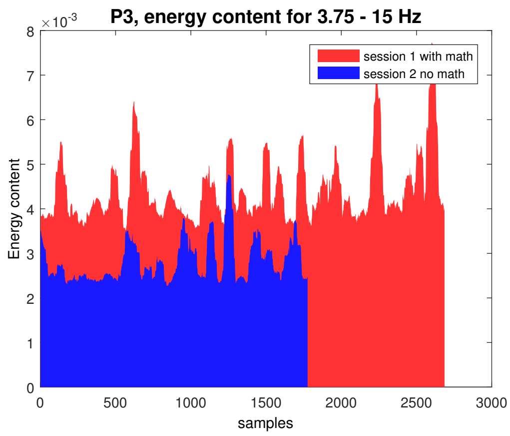
It can be observed that there is more frequency content within the frequency span of 3.75-15Hz for all participants during the first session with math, compared to session 2 without math. This difference is most visible for test persons P3 and P4. It is visible as the amplitude height difference between the red region, which corresponds to math-part during session 1, compared to the amplitude level of the blue region, corresponding to part without math during session 2. The amplitude of the session with math is higher for all participants. The result for session 2 without math does however contain spikes, despite that the participants only are performing a motor task without being disturbed with math questions.



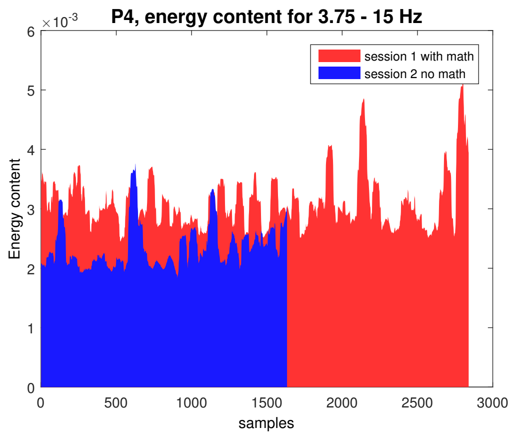
(a) P1 has more energy content within the 3.75-15Hz span during the sequence with math questions (red area).



(b) P2 has more energy content within the 3.75-15Hz span during the sequence with math questions (red area).



(a) P3 has more energy content within the 3.75-15Hz span during the sequence with math questions. (red area)



(b) P4 has more energy content within the 3.75-15Hz span during the sequence with math questions. (red area)

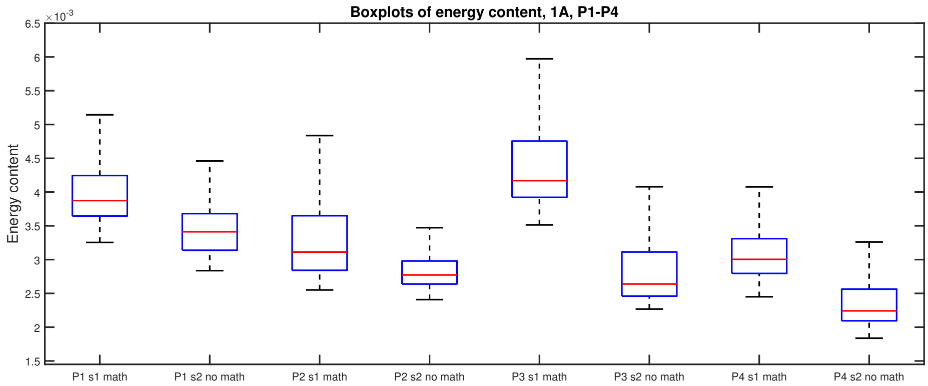


Figure 6.16: Energy content for P1-P4 for session 1 with math and session 2 without math. The central mark in each box is the median and the edges correspond to the 25^{th} and 75^{th} percentiles.

The energy content, within the span of 3.75-15Hz from the wavelet analysis, were for all participants also analyzed statistically using boxplots. These are all together presented in Fig. 6.16, where the energy content for both session 1 and 2 are displayed. The central mark in each box represents the median of the data and the edges correspond to the 25^{th} and 75^{th} percentiles.

Session 1 with math has higher energy content for all four participants, when comparing them individually with session 2 without math. The median of the energy content is higher for all participants during math-part compared to without math, as well as the percentiles span. It can also be observed that the median values among the participants differ a lot when analyzing the same task.

6.2 Initial Conclusions

The results from Pilot 2 fulfill the initial purpose regarding that the test persons should rehearse to a greater extent. All participants rehearsed once a day, which were decided as a reasonable level within this project scope.

It was also perceived that the test persons, in general, did not have trouble with performing the motor task and mental arithmetic at the same time. It is though impossible to determine if this correlated with recruiting a more controlled group of test persons when using only four participants. During the data analysis performed in this project, one cannot observe any significant differences between the data for Experimental set 1 and Pilot 2, as hypothesis 5 purposed. There are some indications in the data, for example the HFC ratio in 6.4, where the ratio with math in general is higher Pilot 2 participants. Also, the energy content amplitude levels, which are visible in Fig. 6.14a-6.15b, do indicate a difference but not significantly strong. For test person P3 and P4, there are more visible differences, but since they only cover 50 % of the participants, more testing with a larger controlled group of test persons has to be performed.

Some general conclusions were though drawn from the findings and results. These are presented below:

- The results from the subjective rankings and trendline slopes indicated, once again, that sub-task 1B is too hard to automatize within the scope and set-up of this thesis. Either a completely different sub-task, more clear instructions or the possibility to perform the sub-task during longer and more frequent occasions would make it possible to automatize the task.
- From the frequency analysis, the peak around 10-17Hz that was observed for most of the participants from Experimental set 1, is now present for all participants, both during session 1 and 2 with math. In comparison to Experimental set 1, the peak was more centered at 10Hz for all test persons during Pilot 2.
- The learning factor from trendline slope values, meaning a slope value closer to zero during session 2 and thus a less deviating pupil, is a vague measure since the difference is small. For some participants it is negligible, meaning below 0.5mm. This indicates that the analysis might not reliable as an indicator of cognitive load.
- The trigger responses now indicate that the pupil dilation varies between the individuals, even though they are exposed to the same recorded math questions. The fluctuations between the math questions are present in Pilot 2 as well, where the pupil diameter sometimes dilates even more than when loaded with a math question.
- The wavelet analysis showed that there are more energy within the high frequency span, 3.75-15Hz, for the harder sub-task. Spikes are however present during session 2 with no math, which should be seen as the easiest session, and one can once again state that the pupil is a sensitive measure.

The first peak values, high frequency content, as well as the wavelet analysis from Pilot 2 do all point in the same direction, i.e there are more HFC when the participants are loaded with math, interpreted as performing a more cognitively demanding task. Despite this, one can also observe pupil behavior that is hard to explain, like fluctuations in the pupil diameter in the trigger responses between math questions, as well as the spikes in energy content in the wavelet analysis during the easier task. Because it is unknown why this pupil behavior origins, it is hard to isolate its appearance. None of the performed analysis in this thesis give any clear results.

7| Conclusion and Discussion

This chapter covers the general conclusions and discussion for the whole project, meaning all data collections and results together. The chapter is then finalized with presenting possible future work.

7.1 General Conclusions and Discussion

After completing all three experiments and concluding the findings from each of them, some general conclusions were able to be drawn. When trying to answer the research questions that the project had objective to find answers to, there are no easy solutions. Regarding the first research question, *What pupillary measurements can be used to measure cognitive activity*, there are according to this project at least two pupillary measures to be considered. The absolute pupil diameter change might show behaviours, like the inverted v-shape due to an increase of mental workload, but also the learning rate and adaption to the task with a trendline over time. To study the fluctuations or "vibrations" in the pupil, meaning the number of changes in pupil diameter per seconds as in the frequency analysis, is also a possible indicator. The results from analyzing these behaviours show that the energy will get distributed over a larger frequency span, with an interesting peak at 10Hz for a significant number of test persons, during more cognitively demanding sequences. Regarding the peak at 10Hz, it was not possible to conclude within this project where it might originate from. The peak was centered around 10Hz for Pilot 2, and had a wider span than for Experimental set 1. One cannot exclude that the more centered peak in Pilot 2 arises because of the test persons narrower (and younger) age span, and thus is age dependent. On the other hand, it is observable that the peak arises when the test persons are more cognitively loaded, but does not necessarily evoke from cognitive load. There might be other factors, such as mental processes and physiological factors, that can be correlated with the cognitively controlled processes seen at 10Hz, and therefore mislead the results. The peak has to be studied further with more controlled experiments, to be able to correlate it to mental workload.

A conclusion is though that the pupillary measures found in the project are not valid or strong enough to be able to indicate the level of cognitive load, neither alone nor together. As far as the project has come, they can only indicate characteristics that might help distinguish between very little loaded and heavy loaded. We cannot linearly or non-linearly correlate the measures we found with the predicted/assumed cognitive load level yet. However, this might be possible with a carefully designed experiment involving more participants. More experimental results are needed in order to map the changes in the observed measures (i.e. pupil diameter change and oscillation frequency characteristics) to the presumed/designed cognitive load of the task.

Since pupillary measures have been found sensitive to a great number of parameters, that have to be isolated and/or under control in experiments to produce reliable results, they do not, from the project's perspective, seem to be a realistic measure to apply in the real world. There are too many parameters that might affect the pupil. First, methods to compensate for these factors have to be created. Even in controlled experiments, for example in simulators, different illuminance conditions will be presented to the test persons and this must be taken into account when analyzing the pupil behavior. The pupillary measure that, according to the observations in this projects, is most likely to be used as an indicator of cognitive load is the frequency and energy content.

An additional finding is that, in contrast to for example fingerprints, pupillary measures are not a consistent identification measurement, both among and within individuals. It was observed that the pupil dilation varied a lot within each individual, even when a test person performed the same task twice. The conclusion is then that when analyzing the pupillary measures, it might be preferable to analyze general behaviours from a large set of data, and also that the context and design of the experiments will be of great importance when performing tests. With that means ensuring that the characteristics in the measures actually display changes in mental workload, and not other factors like illumination changes, stress and accumulation reflexes. The experiments have to be very controlled to ensure that the test persons do not load their working memory more than planned, meaning they respond to cues such that they might display mental images with the visuospatial sketchpad or repeat information in their phonological loop. What has been experienced during the project is that the more factors one isolates, the more one learn that there are additional factors to isolate and the more unnatural the test procedure will become for the participants. It is an ongoing challenge to discover such a task, that also is cognitively demanding at the first try for everyone within a large set of test persons, but then possible to (to some extent) automatize within a short period of time. The underlying thought with the developed experimental design was therefore based on a motor task, which did not include any cues itself and could possibly become learned by the motor memory, within the time of one week. During the math-part sequences, cues were included, with the intention that it would be a clear and visible change in pupil behavior in contrast to the motor task itself.

The designed experimental method did, as mentioned above, contain some factors which were not able to be controlled. An example is the math questions, where the test persons had different levels of prior knowledge and thus prerequisites, which most likely induced different levels of mental workload. A difficulty was also to control that all math questions induced the same level of cognitive demand and thus were processed equally, both among and within individuals. An attempt to improve this factor was done by the corrections in the experimental test design for Pilot 2, where the math questions were recorded beforehand. This ensured that all test persons got the same auditory cue, at the exact same time stamp. Another factor that were not fully controlled was the focus/gaze of the test persons, which might origin accumulation reflexes, meaning change of focus. In the experimental design, this factor was attempted to be controlled with the white cross in the test set-up, which the test persons were asked to focus on during the whole test session. Some of the test persons did however change their focus from the cross, especially when responding to the subjective ratings, but also when getting stuck with a math question or the motor key pressings. These changes of focus were hard to control, but seemed to be improved for Pilot 2. A conclusion was drawn that it might be due to that these participants were in general more used to not looking at the keyboard when pressing the buttons, and also that they are more used to taking in a lot of stimulus at the same time. These factors, the number of errors of the key pressings but also the gaze distribution, would have been interesting to analyze to test the hypothesis. If additional testings should be performed with this method, a reference group should preferably be recruited with pianists to see if their data differs. Pianists will have different prerequisites, since they are used to using both hands simultaneously, and the task will most likely be interpreted in a different way compared to other persons.

Also, when looking at the bigger picture, eye measures such as blinks and fixations might also be interesting indicators. Since in the literature review it was found that the number of blinks increase when the cognitive demand increases, the math-parts did include more data samples categorized as "Blinks" by the equipment, and thus were filtered out. The consequence of that made the number of valid data-points become less for the math-part in contrast to the sequences without math. Therefore, to analyze the blinks itself might be an interesting parameter. In combination with the blinks, the factor of fixation might, as mentioned above be an interesting aspect, maybe in combination with the pupillometry.

Regarding the second research question in the objective, *can pupillary measurements be indicators of whether a task is cognitively demanding or automatically performed*, there is no valid answer based on this study. At first, one can discuss what an automatic task actually is, and to what extent a task has to be cognitively ”non-loading” to be called automatic. The next issue, when testing this hypothesis, is how to identify a task as automatic without having any data from an automatized task to compare with. It would be preferable to collect data for every test person, when one is performing an automatized task, to later compare other pupillary data with. Automatized tasks might from the projects knowledge and findings, not be easy to generate in such a controlled environment that it could be collected with eye tracking in an experimental, and thus rather unnatural, situation. It is also difficult to force test persons to perform something natural, automatized, and also ensuring that they do not think of anything that loads their working memory. Therefore, the pupillary measures analyzed within the project will not be defined as measures that could distinguish between automatized or cognitively controlled tasks, but rather an indication of what different characteristics that the pupil have during these situations.

7.2 Future Work

Because of the time limit of the project, performance rating for Experimental set 1 and Pilot 2 had to be excluded from the analysis. A future work would be to analyze the key pressings and look for individual patterns and deviations. Deviations or errors in key pressings and the corresponding time stamps could be analyzed, together with the size of the pupil diameter, over time to see how these affects the pupil behavior.

Glance behavior from each participant can be investigated and analyzed together with the performance rating and pupil behavior. There could be a correlation between fluctuations in pupil diameter and glancing during the math-part. The glance behavior could also be analyzed in order to see how much each test person actually focused on the white cross during Experimental set 1 and Pilot 2.

Each participants' blink behavior could be analyzed in order to see if there is any connection between blink ratio and more cognitively demanding tasks for the applied method.

Future work could also include finding a fixed frequency where low and high frequency distinction can be drawn, based on physiological evidence. Further investigations of what happens around 10Hz would also be an interesting field, where a peak was observed for most of the participants during the sequences with math questions and not as distinct for sequences without math. According to findings in the literature review, pupillary fluctuations could appear up to these frequencies, which indicate that the behaviour could be realistic pupil vibrations and not noise.

The velocity of the pupil fluctuations, meaning the derivative of the pupil diameter change, could also be a possible direction for future work.

In general, the experiments and data collection must be conducted on more participants in order to achieve a more valid and statistically significant result. As for now, four participants, as in Pilot 2, are too few to draw any general conclusions from.

Bibliography

[1] G. Marquart, C. Cabrall, and J. de Winter, “Review of Eye-related Measures of Drivers’ Mental Workload”, *Procedia Manufacturing*, vol. 3, pp. 2854–2861, Jan. 2015, ISSN: 2351-9789. DOI: 10.1016/J.PROMFG.2015.07.783. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2351978915007842.

[2] T. M. Gelfeld, “Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world”, *International Journal of Psychophysiology*, vol. 93, no. 1, pp. 38–44, Jul. 2014, ISSN: 0167-8760. DOI: 10.1016/J.IJPSYCHO.2013.11.002. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167876013003267.

[3] P. Philip, P. Sagaspe, N. Moore, J. Taillard, A. Charles, C. Guilleminault, and B. Bioulac, “Fatigue, sleep restriction and driving performance”, *Accident Analysis and Prevention*, 2005, ISSN: 00014575. DOI: 10.1016/j.aap.2004.07.007.

[4] J. Engström, G. Markkula, T. Victor, and N. Merat, “Effects of Cognitive Load on Driving Performance: The Cognitive Control Hypothesis”, *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 59, no. 5, pp. 734–764, Aug. 2017, ISSN: 0018-7208. DOI: 10.1177/0018720817690639. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0018720817690639.

[5] Emma Nilsson, *Pupillary measures as indicators of whether a task is cognitively or automatically controlled / Chalmers*, 2017. [Online]. Available: https://www.chalmers.se/en/departments/e2/education/masters-programmes/Pages/Pupillary-measures-as-indicators-of-whether-a-task-is-cognitively-or-automatically-controlled.aspx (visited on Feb. 5, 2018).

[6] E. Nilsson, “Mind off driving: Effects of cognitive load on driver glance behaviour and response times”, PhD thesis, Chalmers University of Technology, 2018. [Online]. Available: https://research.chalmers.se/publication/501276/file/501276%7B%5C_%7DFulltext.pdf.

[7] N. Holt, A. Bremner, E. Sutherland, M. Vliek, M. W. Passer, and R. E. Smith, *Psychology: The Science of Mind and Behaviour*, 3rd. Berkshire: McGraw-Hill Education, 2015, ISBN: 9780077169848.

[8] G. J. Tortora and B. Derrickson, *Introduction to the Human Body*, 10th, B. Roesch, B. Cheetam, and K. Trost, Eds. Hoboken: John Wiley and Sons, Inc., 2015, ISBN: 978-1-118-96353-1.

[9] C. W. Oyster, *The Human eye*. Sunderland, Massachuetts, 1999, pp. 418–419, ISBN: 0-87893-645-9.

[10] R. E. Hampson, I. Opris, and S. A. Deadwyler, "Neural correlates of fast pupil dilation in nonhuman primates: Relation to behavioral performance and cognitive workload", *Behavioural Brain Research*, vol. 212, no. 1, pp. 1–11, 2010, ISSN: 01664328. DOI: 10.1016/j.bbr.2010.03.011.

[11] C. D. Wickens and J. G. Hollands, *Engineering Psychology and Human Performance*, 3rd, N. Robert and B. Webber, Eds. New Jersey: Prentice-Hall Inc., 2000, ISBN: 0-321-04711-7.

[12] E. Smith E. and S. Kosslyn M., *Cognitive psychology: mind and brain*, 1 st. New Jersey: Pearson Education, Inc., 2007.

[13] D. Kahneman, *Thinking, fast and slow*, 1st. New York: Farrar, Straus and Giroux, 2011, ISBN: 978-0-374-53355-7.

[14] W. Schneider and J. M. Chein, "Controlled & automatic processing: behavior, theory, and biological mechanisms", *Cognitive Science*, vol. 27, no. 3, pp. 525–559, May 2003, ISSN: 03640213. DOI: 10.1207/s15516709cog2703_8. [Online]. Available: http://doi.wiley.com/10.1207/s15516709cog2703%7B%5C_%7D8.

[15] E. Querino, L. dos Santos, G. Ginani, E. Nicolau, D. Miranda, M. Romano-Silva, and L. Malloy-Diniz, "Translational Neuroscience Cognitive effort and pupil dilation in Controlled and automatiC proCesses", DOI: 10.1515/tnsci-2015-0017. [Online]. Available: https://www.degruyter.com/downloadpdf/j/tnsci.2015.6.issue-1/tnsci-2015-0017/tnsci-2015-0017.pdf.

[16] H. K. Wong and J. Epps, "Pupillary transient responses to within-task cognitive load variation", *Computer Methods and Programs in Biomedicine*, vol. 137, pp. 47–63, Dec. 2016, ISSN: 01692607. DOI: 10.1016/j.cmpb.2016.08.017. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0169260716300025.

[17] A. Korbach, R. Brünken, and B. Park, "Measurement of cognitive load in multimedia learning: a comparison of different objective measures", *Instructional Science*, 2017, ISSN: 1573-1952. DOI: 10.1007/s11251-017-9413-5.

[18] N. Debue and C. van de Leemput, "What does germane load mean? An empirical contribution to the cognitive load theory.", *Frontiers in psychology*, vol. 5, p. 1099, 2014, ISSN: 1664-1078. DOI: 10.3389/fpsyg.2014.01099. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/25324806%20http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4181236.

[19] B. Park and R. Brünken, "The Rhythm Method: A New Method for Measuring Cognitive Load-An Experimental Dual-Task Study", *Applied Cognitive Psychology*, vol. 29, no. 2, pp. 232–243, Mar. 2015, ISSN: 08884080. DOI: 10.1002/acp.3100. [Online]. Available: http://doi.wiley.com/10.1002/acp.3100.

[20] O. Palinko, A. L. Kun, A. Shyrokov, and P. Heeman, "Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator", 2010. [Online]. Available: http://delivery.acm.org.proxy.lib.chalmers.se/10.1145/1750000/1743701/p141-palinko.pdf?ip=129.16.69.49%7B%5C&%7Did=1743701%7B%5C&%7Dacc=ACTIVE%20SERVICE%7B%5C&%7Dkey=74F7687761D7AE37.3C5D6C4574200C81.4D4702B0C3E38B35.4D4702B0C3E38B35%7B%5C&%7D%7B%5C_%7D%7B%5C_%7Dacm%7B%5C_%7D%7B%5C_%7D=1519587771%7B%5C_%7D0d565deb4eeaad577212ec916ebb7b1.

[21] A. M. Colman, *A Dictionary of Psychology*, 4th. Oxford University Press, 2015, ISBN: 9780199657681. DOI: 10.1093/acref/9780199657681.001.0001.

[22] B. Laeng, M. Ørbo, T. Holmlund, and M. Miozzo, "Pupillary Stroop effects", *Cognitive Processing*, vol. 12, no. 1, pp. 13–21, Feb. 2011, ISSN: 1612-4782. DOI: 10.1007/s10339-010-0370-z. [Online]. Available: http://link.springer.com/10.1007/s10339-010-0370-z.

[23] J. Xu, Y. Wang, F. Chen, and E. Choi, "Pupillary Response Based Cognitive Workload Measurement under Luminance Changes", in, Springer, Berlin, Heidelberg, 2011, pp. 178–185. DOI: 10.1007/978-3-642-23771-3_14. [Online]. Available: http://link.springer.com/10.1007/978-3-642-23771-3%7B%5C_%7D14.

[24] J. Fong, "The Meaning of Pupil Dilation", 2012. [Online]. Available: https://www.the-scientist.com/?articles.view/articleNo/33563/title/The-Meaning-of-Pupil-Dilation/.

[25] A. Szulewski, N. Roth, and D. Howes, "The Use of Task-Evoked Pupillary Response as an Objective Measure of Cognitive Load in Novices and Trained Physicians: A New Tool for the Assessment of Expertise", *Academic Medicine*, 2015, ISSN: 1938808X. DOI: 10.1097/ACM.0000000000000677.

[26] J. W. Krakauer and P. Mazzoni, "Human sensorimotor learning: Adaptation, skill, and beyond", *Current Opinion in Neurobiology*, 2011, ISSN: 09594388. DOI: 10.1016/j.conb.2011.06.012. arXiv: f.

[27] E. L. Abrahamse, M. F. L. Ruitenberg, E. de Kleine, and W. B. Verwey, "Control of automated behavior: insights from the discrete sequence production task", *Frontiers in Human Neuroscience*, vol. 7, p. 82, Mar. 2013, ISSN: 1662-5161. DOI: 10.3389/fnhum.2013.00082. [Online]. Available: http:

//journal.frontiersin.org/article/10.3389/fnhum.2013.00082/abstract.

[28] *N E W N a t u r a l G a z e T M D e s i g n SMI Eye Tracking Glasses 2 Wireless Mobile eye tracking made easy, robust, efficient and versatile SMI Eye Tracking Glasses 2 Wireless 3hrs+ Data Collection*, 2014. [Online]. Available: http://www.mindmetriks.com/uploads/4/4/6/0/44607631/final%7B%5C_%7Dsmi%7B%5C_%7Detg2w%7B%5C_%7Dnaturalgaze.pdf (visited on Feb. 1, 2018).

[29] P. R. Mosaly, L. M. Mazur, and L. B. Marks, "Quantification of baseline pupillary response and task-evoked pupillary response during constant and incremental task load", *Ergonomics*, vol. 60, no. 10, pp. 1369–1375, 2017. DOI: 10.1080/00140139.2017.1288930. [Online]. Available: https://doi.org/10.1080/00140139.2017.1288930.

[30] I. The MathWorks, *Moving Average Filter - MATLAB & Simulink - MathWorks Nordic*. [Online]. Available: https://se.mathworks.com/help/econ/filtering.html (visited on Apr. 27, 2018).

[31] S. Golestan, M. Ramezani, J. M. Guerrero, F. D. Freijedo, and M. Monfared, "Moving Average Filter Based Phase-Locked Loops: Performance Analysis and Design Guidelines", *IEEE Transactions on Power Electronics*, vol. 29, no. 6, pp. 2750–2763, Jun. 2014, ISSN: 0885-8993. DOI: 10.1109/TPEL.2013.2273461. [Online]. Available: http://ieeexplore.ieee.org/document/6560420/.

[32] I. The MathWorks, *Savitzky-Golay filtering - MATLAB sgolayfilt - MathWorks Nordic*. [Online]. Available: https://se.mathworks.com/help/signal/ref/sgolayfilt.html?s%7B%5C_%7Dtid=doc%7B%5C_%7Dta (visited on May 1, 2018).

[33] S. R. Krishnan and C. S. Seelamantula, "On the Selection of Optimum Savitzky-Golay Filters", *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 380–391, Jan. 2013, ISSN: 1053-587X. DOI: 10.1109/TSP.2012.2225055. [Online]. Available: http://ieeexplore.ieee.org/document/6331560/.

[34] A. Graps, *Introduction to wavelets*, 1995. DOI: 10.1109/99.388960. arXiv: 1011.1669v3.

[35] *Periodogram power spectral density estimate - MATLAB periodogram - MathWorks Nordic*. [Online]. Available: https://se.mathworks.com/help/signal/ref/periodogram.html (visited on May 16, 2018).

[36] H. Adeli, Z. Zhou, and N. Dadmehr, "Analysis of EEG records in an epileptic patient using wa v elet transform", vol. 123, pp. 69–87, 2003.

[37] I. Daubechies, "The Wavelet Transform , Time-Frequency Localization and Signal Analysis", no. 5, pp. 961–1005, 1990.

[38] G. Strang, "Wavelet transforms versus fourier transforms", vol. 28, no. 2, 1993.

[39] S. Town, *Free virtual piano: Online Keyboard*. [Online]. Available: `https://www.sessiontown.com/en/music-games-apps/online-virtual-keyboard-piano`.

A| Appendix - Method

A.1 Experimental Set-up for the Project

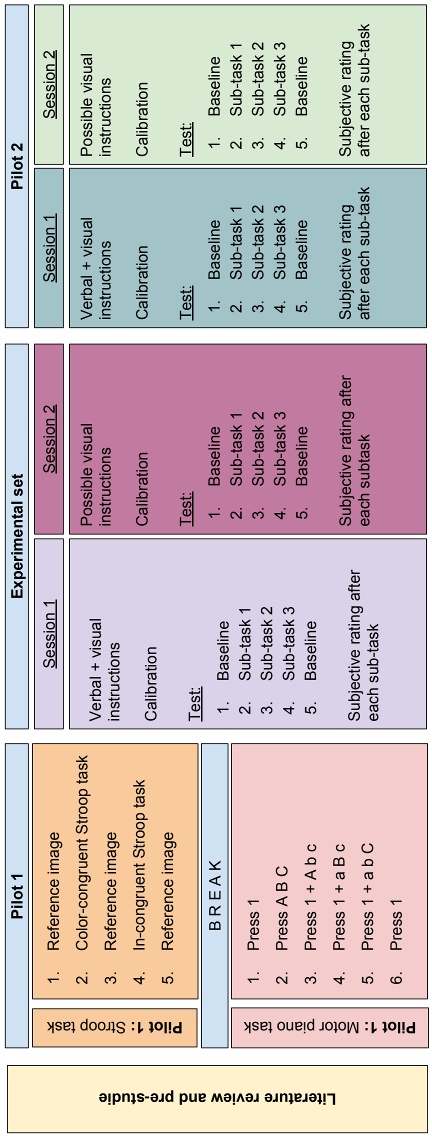


Figure A.1: The experimental set-up for the thesis, including both pre-study, Pilot 1, Experimental setup 1 and Pilot 2.

A.2 Calibration Image - Pilot 1

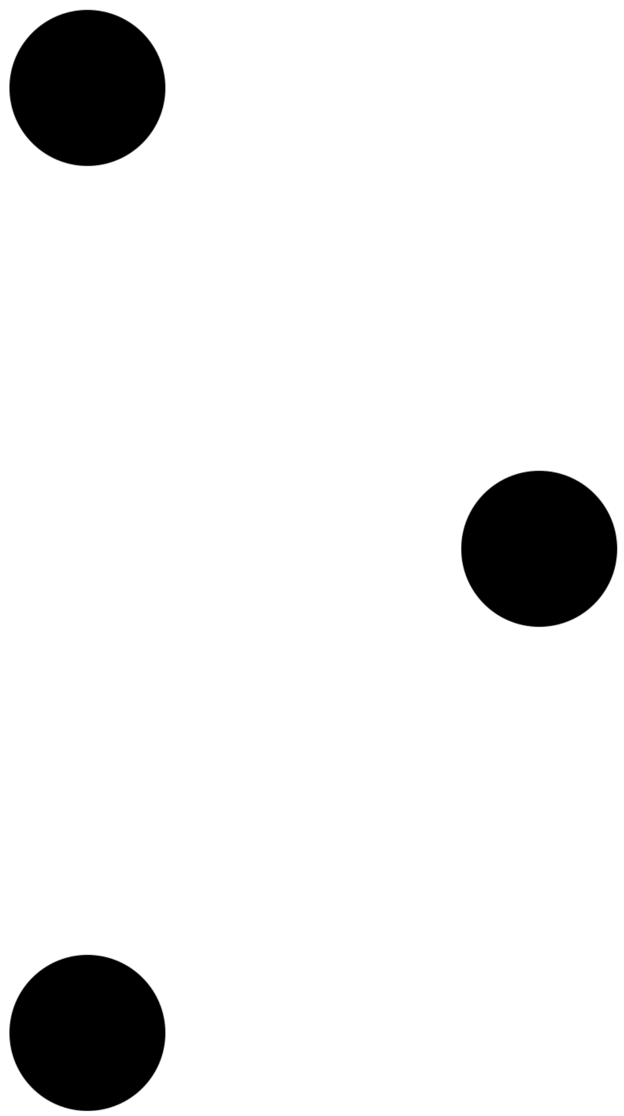


Figure A.2: The three-point calibration image used in the Pilot 1.

A.3 Reference Images - Pilot 1

Figure A.3: Reference image used before the Stroop task, to ensure that the test persons not were color-blind and had the same perception of colors as planned.

Figure A.4: Reference image used in between sub-tasks in Pilot 1. This was used to let the test persons focus on before and after sub-tasks, so that the pupil size would stabilize.

A.4 Stroop Task - Images

Figure A.5: Stroop task - The congruent version where the color of the word and the word it is describing are matching.

Figure A.6: Stroop task - Incongruent version where the color of the word and the word it is describing do not match

A.5 Calibration Image - Experimental set 1

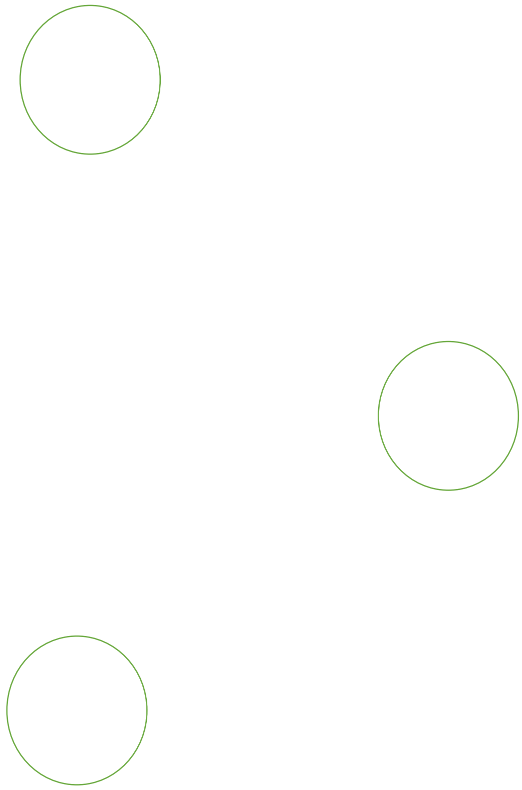


Figure A.7: Image used for the three-point calibration in the Experimental set 1.

A.6 Key Combinations - Experimental Set 1

Kommando	Exkeverings-ordning	Knappar att trycka ner	
		Vänster hand	Höger hand
1	1	1	
	2	1	
	3	1	
A B C	1		A
	2		B
	3		C
1A	1	1	A
	2		B
	3		C
1B	1		A
	2	1	B
	3		C
1C	1		A
	2		B
	3	1	C

Figure A.8: Descriptive image of key pressing combinations used in Experimental set 1

A.7 Math Questions - Experimental Set 1

Mattetal

5 + 2 = 7
10 + 4 = 14
18 + 20 = 38
9 + 7 = 16
10 + 10 = 20
38 + 10 = 48
4 + 14 = 18

8 + 3 = 11
3 + 14 = 17
18 + 4 = 22
33 + 1 = 34
5 + 15 = 20
23 + 4 = 27
12 + 2 = 24

3 + 15 = 18
22 + 10 = 32
8 + 5 = 13
3 + 4 = 7
9 + 1 = 10
16 + 4 = 20
7 + 7 = 14

4 + 5 = 9
17 + 3 = 20
4 + 2 = 6
8 + 16 = 24
21 + 20 = 41
9 + 2 = 11
28 + 30 = 58

14 + 14 = 28
9 + 7 = 16
4 + 4 = 8
11 + 4 = 15
7 + 3 = 10
12 + 8 = 20
6 + 5 = 11

9 + 3 = 12
11 + 5 = 16
25 + 25 = 50
14 + 4 = 18
20 + 20 = 40
6 + 3 = 9
14 + 7 = 21
15 + 5 = 20

Figure A.9: List of the numbers used in the mental arithmetic during Experimental set 1

A.8 Self-assessment Questionnaire

Självskattningsformulär

Undersökning om pupillmått kan användas för att bedöma kognitiv belastning

Nedan följer 2 skattningsskalor som olika aspekter av arbetsbelastningen som applicerades under testet. Indikera Din upplevda arbetsbelastning under det gångna testet med en siffra mellan 1-7, där 1 är lägsta och 7 högsta mentala belastning. Med mental belastning menas ditt koncentrationsdjup, alltså hur mycket du behövde koncentrera dig för att kunna utföra den motoriska uppgiften.

Mental belastning: Hur mentalt belastande var den motoriska uppgiften utan mattetal?

Låg Hög

Mental belastning: Hur mentalt belastande var den motoriska uppgiften med mattetal?

Låg Hög

Figure A.10: Self-assessment questionnaire used for estimating mental workload after each sub-task for Experimental set 1 and Pilot 2

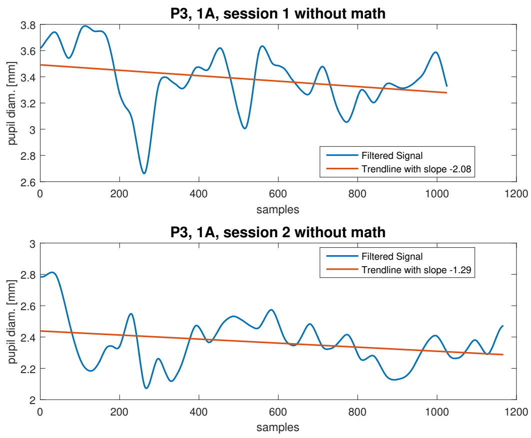
B| Appendix - Experimental Set 1

B.1 Subjective Self-assessment Rankings

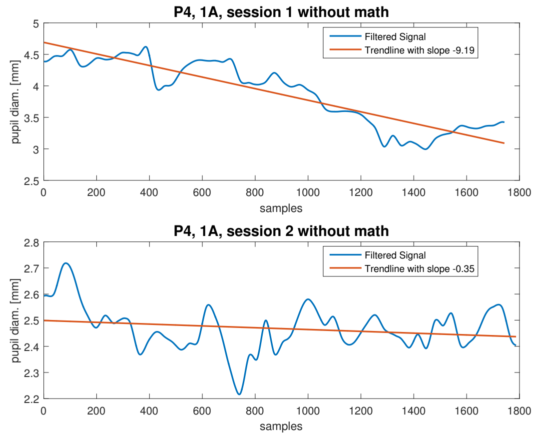
Test object	Subtask	Self assessment, without math	Self assessment, with math	Sum, mental workload	Rank, mental workload	Self assessment, without math	Self assessment, with math	Sum, mental workload	Rank, mental workload
		Session 1				Session 2			
P1	1.1	1	3	4	1	1	2	3	1
	1A	2	4	6	3	3	4	7	3
	1B	4	6	10	5	5	6	11	5
	1C	3	5	8	4	4	5	9	4
	1.2	1	3	4	1	1	2	3	1
P2	1.1	1	5	6	3	1	3	4	2
	1A	2	5	7	5	1,5	3,5	5	3
	1B	2	5	7	5	3	5	8	5
	1C	1	4	5	1	2	4	6	4
	1.2	1	4	5	1	1	2	3	1
P3	1.1	1	3	4	1	1	2	3	1
	1A	4	6	10	4	3	6	9	3
	1B	6	7	13	5	5	7	12	5
	1C	3	5	8	3	3	6	9	3
	1.2	1	3	4	1	1	2	3	1
P4	1.1	2	3	5	2	1	2	3	1
	1A	3	5	8	3	1	3	4	3
	1B	6	7	13	5	4	7	11	5
	1C	6	6	12	4	2	5	7	4
	1.2	1	3	4	1	1	2	3	1
P5	1.1	4	6	10	3	2	3	5	1
	1A	5	6	11	4	3	5	8	2
	1B	7	7	14	5	5	6	11	5
	1C	4	5	9	1	4	6	10	4
	1.2	4	5	9	1	3	5	8	2
P6	1.1	1	2	3	2	1	1	2	1
	1A	2	3	5	3	2	4	6	4
	1B	3	5	8	5	3	6	9	5
	1C	2	4	6	4	2	3	5	3
	1.2	1	1	2	1	1	2	3	2
P7	1.1	1	3	4	1	1	3	4	1
	1A	2	4	6	3	2	5	7	3
	1B	3	5	8	5	3	6	9	5
	1C	3	5	8	5	2	5	7	3
	1.2	1	4	5	2	1	4	5	2
P8	1.1	1	4	5	1	2	3	5	1
	1A	5	7	12	5	4	7	11	5
	1B	5	6	11	4	3	6	9	2
	1C	3	5	8	3	3	7	10	3
	1.2	2	3	5	1	3	7	10	3
P9	1.1	1	4	5	2	1	2	3	1
	1A	3	6	9	3	3	6	9	4
	1B	4	7	11	5	4	7	11	5
	1C	3	6	9	3	2	4	6	3
	1.2			0		1	2	3	1
P10	1.1	1	4	5	1	1	3	4	1
	1A	3	5	8	3	3	5	8	3
	1B	5	7	12	5	5	7	12	5
	1C	4	6	10	4	4	6	10	4
	1.2	1	5	6	2	1	4	5	2
P11	1.1	1	2	3	1	1	2	3	2
	1A	2	4	6	3	2	3	5	3
	1B	3	5	8	5	3	4	7	5
	1C	2	4	6	3	3	4	7	5
	1.2	1	2	3	1	1	1	2	1
P12	1.1	1	4	5	4	1	2	3	1
	1A	1	2	3	1	2	5	7	4
	1B	1	5	6	5	2	6	8	5
	1C	1	3	4	2	1	3	4	2
	1.2	1	3	4	2	1	4	5	3
P13	1.1	2	5	7	2	2	4	6	2
	1A	3	4	7	2	3	4	7	3
	1B	5	6	11	5	5	6	11	5
	1C	4	5	9	4	4	5	9	4
	1.2	2	3	5	1	1	2	3	1

Figure B.1: Summary of the test objects subjective self assessment rankings

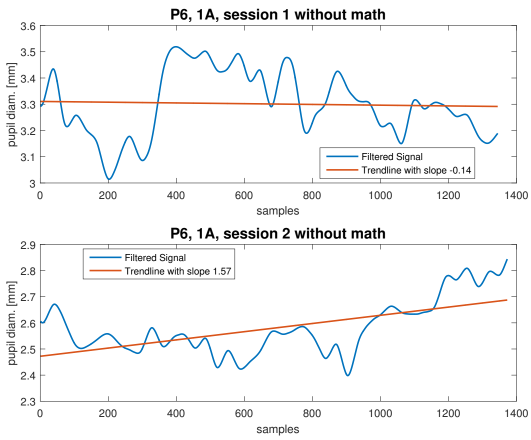
B.2 Trendlines Sub-task 1A



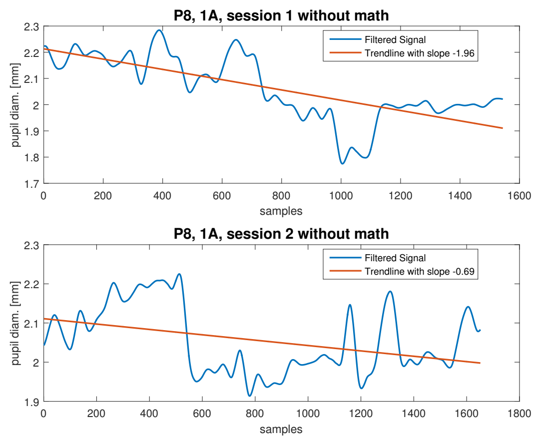
(a) Trendlines for P3, where one can observe a steeper downward slope for session 1 compared to session 2.



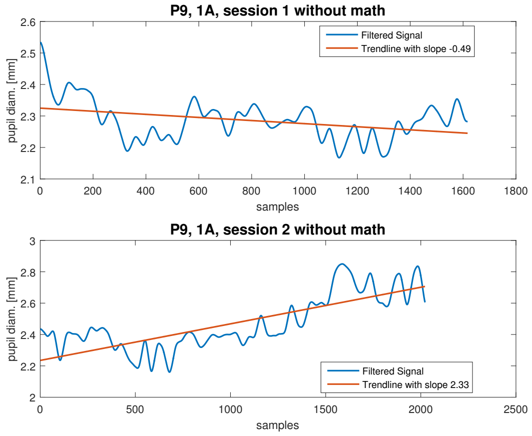
(b) Trendlines for P4, where pupil diameter is constantly decreasing during session 1 and the slope is almost flat for session 2.



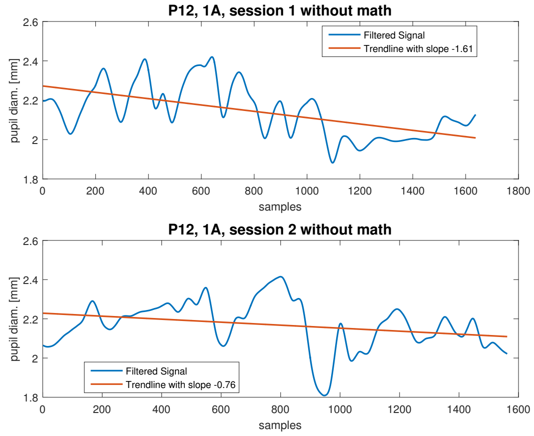
(a) Trendlines for P6 where the slope is almost zero during session 1 and the pupil diameter has an increasing behavior for session 2.



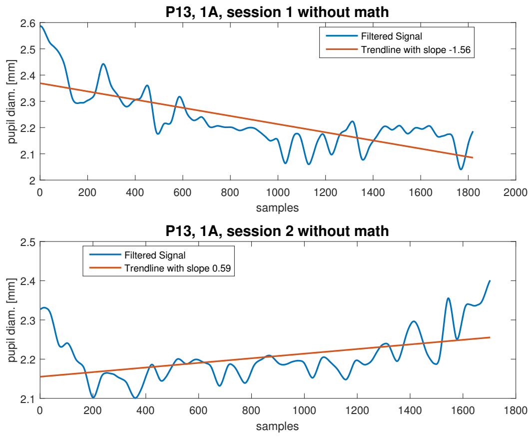
(b) Trendlines for P8 where the pupil diameter is decreasing for session 2 and the slope value is pretty close to zero for session 2.



(a) Trendlines for P9 where the pupil size is very stable for session 1 and increasing for session 2.

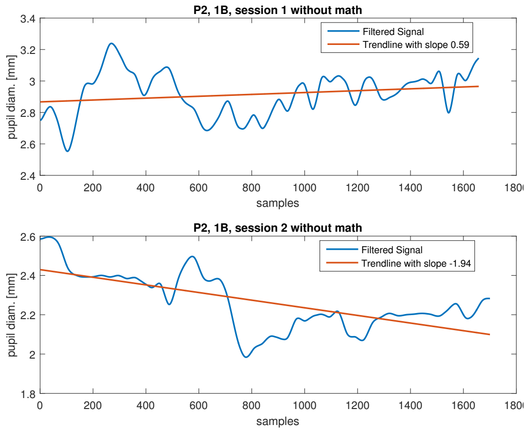


(b) Trendlines for P12 where the pupil size has a decreasing behavior for session 1 and more constant during session 2.

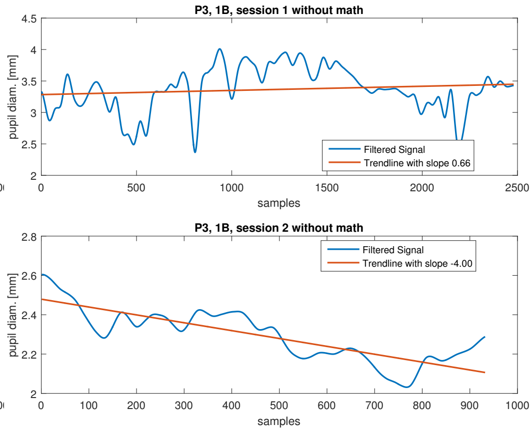


(a) Trendlines for P9 with a decreasing behavior in pupil size for session 1 and a pupil size that is relatively constant during session 2.

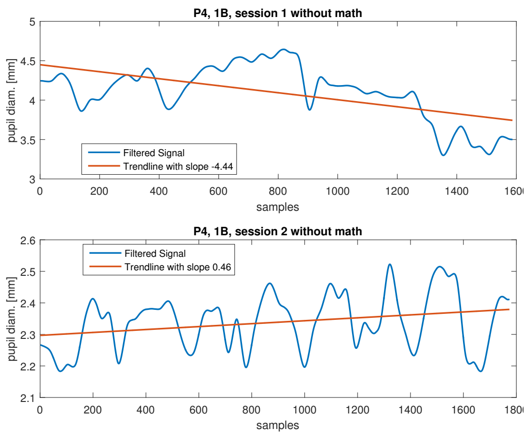
B.3 Trendlines Sub-task 1B



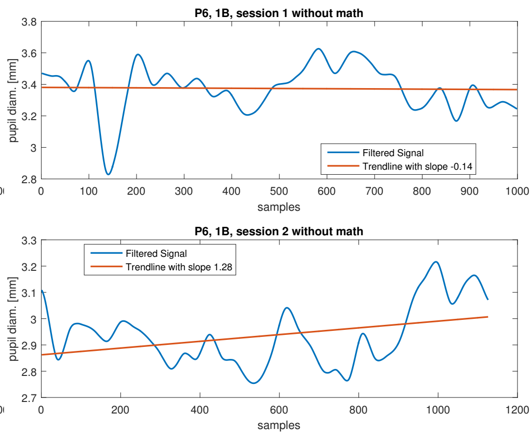
(a) Trendlines for P2, where the pupil diameter is relatively stable for session 1 and decreasing for session 2.



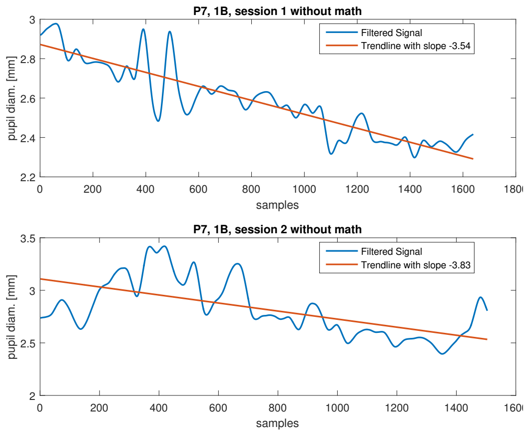
(b) Trendlines for P3, which has a stable behavior for session 1 and a decreasing behavior for session 2.



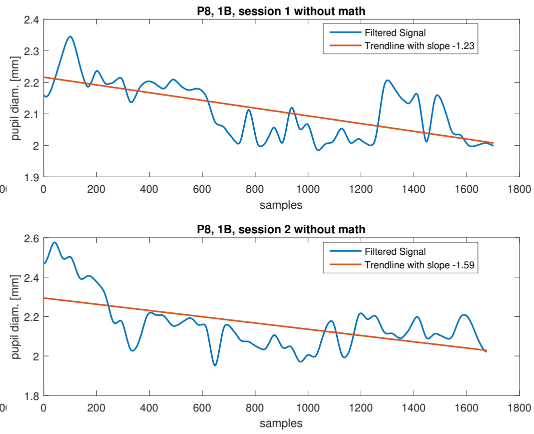
(a) Trendlines for P4 where one can observe a decreasing pupil size for session 1 and a small increase for session 2.



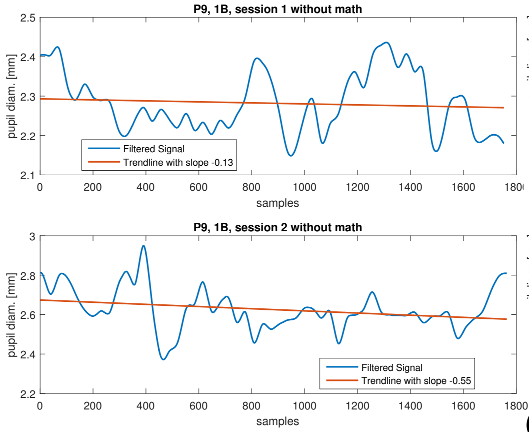
(b) Trendlines for P6 where one can observe a stable pupil behavior during session 1 and an increasing pupil diameter for session 2.



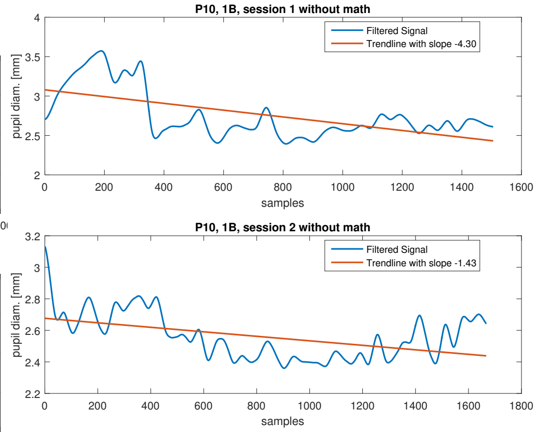
(a) Trendlines for P7 where one can observe a decreasing pupil size for both session 1 and 2.



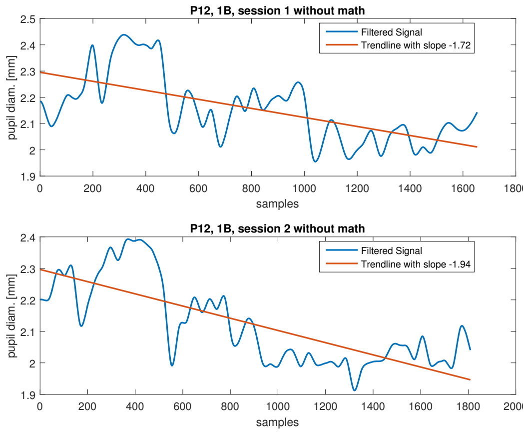
(b) Trendlines for P8 where one can observe a decreasing pupil size for both session 1 and 2.



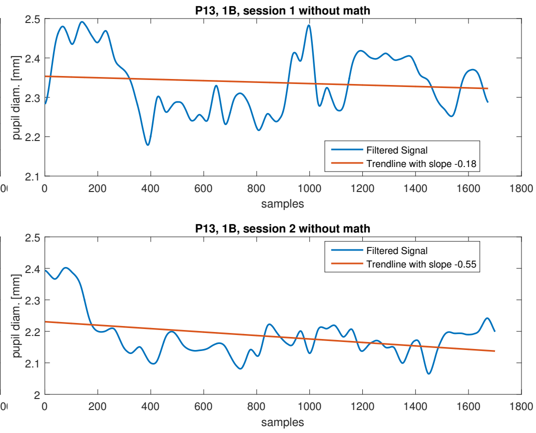
(a) Trendlines for P9 where one can observe a stable pupil behavior for both session 1 and 2.



(b) Trendlines for P10 where one can observe a decreasing pupil size for both sessions, specially during the first one.

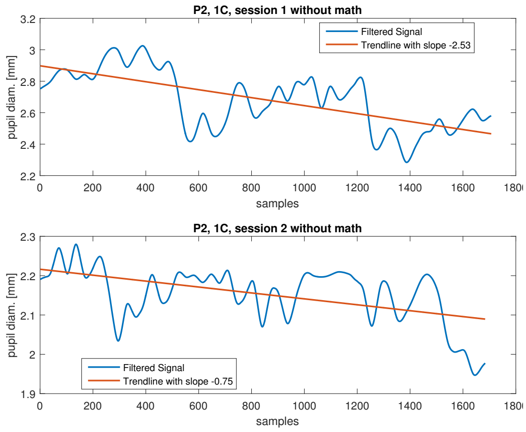


(a) Trendlines for P12 where one can observe a decreasing pupil size for both sessions.

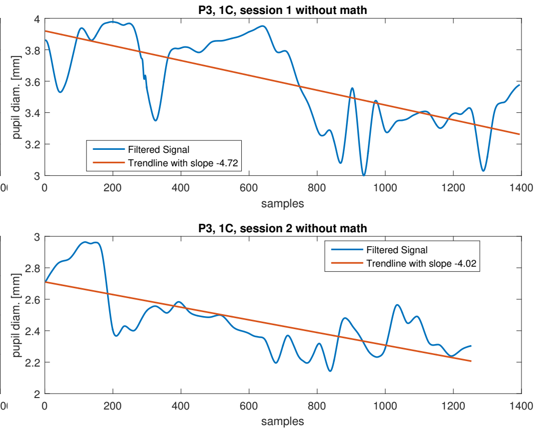


(b) Trendlines for P13 where one can observe a stable pupil size for both sessions, specially the first one.

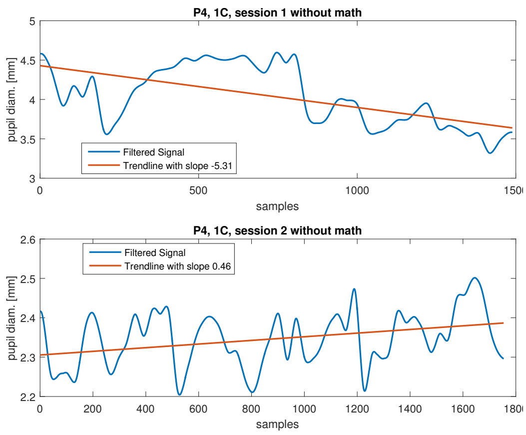
B.4 Trendlines Sub-task 1C



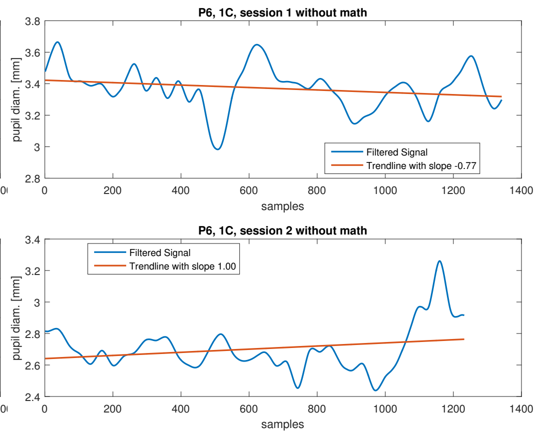
(a) Trendlines for P2, where one can observe a more steep downward slope for session 1, in comparison to session 2.



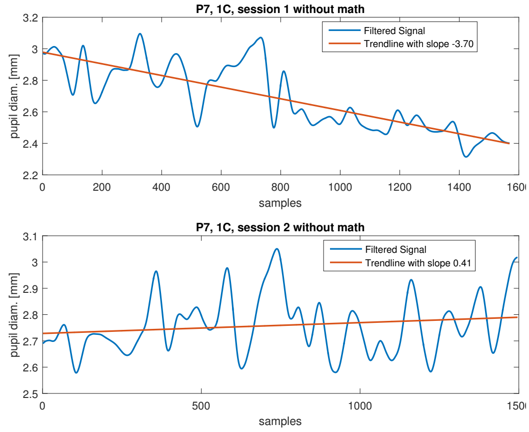
(b) Trendlines for P3, where one can oberev a (marginally) more steep downward slope for session 1, compared to session 2



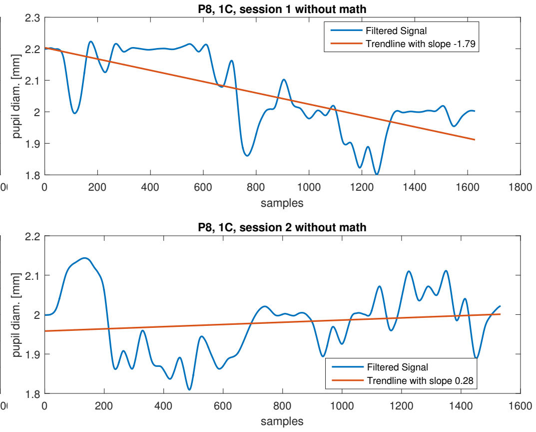
(a) Trendlines for P4, where one can observe a downward slope for session 1 and an almost stable trendline for session 2.



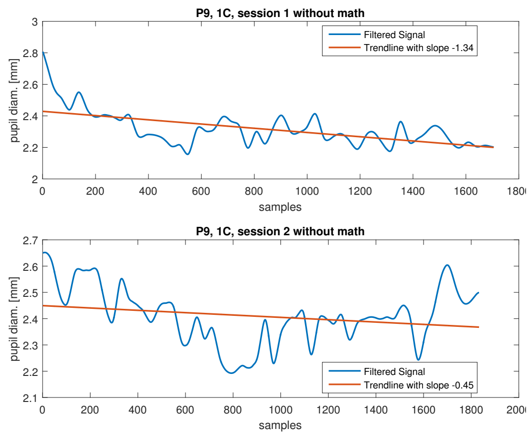
(b) Trendlines for P6, where one can observe a downward slope for session 1, and an upward slope for session 2, with almost the same absolute slope values.



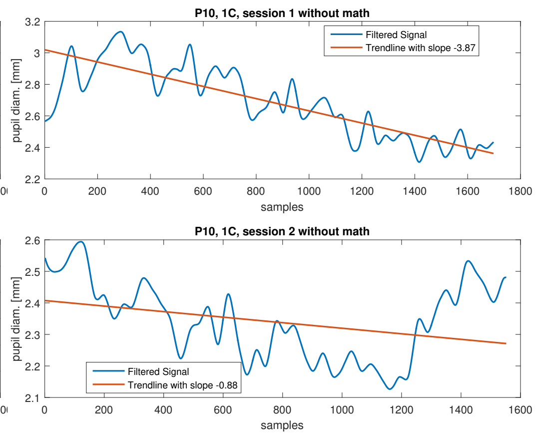
(a) Trendlines for P7, where a downward slope is observable for session 1, and an almost stable trendline for session 2.



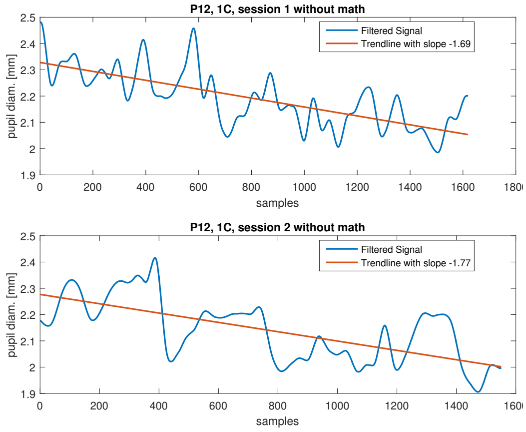
(b) Trendlines for P8, where one can observe a downward slope for session 1, and an almost stable line for session 2.



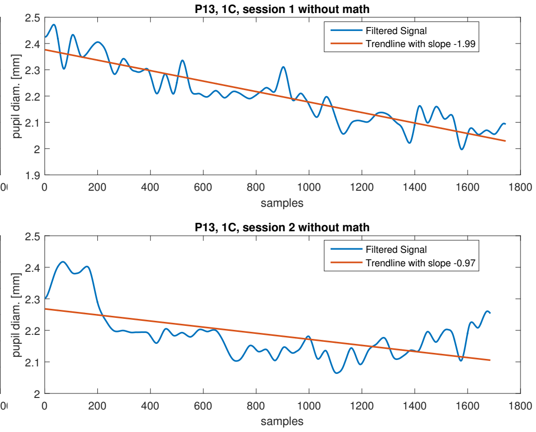
(a) Trendlines for P9, where one can observe a more steep downward slope for session 1, in comparison to session 2.



(b) Trendlines for P10, where one can observe a more steep downward slope for session 1, compared to session 2.

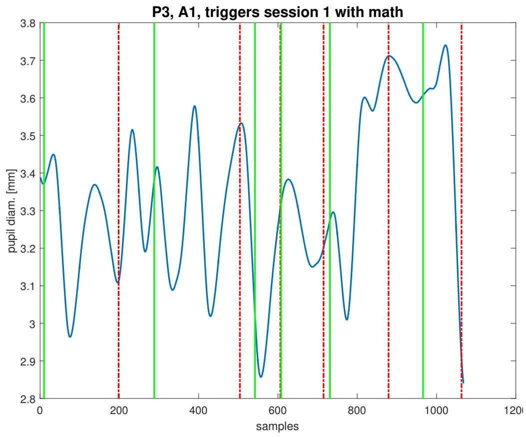


(a) Trendlines for P12, where one can observe almost the same downward slope for both session 1 and 2.

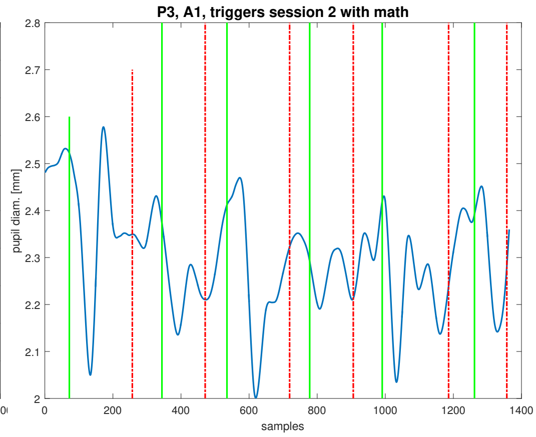


(b) Trendlines for P13, where the trendline slope value for session 1 has a more steep downward slope value, compared to session 2.

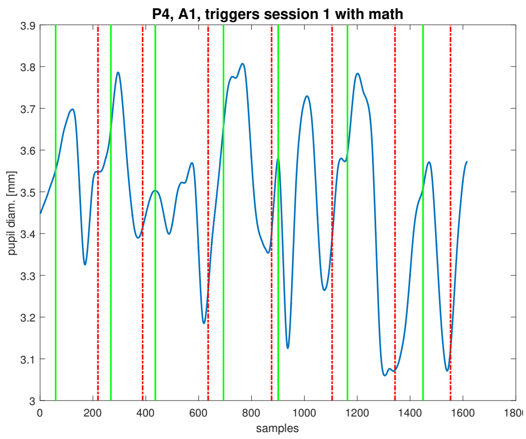
B.5 Triggers for Each Participant



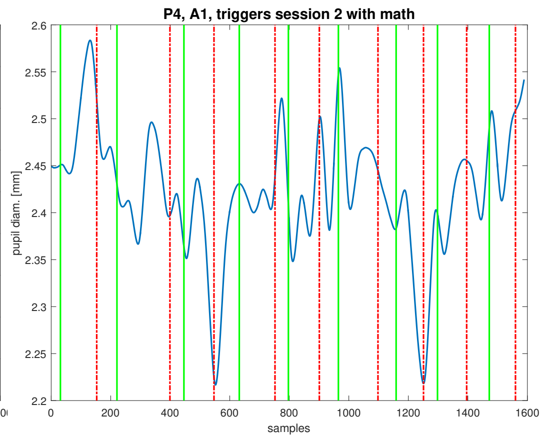
(a) Sub-task 1A with math during session 1 for test person P3



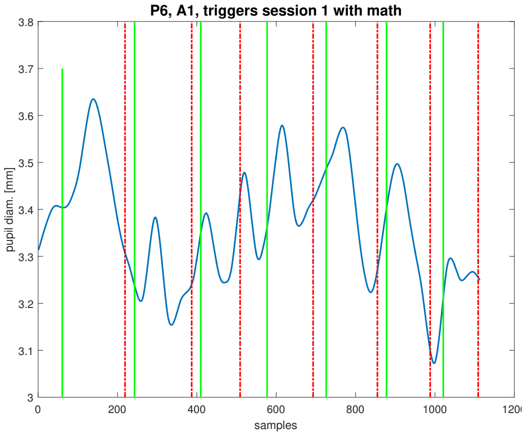
(b) Sub-task 1A with math during session 2 for test person P3.



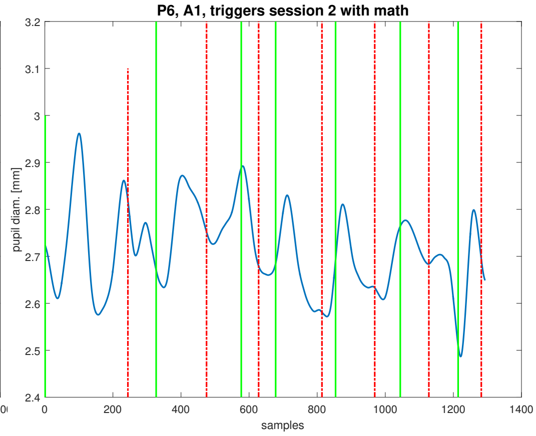
(a) Sub-task 1A with math during session 1 for test person P4



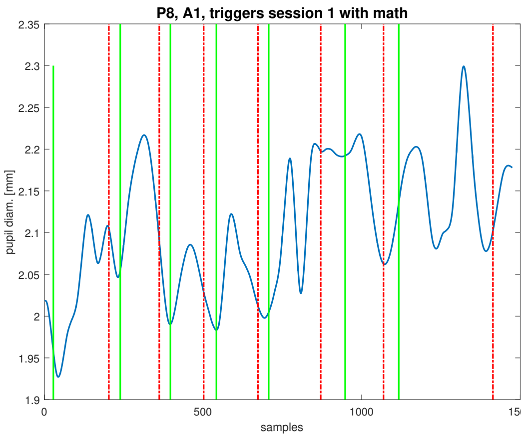
(b) Sub-task 1A with math during session 2 for test person P4.



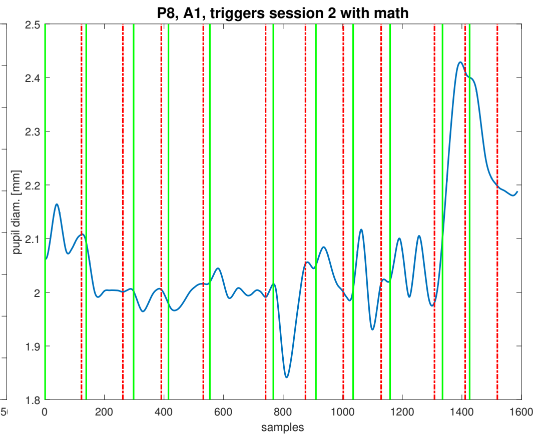
(a) Sub-task 1A with math during session 1 for test person P6



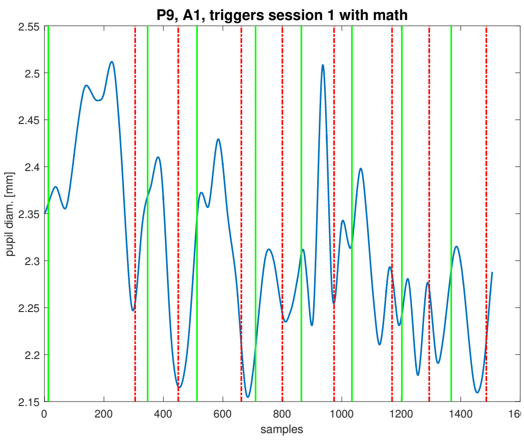
(b) Sub-task 1A with math during session 2 for test person P6.



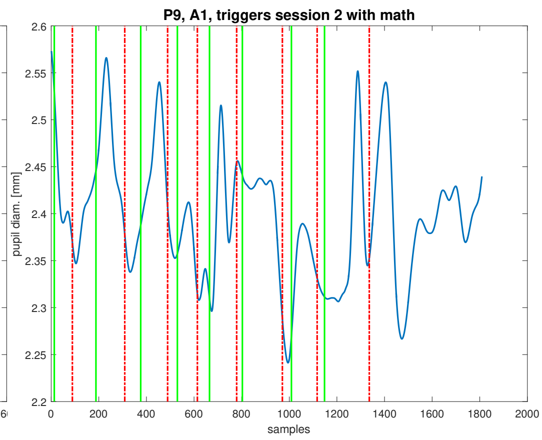
(a) Sub-task 1A with math during session 1 for test person P8



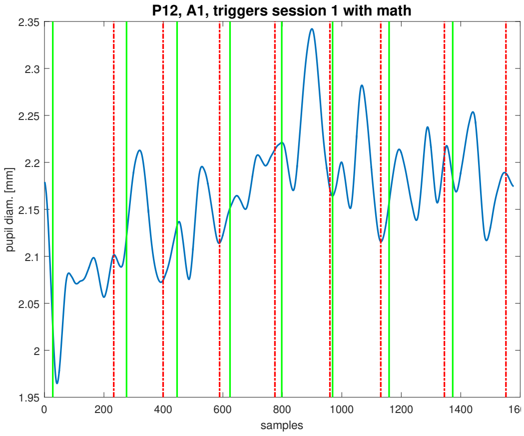
(b) Sub-task 1A with math during session 2 for test person P8.



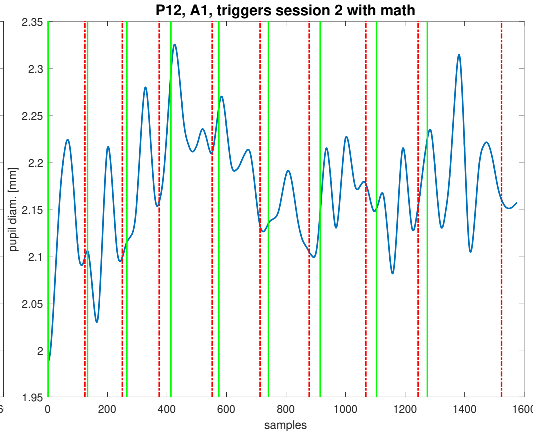
(a) Sub-task 1A with math during session 1 for test person P9



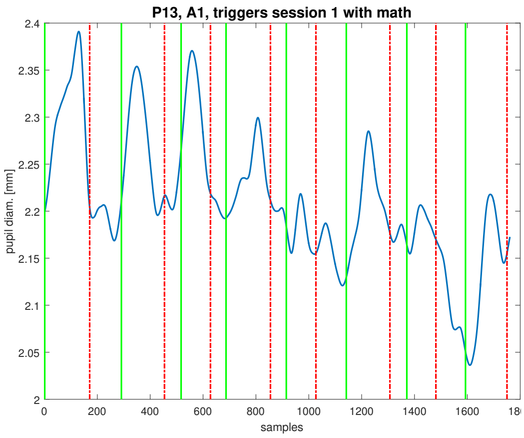
(b) Sub-task 1A with math during session 2 for test person P9.



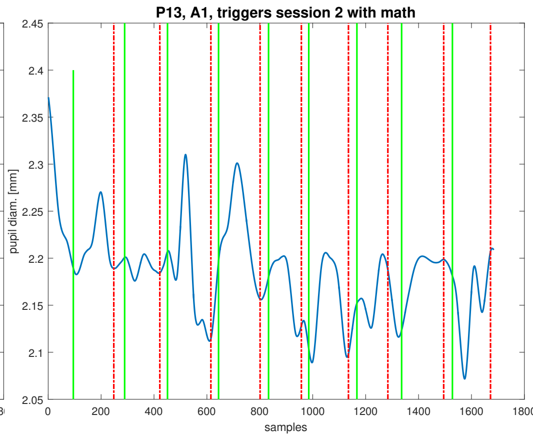
(a) Sub-task 1A with math during session 1 for test person P12



(b) Sub-task 1A with math during session 2 for test person 1P2.



(a) Sub-task 1A with math during session 1 for test person P13



(b) Sub-task 1A with math during session 2 for test person P13.

B.6 Triggers

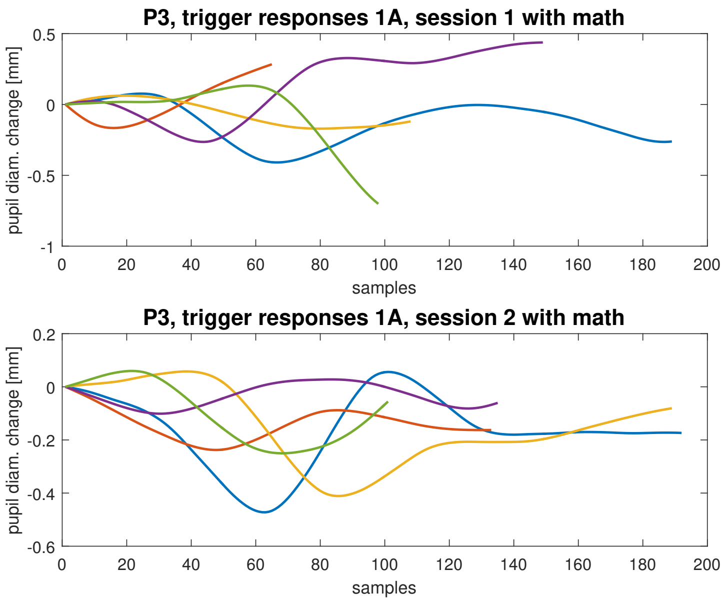


Figure B.23: Trigger responses for P3 when performing sub-task 1A, where each line represents a math question. The upper plot illustrates the trigger responses for session 1 and the lower for session 2.

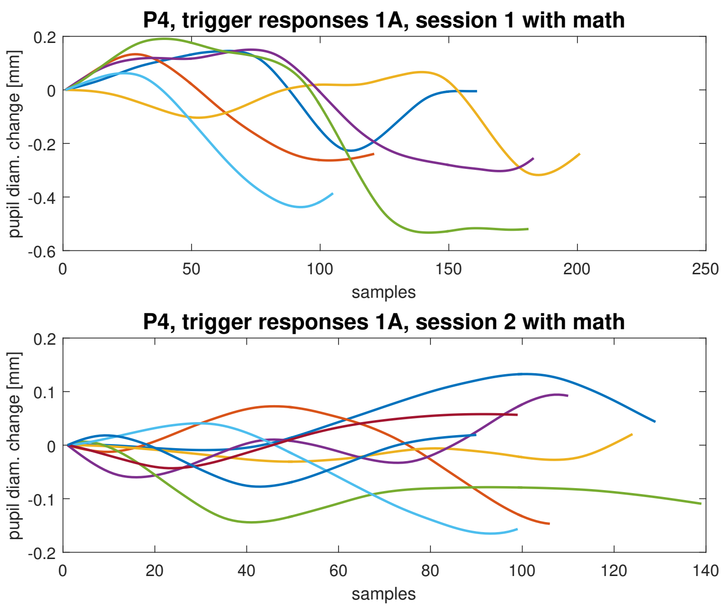


Figure B.24: Trigger responses for P4 when performing sub-task 1A, where each line represents a math question. The upper plot illustrates the trigger responses for session 1 and the lower for session 2.

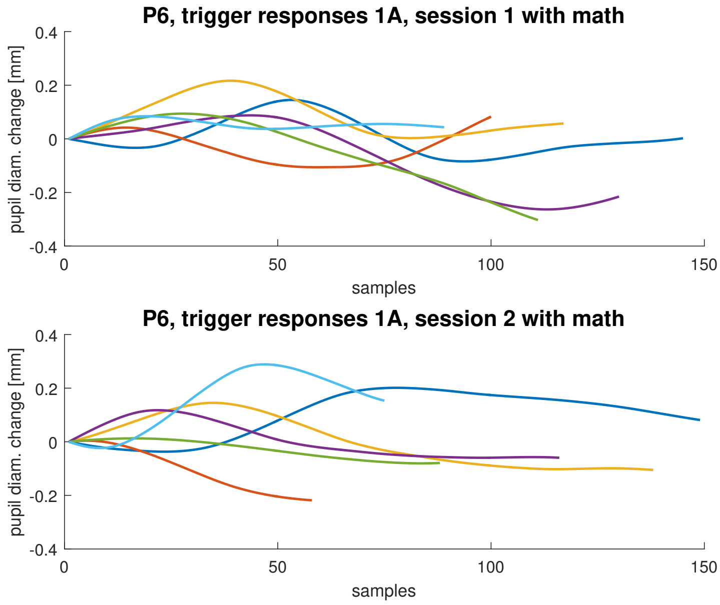


Figure B.25: Trigger responses for P6 when performing sub-task 1A, where each line represents a math question. The upper plot illustrates the trigger responses for session 1 and the lower for session 2.

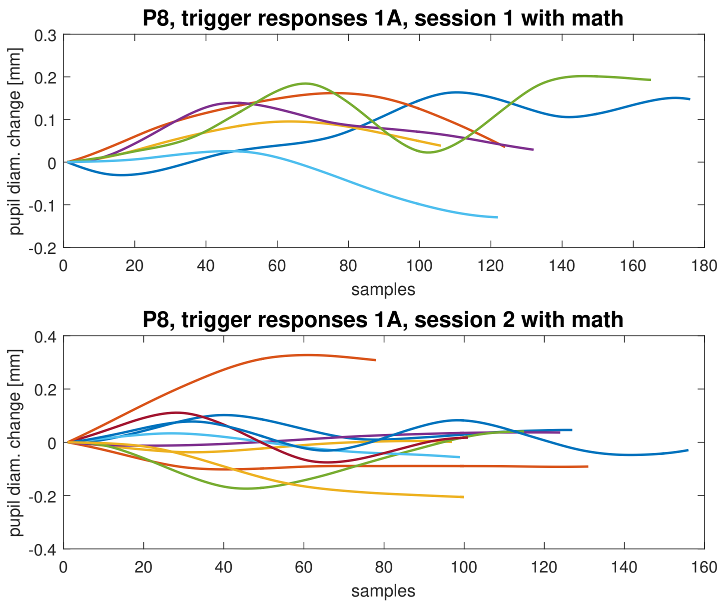


Figure B.26: Trigger responses for P8 when performing sub-task 1A, where each line represents a math question. The upper plot illustrates the trigger responses for session 1 and the lower for session 2.

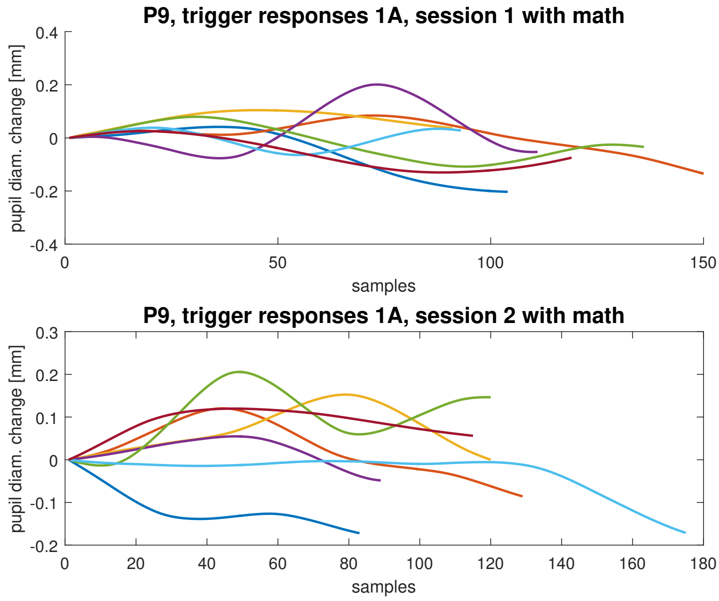


Figure B.27: Trigger responses for P9 when performing sub-task 1A, where each line represents a math question. The upper plot illustrates the trigger responses for session 1 and the lower for session 2.

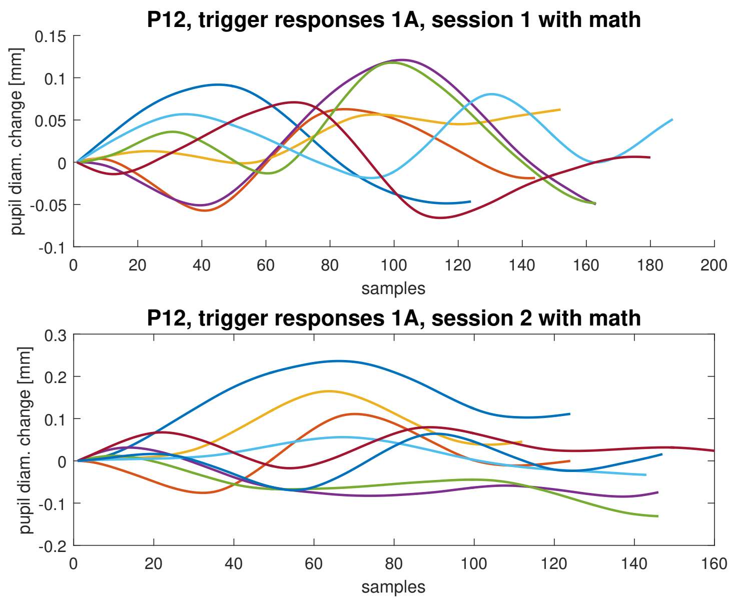


Figure B.28: Trigger responses for P12 when performing sub-task 1A, where each line represents a math question. The upper plot illustrates the trigger responses for session 1 and the lower for session 2.

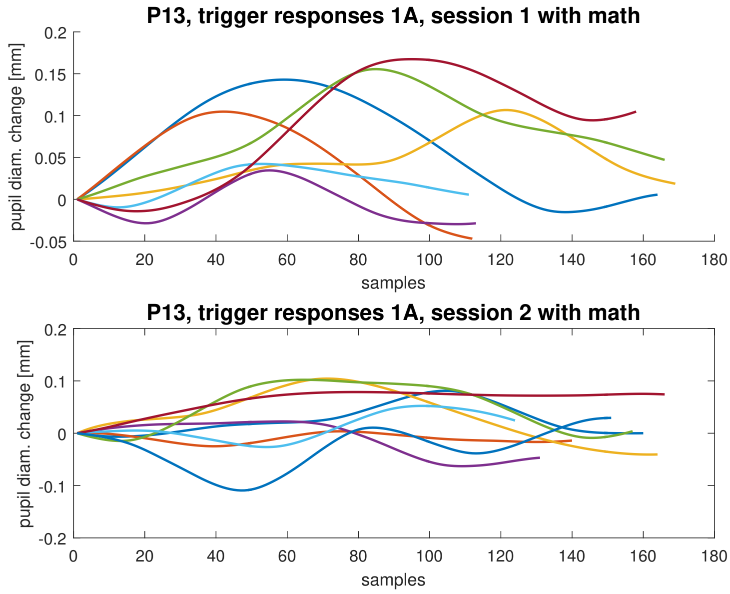
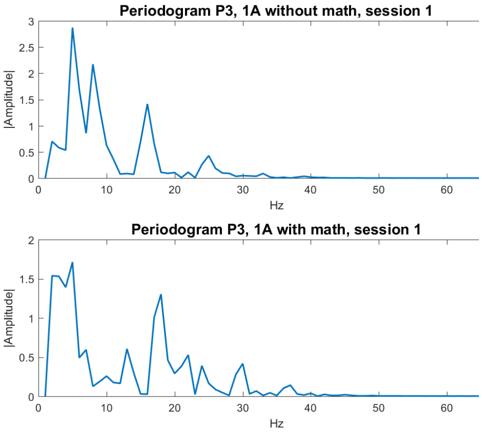
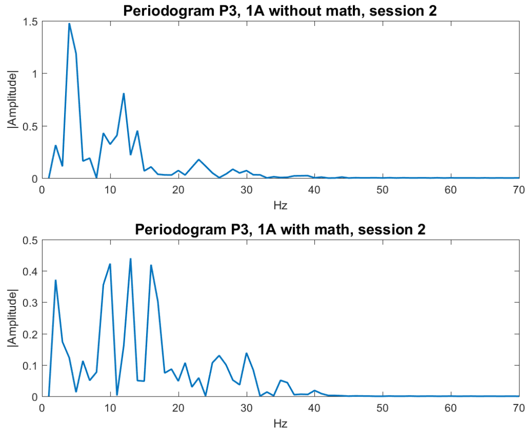


Figure B.29: Trigger responses for P13 when performing sub-task 1A, where each line represents a math question. The upper plot illustrates the trigger responses for session 1 and the lower for session 2.

B.7 Periodograms

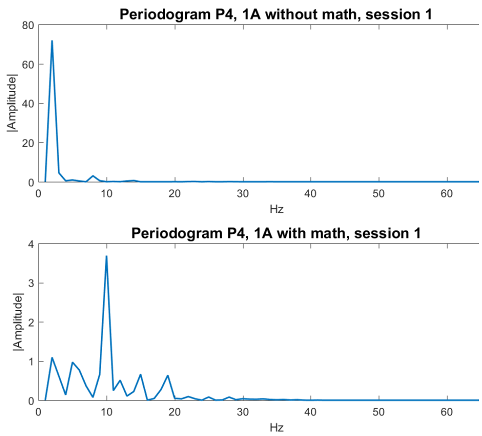


(a) Periodogram for P3, session 1

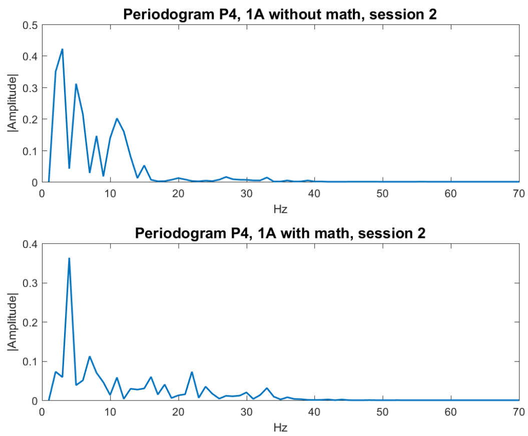


(b) Periodogram for P3, session 2

Figure B.30: Periodogram for P3, session 1 and 2, where the upper plots represents frequency content for the sub-task without math and the lower ones with math. The left plots show session 1 and the right ones session 2.

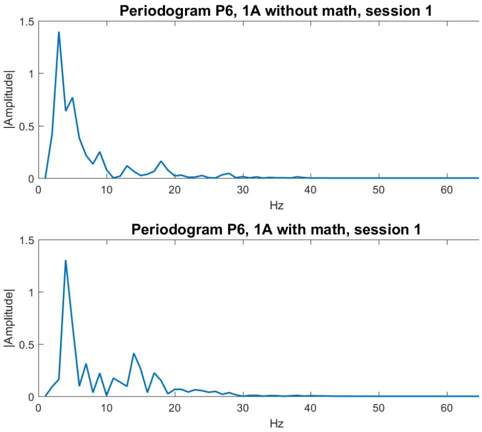


(a) Periodogram for P4, session 1

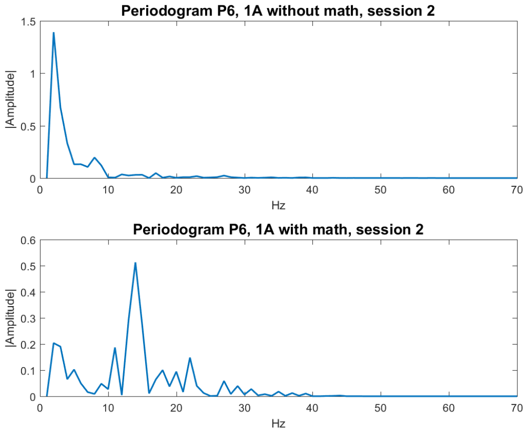


(b) Periodogram for P4, session 2

Figure B.31: Periodogram for P4, session 1 and 2, where the upper plots represents frequency content for the sub-task without math and the lower ones with math. The left plots show session 1 and the right ones session 2.

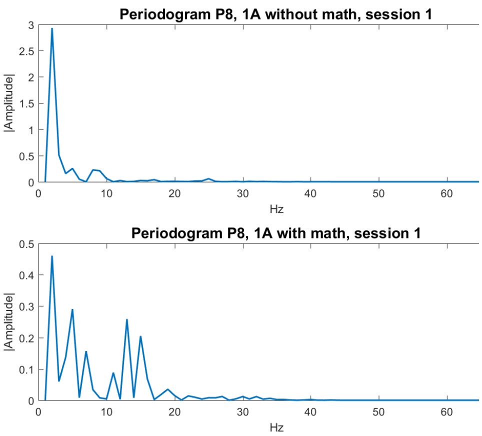


(a) Periodogram for P6, session 1

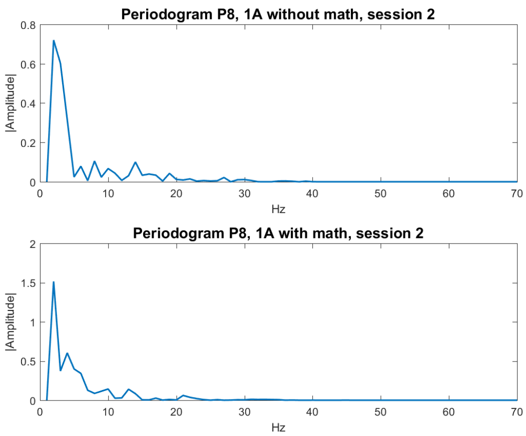


(b) Periodogram for P6, session 2

Figure B.32: Periodogram for P6, session 1 and 2, where the upper plots represents frequency content for the sub-task without math and the lower ones with math. The left plots show session 1 and the right ones session 2.

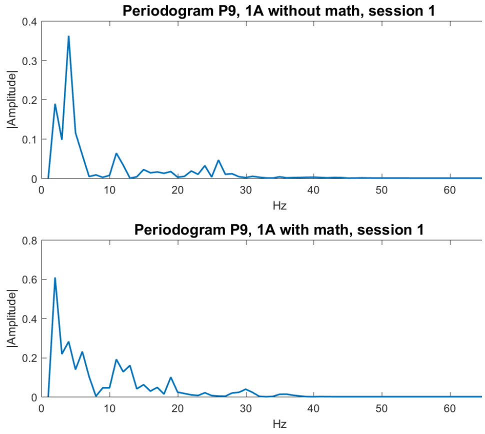


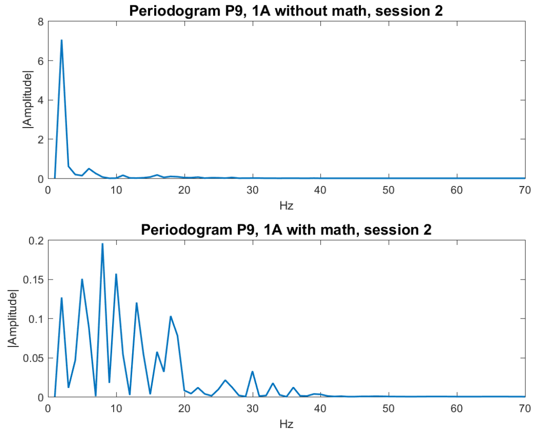
(a) Periodogram for P8, session 1



(b) Periodogram for P8, session 2

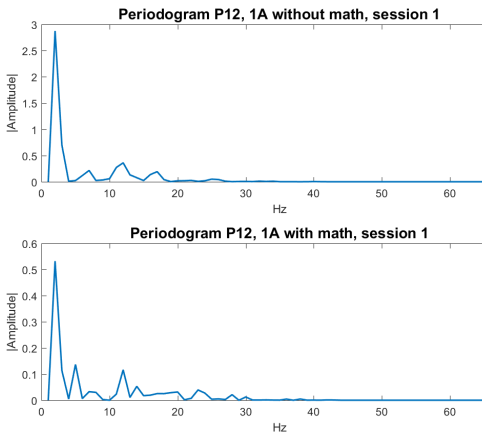
Figure B.33: Periodogram for P8, session 1 and 2, where the upper plots represents frequency content for the sub-task without math and the lower ones with math. The left plots show session 1 and the right ones session 2.

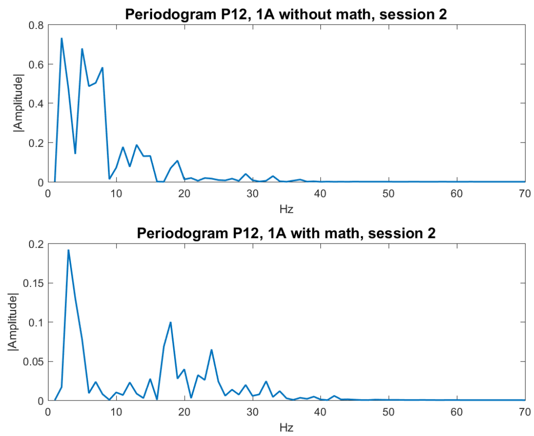




(a) Periodogram for P9, session 1 **(b)** Periodogram for P9, session 2

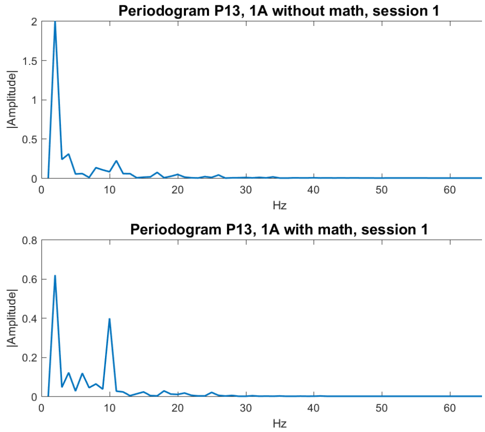
Figure B.34: Periodogram for P9, session 1 and 2, where the upper plots represents frequency content for the sub-task without math and the lower ones with math. The left plots show session 1 and the right ones session 2.



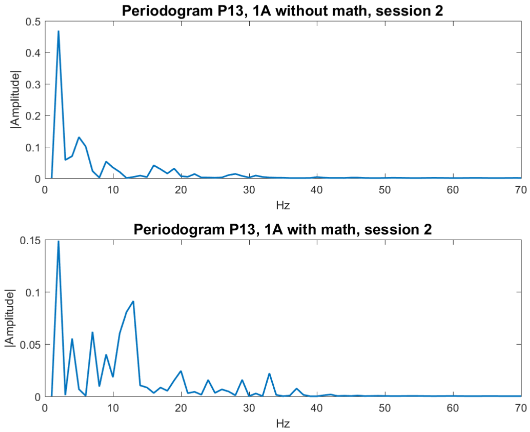


(a) Periodogram for P12, session 1 **(b)** Periodogram for P12, session 2

Figure B.35: Periodogram for P12, session 1 and 2, where the upper plots represents frequency content for the sub-task without math and the lower ones with math. The left plots show session 1 and the right ones session 2.



(a) Periodogram for P13, session 1



(b) Periodogram for P13, session 2

Figure B.36: Periodogram for P13, session 1 and 2, where the upper plots represents frequency content for the sub-task without math and the lower ones with math. The left plots show session 1 and the right ones session 2.

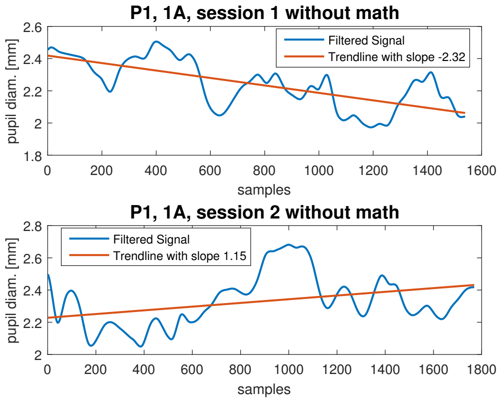
C| Appendix - Pilot 2

C.1 Subjective Self Assessment Rankings

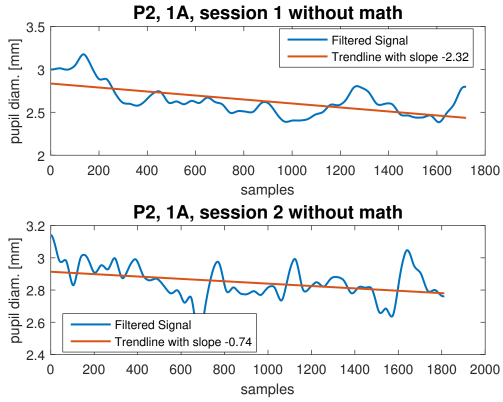
Test object	Subtask	Session 1				Session 2			
		Self assessment, without math	Self assessment, with math	Sum, mental workload	Rank, mental workload	Self assessment, without math	Self assessment, with math	Sum, mental workload	Rank, mental workload
P14	1.1	1	4	5	2	1	2	3	**1**
	1A	3	5	**8**	4	2	3	5	3
	1B	4	6	**10**	**5**	3	4	7	**5**
	1C	2	4	6	3	2	3	5	3
	1.2	1	3	4	**1**	1	2	3	**1**
P15	1.1	1	2	**3**	1	1	2	3	2
	1A	2	4	6	3	3	5	8	3
	1B	2	4	6	**3**	5	6	11	**5**
	1C	3	4	**7**	**5**	3	5	8	3
	1.2	1	2	3	**1**	1	1	2	**1**
P16	1.1	2	3	**5**	**1**	2	3	5	2
	1A	3	5	8	3	3	5	8	4
	1B	5	6	**11**	**5**	4	6	10	**5**
	1C	4	6	10	4	3	4	7	3
	1.2	2	3	**5**	**1**	1	2	3	**1**
P17	1.1	1	3	**4**	1	1	3	4	2
	1A	2	5	7	4	3	5	8	3
	1B	3	5	**8**	**5**	4	5	9	**5**
	1C	3	3	6	3	4	5	9	5
	1.2	1	3	**4**	**1**	1	2	3	**1**

Figure C.1: Summary of the test objects subjective self assessment rankings from Pilot 2

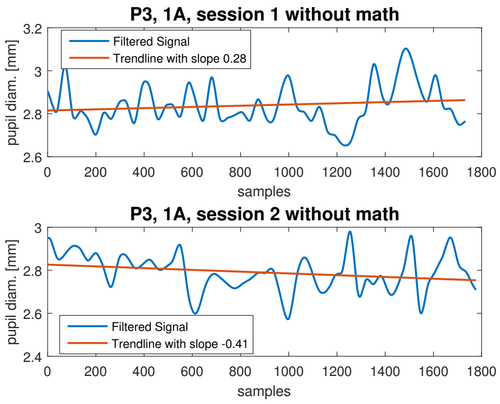
C.2 Trendlines, Sub-task 1A



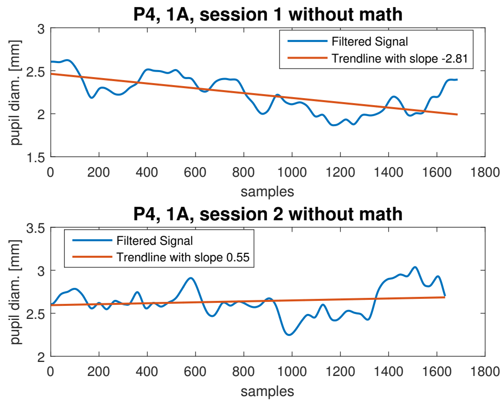
(a) Trendlines for P1, Pilot 2, where one can observe a decreasing trend for session 1 compared to session 2



(b) Trendlines for P2, Pilot 2, where one can observe a more decreasing behavior for session 1 than session 2

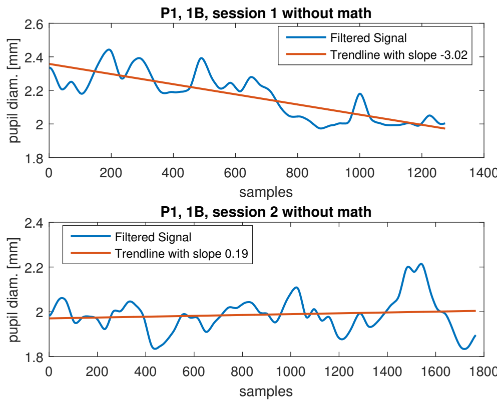


(a) Trendlines for P3, session 2, where the trend of the pupil diameter is almost flat for session 2

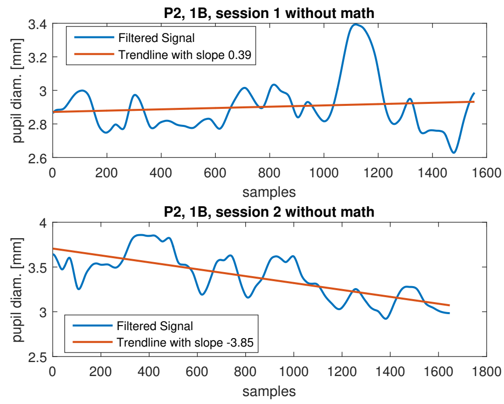


(b) Trendlines for P4, session 2, where the trend of the pupil diameter is almost flat for session 1 and 2.

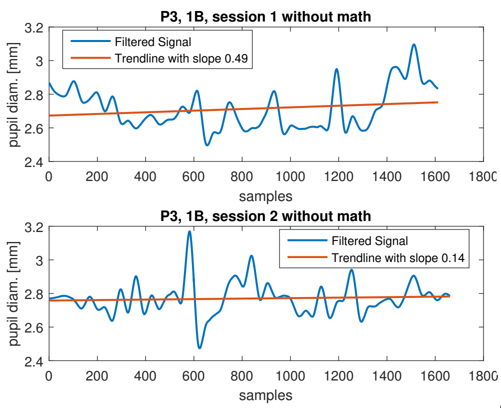
C.3 Trendlines, Sub-task 1B



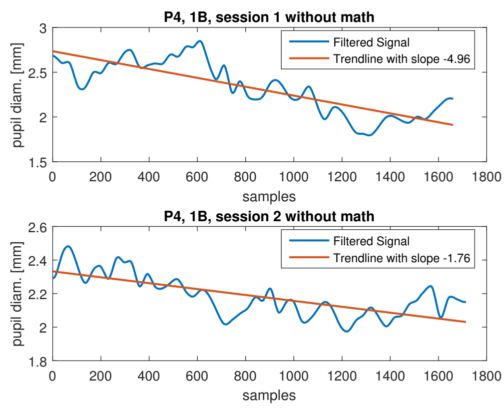
(a) Trendlines for P1, where the trendline for session 2 is almost stable, and one can observe a downward slope for session 1.



(b) Trendlines for P2, where one can observe an almost stable trendline for session 1, and a steep downward slope for session 2.

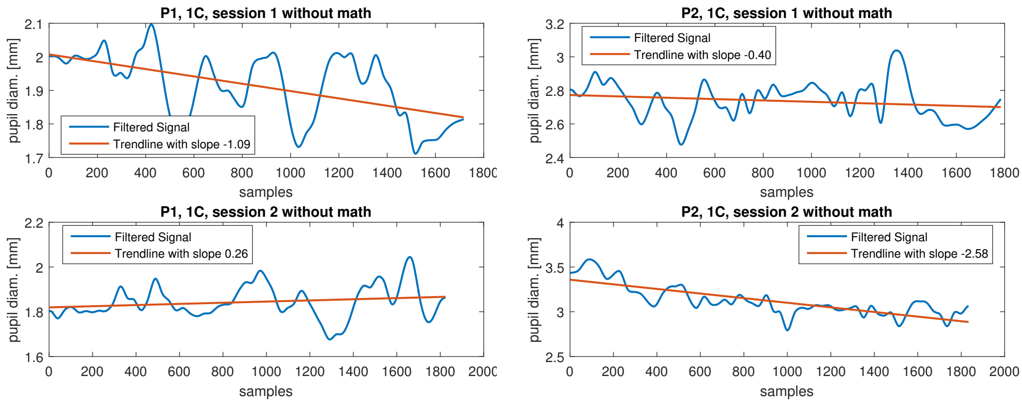


(a) Trendlines for P3, where one can observe almost stable trendlines for both session 1 and 2.



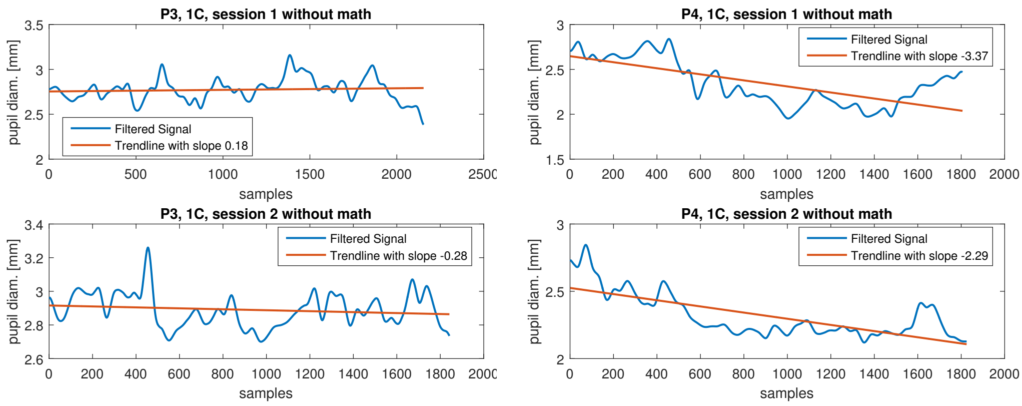
(b) Trendlines for P4, where one can observe a more steep downward slope for session 1, in comparison to session 2.

C.4 Trendlines Sub-task 1C



(a) Trendlines for P1, where one can observe a downward slope for session 1 and an almost stable slope for session 2.

(b) Trendlines for P2, where one can observe an almost stable line for session 1, and a steeper downward slope for session 2.



(a) Trendlines for P3, where one can observe almost stable trendlines for both session 1 and 2.

(b) Trendlines for P3, where one can observe almost stable trendlines for both session 1 and 2.