

Ville Ahonen

# **ELECTROENCEPHALOGRAPHY IN EVALUATING MENTAL WORKLOAD OF GAMING**

Master of Science Thesis  
Faculty of Engineering and Natural Sciences  
November 2020

# ABSTRACT

Ville Ahonen: Electroencephalography in evaluating mental workload of gaming  
Master of Science Thesis  
Tampere University  
Degree Programme in Management and Information Technology  
November 2020

---

In this thesis, the feasibility of the electroencephalography (EEG) analysis in evaluating the mental workload of gaming was studied, primarily by giving an overview on the related research, and secondarily as a proof-of-concept type experiment on existing EEG recordings, with a tool implemented for the purpose. In a relation to the EEG analysis, a selective overview on underlying mathematical methods and techniques was given as well.

As a result of the review on various studies and their outcome, it was evident that the EEG analysis provides a plausible means to objectively measure and evaluate mental load imposed by gaming. The EEG indicators, that had been successfully deployed in mental load evaluation in the reviewed studies, utilized power spectrum, event related potential and brain connectivity related measurement methods.

In the experimental part of the thesis, a tool to process EEG signals and to calculate EEG metrics, was implemented in Matlab environment. The existing EEG recordings (20 recordings in total), that were used in the experiment, had been acquired by groups of students and staff of Tampere University during n-back gaming sessions, as a part of course projects. The ratio of theta and alpha power, calculated over the EEG signal segments that were time-locked to game events, was selected as EEG metrics for mental load evaluation. The expectation, based on the reviewed studies, was that the value of the calculated ratio should increase with increasing mental load. The Wilcoxon rank-sum test was applied to test this hypothesis for the ratio values combined from all recordings. The rank-sum test results revealed that the theta-alpha power ratio performed as a confident indicator for the evaluation and comparison of mental load. It should be noted that this was valid only for the frontal channels Fp1 and Fp2, of the recordings, and at the highest game difficulty level the calculated ratio values started to appear inconsistent, which could be a consequence of possible concentration issues, as the task became too demanding.

Keywords: EEG, electroencephalography, mental load, gaming, n-back game, EEG analysis, EEG indicator

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

# TIIVISTELMÄ

Ville Ahonen: Aivosähkökäyrätutkimuksen soveltaminen kuormittavuuden arvioinnissa pelien yhteydessä  
Diplomityö  
Tampereen yliopisto  
Johtamisen ja tietotekniikan DI-tutkinto-ohjelma  
Marraskuu 2020

---

Tässä diplomityössä selvitettiin aivosähkökäyräanalyysin soveltuvuutta pelaamisesta aiheutuvan kuormituksen arvioinnissa. Ensisijaisesti selvitys toteutettiin katsauksena aiheesta julkaistuihin tutkimuksiin. Tämän lisäksi työn kokeellisessa osuudessa analysoitiin olemassa olevia EEG-rekisteröintejä tätä tarkoitusta varten toteutetun ohjelmistotyökalun avulla. Työssä myös esiteltiin valikoidusti EEG-analyysiin liittyviä matemaattisia menetelmiä ja tekniikoita.

Useita tutkimuksia käsittäneen katsauksen perusteella oli ilmeistä, että EEG-analyysi on käytökelpoinen menetelmä pelaamisesta aiheutuvan kuormituksen objektiiviseen mittaamiseen ja arviointiin. Näissä tutkimuksissa kuormituksen arvioinnissa käytetyt EEG-mittarit perustuivat joko EEG:n tehospektriä, tapahtumasidonnaista herätepotentiaalia tai aivojen kytketyneisyyttä hyödyntäviin mittausmenetelmiin.

Kokeellisessa osuudessa toteutettiin Matlab-ympäristössä ohjelmistotyökalu EEG-signaalien prosessoimiseksi ja EEG-mittareiden laskemiseksi. Kokeilussa käytettiin Tampereen yliopiston opiskelijoiden ja henkilökunnan aiemmissa kurssiprojekteissa "n-back"-pelin pelaamisen yhteydessä taltioimia EEG-rekisteröintejä, joita oli yhteensä 20. Kuormituksen arviointiin käytettäväksi EEG-mittariksi valittiin theeta- ja alfa-taajuuskaistojen tehojen välinen suhde. Tehosuhte laskettiin EEG-signaalien segmenteille, jotka olivat sidoksissa pelitapahtumiin. Tutkimuskatsauksen perusteella odotuksena oli, että tehosuhde kasvaa kuormituksen kasvaessa. Tämän hypoteesin paikansapitävyyttä arvioitiin Wilcoxonin järjestyssummatestillä. Testin tulokset osoittivat, että theeta- ja alfa-taajuuskaistojen välinen tehosuhde on varsin luotettava mittari kuormituksen arviointiin ja kuormitusten väliseen vertailuun. Huomioitavaa on, että mitatuista EEG-kanavista tämä tulos pätee ainoastaan etukanaville Fp1 ja Fp2. Lisäksi korkeimmalla pelin vaikeustasolla mittarin antamat arvot olivat epäyhtenäisiä, mikä saattoi olla seurausta mahdollisista keskittymisongelmista pelin vaikeustason kasvaessa liiallisesti.

Avainsanat: EEG, aivosähkökäyrä, henkinen kuormitus, pelaaminen, n-back-peli, EEG-analyysi, EEG-mittari

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

## PREFACE

The EEG provides a window for studying the human brain activity. I feel quite lucky having been given an opportunity to look through that window, and to learn how that scenery can be interpreted.

I wish to give my sincere thanks to professor Tarmo Lipping. First of all, for introducing this most interesting subject for the thesis, and then, for all the guidance, advice, comments and support throughout the course of writing the thesis, those have been invaluable. I also want to thank MSc Marko Leino, for the comments, and as the second examiner for this work.

Last, but definitely not least, I want to express my gratitude to the one who made this all possible, to my lovely wife Marika. You were the driving force, when I was in doubt should I even start this challenging journey for a master's degree, and you gave your full support and understanding for those countless evening and weekend hours I needed for my studies. But now, as strange as it may seem, this journey is coming to its end, and I will be fully back in the team... just the very last period, right here.

Riihimäki, 27th November 2020

Ville Ahonen

# CONTENTS

1	Introduction . . . . .	1
2	Electroencephalography . . . . .	3
2.1	Origin of the EEG . . . . .	3
2.2	Cerebral cortex . . . . .	5
2.3	EEG measuring . . . . .	7
2.4	Interpretation of the brain waves . . . . .	9
2.5	Evoked potentials and event related potentials . . . . .	10
2.6	Artefacts . . . . .	11
2.6.1	Sources of artefacts . . . . .	12
2.6.2	Artefact detection and removal . . . . .	12
3	Mental load, gaming and EEG . . . . .	14
3.1	Mental workload . . . . .	14
3.1.1	Cognitive load theory . . . . .	14
3.1.2	Mental workload measures . . . . .	15
3.2	Mental workload induced by video game playing . . . . .	16
3.2.1	Sources of mental load in gaming . . . . .	16
3.2.2	Learning in games . . . . .	17
3.3	EEG indicators on mental workload . . . . .	18
3.3.1	Power spectrum . . . . .	18
3.3.2	Event related potential . . . . .	20
3.3.3	Brain connectivity . . . . .	20
3.3.4	Adjustment of automation based on mental load assessment . . . . .	21
3.4	Gaming and cognitive skills . . . . .	22
3.5	EEG in virtual reality environment . . . . .	23
3.5.1	EEG measuring during VR session . . . . .	23
3.5.2	Mental load evaluation in VR environment . . . . .	24
3.6	N-back memory game and EEG . . . . .	25
4	EEG analysis methods . . . . .	27
4.1	Power Spectral Density . . . . .	27
4.2	Wavelet transform . . . . .	32
4.3	Coherence . . . . .	39
4.4	Phase-amplitude coupling and phase locking value . . . . .	39
5	Experiment on n-back recordings . . . . .	42
5.1	EEG validator and calculator . . . . .	42
5.1.1	Required functionality for the tool . . . . .	42
5.1.2	Programming language and runtime environment . . . . .	45

5.1.3	Tool architecture . . . . .	45
5.1.4	Tool configuration . . . . .	48
5.1.5	Functional split for the validator . . . . .	49
5.1.6	Functional split for the calculator . . . . .	52
5.2	The n-back experiment . . . . .	53
5.2.1	The setup and procedure . . . . .	53
5.2.2	The EEG processing . . . . .	54
5.2.3	The EEG metrics calculation . . . . .	55
5.2.4	Results and analysis . . . . .	55
6	Conclusions . . . . .	61
	References . . . . .	63
	Appendix A P-values separately for each recording and each channel . . . . .	70

## LIST OF FIGURES

1.1	Number of EEG related publications by year (in PubMed database) . . . . .	1
2.1	Neuron . . . . .	4
2.2	The phases of the action potential . . . . .	4
2.3	Brain lobes . . . . .	5
2.4	Cortical layers . . . . .	6
2.5	The 10-20 system . . . . .	7
2.6	Electrode positions . . . . .	8
2.7	Examples of brain rhythms in an EEG signal . . . . .	9
2.8	Naming convention for EP/ERP components . . . . .	11
3.1	Mental workload types . . . . .	16
3.2	Flow channel . . . . .	17
3.3	Individual alpha frequency . . . . .	19
4.1	Window functions . . . . .	30
4.2	Frequency responses of window functions . . . . .	30
4.3	Welch PSD estimates (window length is 128 samples) . . . . .	32
4.4	Welch PSD estimates (window length is 512 samples) . . . . .	32
4.5	Time-frequency resolution . . . . .	33
4.6	Some common wavelet functions . . . . .	34
4.7	CWT computed with a Mexican hat wavelet . . . . .	35
4.8	Dyadic analysis filter bank implementation of DWT . . . . .	37
4.9	Dyadic synthesis filter bank implementation . . . . .	37
4.10	DWT based decomposition for EEG signal . . . . .	38
4.11	The original EEG signal and the cleaned and reconstructed EEG signal . . . . .	39
5.1	The validator use case diagram . . . . .	44
5.2	The calculator use case diagram . . . . .	45
5.3	The class diagram for the tool . . . . .	46
5.4	The process flow for the validation and calculation . . . . .	47
5.5	Screenshot from an Excel sheet containing information on EDF files . . . . .	49
5.6	The visualization of the validity of the channels . . . . .	50
5.7	Annotations shown in EDFbrowser . . . . .	51
5.8	Theta-alpha power ratio for channels Fp1 and Fp2 . . . . .	56
5.9	Theta-alpha power ratio for channels C3 and C4 . . . . .	57
5.10	Theta-alpha power ratio for channels O1 and O2 . . . . .	57
5.11	Theta-alpha power ratio of combined records . . . . .	58

5.12 Channel specific ROC curves for TAPR measures between the different  
n-back levels . . . . . 60

## LIST OF TABLES

2.1	The characteristics of the cortical layers . . . . .	6
2.2	Brain wave frequency ranges . . . . .	9
5.1	EDF header and data record structure . . . . .	43
5.2	6-channel and 19-channel setups. . . . .	54
5.3	P-values for TAPR comparison between different n-back levels (left-sided Wilcoxon rank-sum test). . . . .	59
A.1	P-values related to TAPR comparison between the different n-back levels .	70

## LIST OF SYMBOLS AND ABBREVIATIONS

2D	Two-Dimensional
3D	Three-Dimensional
AA	Adaptive Automation
ACS	Autocorrelation Sequence
ADC	Analog-to-Digital Converter
AEP	Auditory Evoked Potential
AR	Autoregressive
ARSG	Action Real-time Strategy Gaming
ASCII	American Standard Code for Information Interchange
BAEP	Brainstem Auditory Evoked Potential
BSS	Blind Source Selection
CCA	Canonical Correlation Analysis
CES	Central Executive System
CFC	Cross-Frequency Coupling
CLI	Cognitive Load Index
CLT	Cognitive Load Theory
CNS	Central Neural System
CPU	Central Processing Unit
CWT	Continuous Wavelet Transform
DDR	Double Data Rate
DTFT	Discrete-Time Fourier Transform
DWT	Discrete Wavelet Transform
EC	Effective Connectivity
ECG	Electrocardiogram
EDF	European Data Format
EDFbrowser	An open source EDF viewer and toolbox
EEG	Electroencephalography
EMD	Empirical Mode Decomposition
EMG	Electromyogram

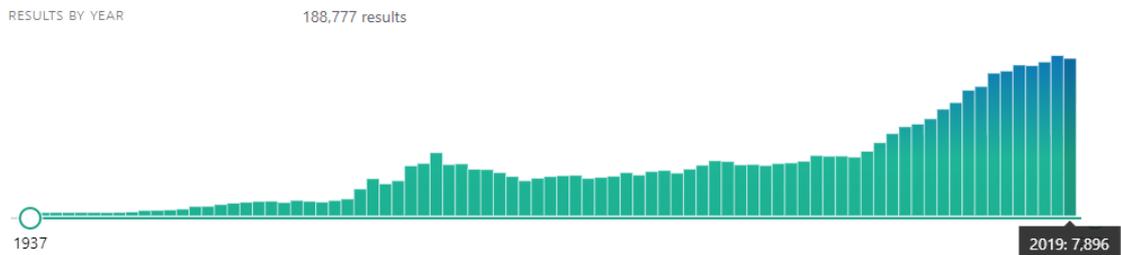
EOG	Electrooculogram
EP	Evoked Potential
ERP	Event Related Potential
ESC	Envelope to Signal Correlation
ESI	EEG Source Imaging
FC	Functional Connectivity
FT	Fourier Transform
HMD	Head-Mounted Display
IAF	Individual Alpha Frequency
ICA	Independent Component Analysis
JSON	JavaScript Object Notation
LLAEP	Long Latency Auditory Evoked Potential
MAD	Median Absolute Deviation
ML	Matlab
MLAEP	Middle Latency Auditory Evoked Potential
MRA	Multiresolution Analysis
MSC	Magnitude Squared Coherence
MVL	Mean Vector Length
NASA	National Aerospace and Space Administration
NASA-TLX	NASA-Task Load Index
NVGP	Non-Video Game Player
NVR	Non-Virtual Reality
OO	Object-Oriented
PAC	Phase-Amplitude Coupling
PCA	Principal Component Analysis
PLV	Phase-Locking Value
PSD	Power Spectral Density
PubMed	A free full-text archive of biomedical and life sciences journal literature at the U.S. National Institutes of Health's National Library of Medicine
ROC	Receiver Operating Characteristics
SCA	Sparse Component Analysis
SEP	Somatosensory Evoked Potential
SVM	Support Vector Machine

SWAT	Subjective Workload Assessment Technique
TAPR	Theta-Alpha Power Ratio
TF	Transition Frequency
USD	United States Dollar
VEP	Visual Evoked Potential
VGP	Video Game Player
VR	Virtual Reality
VSA	Visual Selective Attention
WLAN	Wireless Local Area Network
WT	Wavelet Transform
XML	Extensible Markup Language
$r$	Autocorrelation sequence
$\hat{r}$	Autocorrelation sequence estimate
$S_{xx}(\omega)$	Autospectra of signal $x$
$y^*(\cdot)$	Complex conjugate of $y(\cdot)$
$\hat{\Phi}_c$	Correlogram
$r_{xy}(\cdot)$	Cross-correlation function
$S_{xy}(\omega)$	Cross-spectrum of signals $x$ and $y$
$E[\cdot]$	Expectation operator
$f$	Frequency
$\mathbf{H}[\cdot]$	Hilbert transform
$\phi$	Instantaneous phase
$j$	Imaginary unit ( $\sqrt{-1}$ )
$\delta(\cdot)$	Kronecker delta function
$\mathbf{L}^2\{\mathbb{R}\}$	Space of square-integrable functions in the real-valued domain
$\gamma_{xy}^2(\omega)$	Magnitude squared coherence
$\psi$	Mother wavelet
$\omega$	Angular frequency
$\hat{\Phi}_p$	Periodogram
$\Phi$	Power spectral density
$\mathbb{R}$	Set of all real numbers
$\varphi(\cdot)$	Scaling function
$\hat{\Phi}_W$	Windowed periodogram

$\mathbb{Z}$  Set of all integer numbers

# 1 INTRODUCTION

The birth of the EEG as a scientific discipline can be dated back to the early 20<sup>th</sup> century when Hans Berger made his first EEG recording from the human brain. His invention on the EEG, that is, measuring the electrical activity of the brain, has been said to be "one of the most surprising, remarkable, and momentous developments in the history of clinical neurology" (Millet 2002). The EEG can be measured non-invasively with a portable and relatively low-cost equipment, which makes it a highly attractive tool that is widely used for clinical brain activity monitoring and measuring. It is a very essential and widely used tool also in the field of the brain related research, and it has gained even more interest during the recent years, which is reflected in the number of the EEG related publications (Figure 1.1).



**Figure 1.1.** Number of EEG related publications by year (in PubMed database).

In this thesis, the application of the EEG in measuring the mental load is discussed, in general terms and from the gaming perspective. Roughly said, the mental load is the load that the demands of task(s) being executed impose on a person's cognitive resources. The EEG provides a convenient means for monitoring and evaluating the mental load, without causing any significant interference to the subject under the study. EEG metrics for mental load assessment can be calculated from power spectrum, event related potential or brain connectivity measures. Gaming quite perfectly complements the EEG in the formation of a framework for studying the mental load as, in general, load imposing conditions in a game can be easily adjusted, and basic setups for such study environments are rather simple.

The main goals of this thesis are the following:

1. Give an overview on scientific research related to the feasibility of EEG analysis in the assessment of the effect of gaming on mental activity.

2. Develop a tool for analysing multichannel biosignal recordings of gaming sessions.
3. Conduct an analysis of a set of existing recordings, taken as a part of a course project by university students, for the detection of changes in the EEG during an n-back memory game.

The thesis is structured as follows. The EEG basics are introduced in Chapter 2. Mental load as such and in relation to the EEG and gaming is discussed in Chapter 3, including an overview on related research. Some selected methods for analysing the EEG, employed in the practical work of the thesis, are described in Chapter 4. The practical part, regarding the mental load evaluation in n-back gaming sessions, is considered in Chapter 5. Finally, the conclusion is given in Chapter 6.

## 2 ELECTROENCEPHALOGRAPHY

Richard Caton (1842-1926) can be regarded as the pioneer of EEG. In 1875 he published a short report on electrical phenomena of the brain. In his experiments he used rabbits and monkeys and was able to measure electric currents with a galvanometer having one electrode placed on the exposed cortex and one electrode on the skull surface. (Niedermeyer, Schomer and Lopes da Silva 2011)

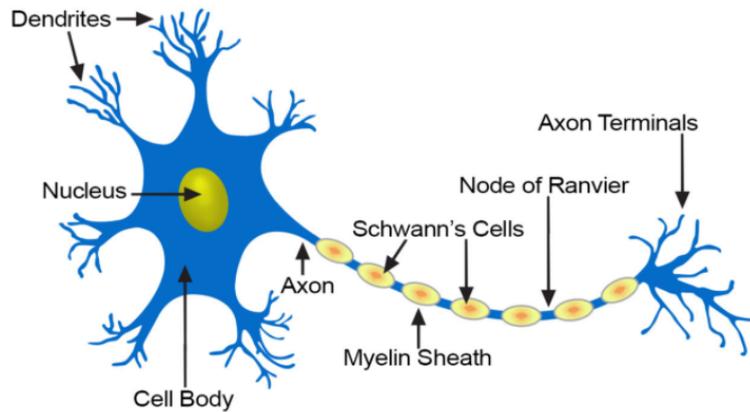
Hans Berger (1873-1941) was a German neuropsychiatrist who started studies on the human EEG. He recorded first EEG from the human brain in 1924 during a neurosurgery operation where the recording was made by connecting electrodes to the cortex. Later he developed a non-invasive recording technique where electrodes were attached to the scalp.

In 1929 Berger published his first paper on EEG (“Über das Elektrenkephalogramm des Menschen”) where he introduced alpha and beta waves, formerly also known as “Berger waves”. He discovered that brainwaves changed during activity and sleep. His experiments became the foundation of electroencephalography which has been since then an important non-invasive tool for better understanding the human brain and for diagnosing cerebral abnormalities. (Sörnmo and Laguna 2005)

### 2.1 Origin of the EEG

Most essential building blocks of the central neural system (CNS) are nerve cells (neurons) and glia cells that are located between neurons. EEG is mainly originated from electrical activities of populations of pyramidal neurons of the cerebral cortex (Mulert and Lemieux 2010). Each neuron consists of a cell body, several dendrites and a single axon (Figure 2.1). The dendrites provide a neuronal reception interface and the axon transmits electrical impulses (Byrne and Roberts 2004).

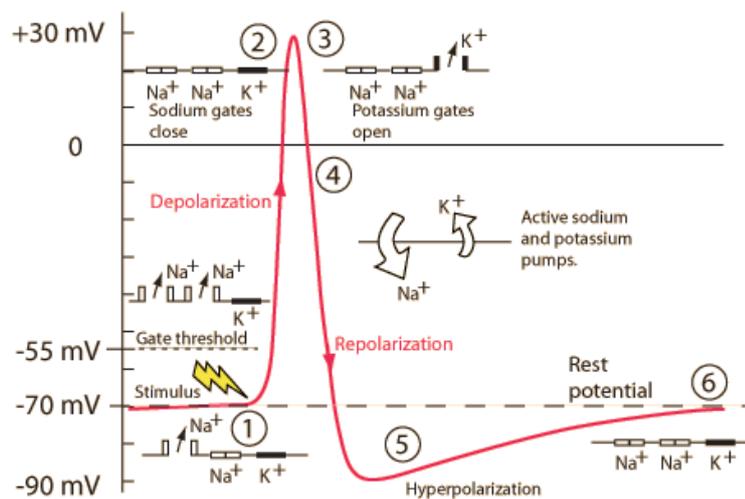
When a neuron is in resting state there are more sodium ions ( $\text{Na}^+$ ) outside than inside the cell and more potassium ions ( $\text{K}^+$ ) inside than outside the cell which results in negative potential over the cell membrane of approximately -70 mV. Due to a stimulus, sodium gates open and sodium ions start to flow into the cell which causes increase in the membrane potential. If the membrane potential reaches the threshold (commonly about -55 mV), the voltage-gated sodium channels open, and the cell membrane becomes depolarized and the action potential occurs which makes the cell to fire. The action potential is an “all or none” type phenomenon, it either happens with the full power or does not



**Figure 2.1.** Neuron <sup>1</sup>

happen at all, only depending on whether the threshold potential has been reached or not. After reaching the peak action potential, approximately +30 mV, sodium gates close and potassium gates open causing potassium ions to start flowing out of the cell. This repolarization phase ends when the membrane potential is lower than the resting potential, that is, the cell membrane becomes hyperpolarized. The neuron eventually reaches the resting potential via ion pump activity as potassium ions are pumped in and sodium ions are pumped out. (OpenStax 2016)

The different phases of the action potential are illustrated in Figure 2.2.



**Figure 2.2.** The phases of the action potential <sup>2</sup>

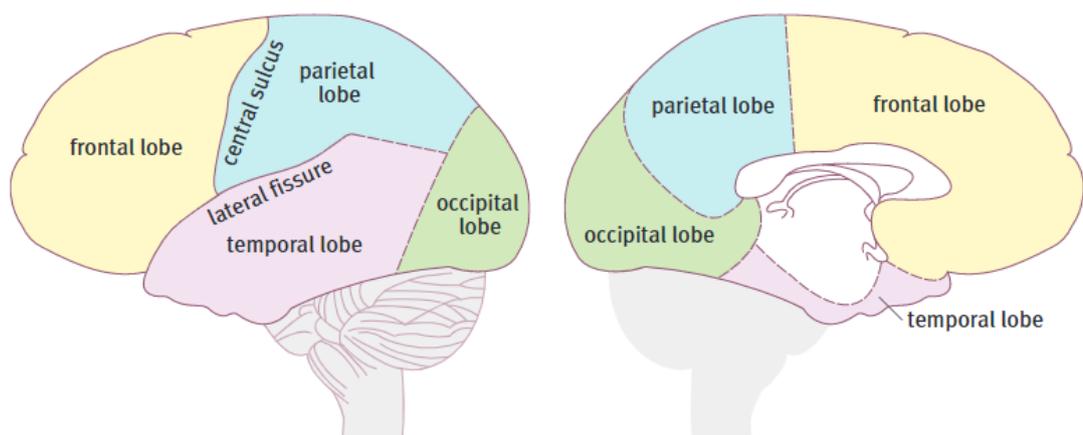
<sup>1</sup><https://training.seer.cancer.gov/anatomy/nervous/tissue.html>

<sup>2</sup><http://neurofeedbackalliance.org/eeg-electrophysiology>

## 2.2 Cerebral cortex

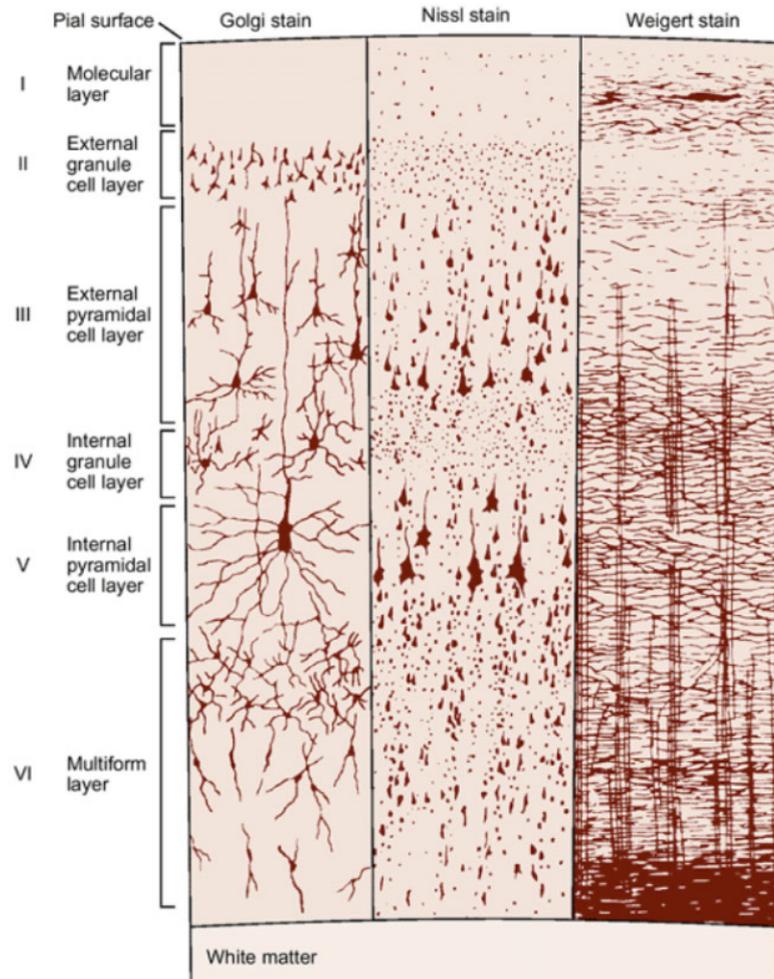
Human neurons and glia cells function quite similarly to those of other mammals. Also most of the genes of the human neural system and those of other mammals are similar to each other. The difference lies in how humans configure and use rather simple functional units to achieve complex behaviour, and this complex behaviour relies especially on the cerebral cortex (Watson, Kirkcaldie and Paxinos 2010). The cerebral cortex is a folded sheet of neurons and glia covering the rest of the forebrain. Its surface area and thickness are, respectively and approximately, 2500 cm<sup>2</sup> and 2-4 mm, and it contains 10<sup>9</sup> or more neurons (Jones 2010). In general terms, the cerebral cortex is responsible for analysing, predicting and responding to environmental events.

The cortical surface is segmented into four lobes named after their covering skull bones: frontal, parietal, temporal and occipital lobe (Figure 2.3). Even though this segmentation is an anatomical one, some functionalities can be localized on the same basis, at least roughly. For example, the visual processing is mainly executed in the occipital lobe.



**Figure 2.3.** Brain lobes (Watson, Kirkcaldie and Paxinos 2010)

With the aid of the microscopy and different staining techniques it can be observed that the neurons of the cortex are arranged into six distinct layers (Watson, Kirkcaldie and Paxinos 2010). These cortical layers are numbered from one to six starting from the surface (Figure 2.4). The characteristics of the cortical layers are described in Table 2.1. EEG is primarily originated from large vertically oriented pyramidal neurons located in the cortical layers III, V and VI (Olejniczak 2006).



**Figure 2.4.** Cortical layers (Kandel 2013)<sup>3</sup>

**Table 2.1.** The characteristics of the cortical layers (Koeppen and Stanton 2018)

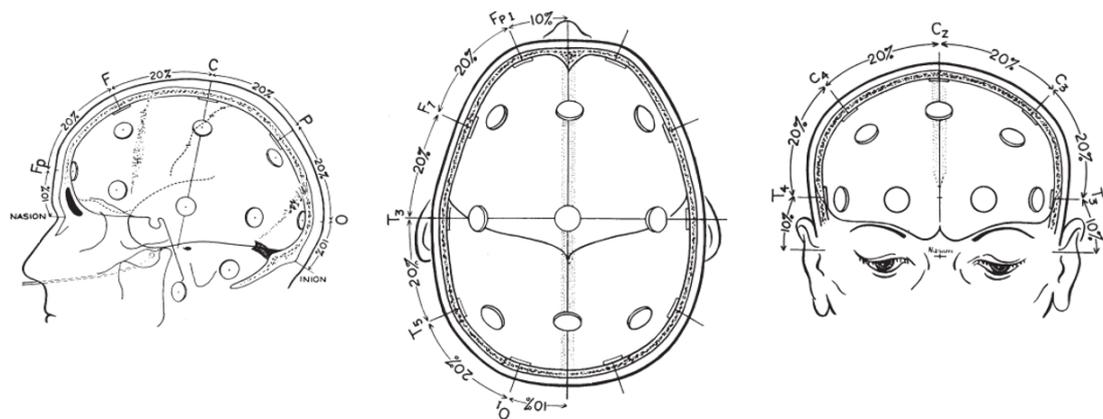
Layer	Description
I	Molecular layer containing few scattered neurons
II	External granular layer containing mostly stellate cells
III	External pyramidal layer containing predominantly small pyramidal neurons
IV	Internal granular layer containing different types of stellate and pyramidal cells
V	Internal pyramidal layer containing large pyramidal neurons
VI	Multiform layer containing few large pyramidal neurons

<sup>3</sup><https://neurones.co.uk/Neurosciences/Tutorials/M4/M.4.1CerebralCortex.html>

## 2.3 EEG measuring

Cerebral electrical signals recorded as EEG are oscillations composed of frequency components varying between 0.05 and 600 Hz. Higher frequency EEG signals have typically lower amplitude in comparison to lower frequency EEG signals. Furthermore, the spatial extent of the EEG signal is inversely proportional to its frequency, that is, higher frequency oscillations are fading faster than lower frequency oscillations when propagating within cortical volume. Due to these signal characteristics, the EEG recorded from the scalp is usually limited to the frequencies under about 30 Hz as higher frequency signals are effectively attenuated by the skull and underlying tissues and they are also interfered by non-cerebral sources like a muscle activity. (Ebersole, Husain and Nordli 2015)

Usually multiple electrodes are placed on different locations on the scalp to have better coverage and spatial resolution when measuring EEG. Quite commonly used electrode placement is according to the standard 10-20 measurement system. In this system, the scalp and thus the underlying cortex as well is divided into 10% and 20% sections between the nasion and inion, and between the left and right preauricular points, as illustrated in Figure 2.5 (Klem et al. 1999).



**Figure 2.5.** The 10-20 system. Adapted from (Klem et al. 1999).

Electrode positions, with the exception of the central position, are named after cortical lobes: Fronto polar (Fp), Frontal (F), Central (C), Temporal (T) and Occipital (O). The number indicates the lateral position of the electrode, odd numbers are used for the left hemisphere (Fp1, F3, F7, C3, T3, P3, T5, O1) and even numbers are used for the right hemisphere (Fp2, F4, F8, C4, T4, P4, T6, O2). The letter "z" in vertical electrodes Fz, Cz and Pz refers to "zero". (Klem et al. 1999)

The loose numbering used for the basic 10-20 layout allows to use the same numbering scheme for electrode positions in the extended layout. The modified combinatorial nomenclature (10-10 system) increases the number of electrodes from 21 up to 74. Both layouts are illustrated in Figure 2.6. In addition to the 10-20 and 10-10 systems also



In a modern EEG equipment, data can be transferred wirelessly, e.g. over WLAN, from the measuring device to the recording device.

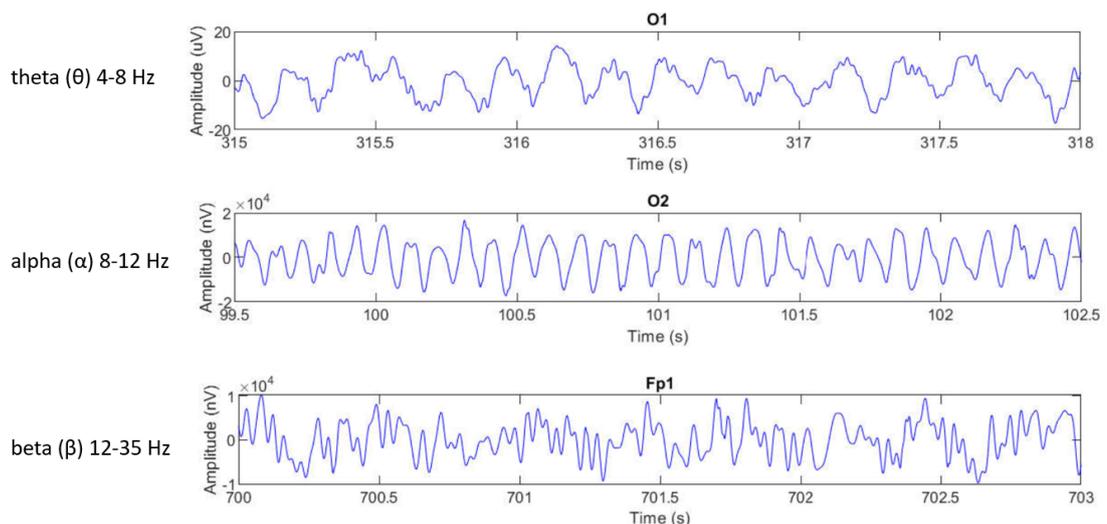
## 2.4 Interpretation of the brain waves

EEG interpretation can be based on a pattern recognition in the measured signal. Brain waves are commonly described by the rhythms corresponding to their frequency ranges. These rhythms are divided into five major categories named as delta, theta, alpha, beta and gamma (Table 2.2).

**Table 2.2.** Brain wave frequency ranges

Name	Frequency range (Hz)
delta	0.5 - 4
theta	4 - 8
alpha	8 - 12
beta	12 - 35
gamma	> 35

It should be noted that the brain wave frequency ranges are not strictly defined or standardized. Regardless of the absence of the standards, there is usually not that much variation between the different range definitions found in the literature, except for the gamma frequency range. However, the gamma waves are clinically not that significant compared to the other major brain waves. Examples of brain rhythms in an EEG signal are presented in Figure 2.7.



**Figure 2.7.** Examples of brain rhythms in an EEG signal.

Delta waves are commonly seen in the EEG and they can either indicate normal or abnormal brain function. Normal delta waves are typically found during deep sleep. (Nayak and Anilkumar 2020)

Theta waves are typically found in the EEG in case of drowsiness or at the early stages of sleep and also with cognitive tasks, where it may be related to the working memory. (Stern and Engel 2013)

Alpha waves are associated to relaxed wakefulness with eyes closed. They decrease or disappear with drowsiness, concentration, stimulation or visual fixation. A sudden loss of the alpha rhythm because of a visual or cognitive activity is called the alpha blocking (Stern and Engel 2013). The alpha rhythm is the dominant brain rhythm of adults (Klimesch 1999).

Beta waves were named by Hans Berger after having discovered and named the alpha waves. Originally beta waves were considered to have duration in the range of 30 to 40 ms that equals to 25 to 33 Hz in a rhythmic oscillation. Currently, beta waves are defined as oscillations having the frequency of about 12 Hz or higher. Frontal-central beta activity occurs most commonly with drowsiness or sleep onset, however, it has been also interpreted to indicate cognitive processing. Even though this latter interpretation may appear contradictory to the previous consideration of beta waves being related to drowsiness, it may be due to the increased visibility of beta waves when other EEG activities attenuate with alerted wakefulness. (Stern and Engel 2013)

Gamma waves are thought to be related to sensory perception, or cognitive activity, where different functional brain areas are connected. (Nayak and Anilkumar 2020)

## 2.5 Evoked potentials and event related potentials

Evoked potentials (EP) and event related potentials (ERP) are transient EEG waveforms generated by a stimulus, such as auditory or visual stimuli. The terms EP and ERP are quite often used interchangeably in the literature, but also a more exact terminology exists (Luck 2014) defining EP as an obligatory response to stimulus, and ERP, elicited by EP, as a non-obligatory potential related to cognitive activity. The common categories for EPs are visual evoked potential (VEP), auditory evoked potential (AEP) and somatosensory evoked potential (SEP) (Sörnmo and Laguna 2005). EPs have been found useful in practical applications, e.g. AEPs are used in monitoring the depth of anesthesia (Paulraj et al. 2015).

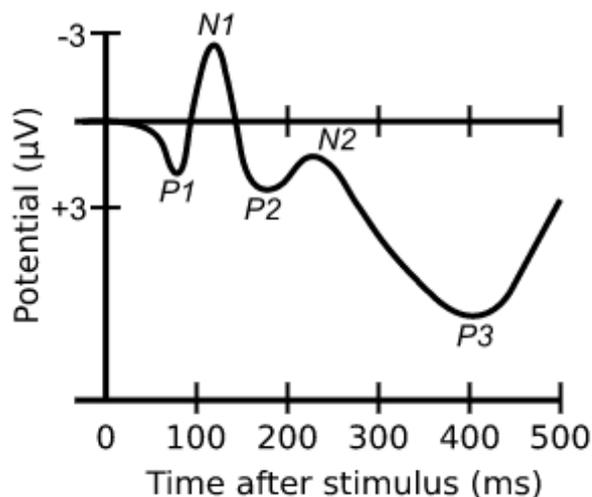
EPs/ERPs are time-locked to stimulus occurrence which means they are visible in EEG after a certain latency period. For example, the components of AEPs have the following approximate latencies:

- brainstem AEP (BAEP): 2-12 ms
- middle latency AEP (MLAEP): 12-50 ms
- long latency AEP (LLAEP): 50-600 ms.

BAEPs are related to the activation of the acoustic nerve, MLAEPs are considered to be generated in thalamic and cortical auditory structures, and LLAEPs reflect the activation

of the association areas of the cerebral cortex. (Beer et al. 1996)

EPs/ERPs have a considerably low amplitude in comparison to the background EEG activity and thus noise reduction, that is, the removal of the background EEG, is an important issue in EP/ERP analysis. Peak wave components of EPs/ERPs are named after their potentials, letter "P" is used for EPs/ERPs with positive potential and letter "N" is used for EPs/ERPs with negative potential. In addition, a number after the letter is used to reflect the latency in milliseconds after event occurrence. Alternatively, this postfix number can reflect the timely order of the component and in that case the number is less than ten. This short numbering convention is usually used for EPs only. For example, N2 refers to the second negative component and P300 refers to a positive peak occurring approximately 300 ms after a stimulus. As an exception to this naming rule, P3 is commonly used instead of P300. The naming convention for EP/ERP components is illustrated in Figure 2.8. (Sörnmo and Laguna 2005)



**Figure 2.8.** Visual example for naming convention of EP/ERP components (N1 = first negative, N2 = second negative, ..., P1 = first positive, ...). Note that potential on the potential axis decreases in upwards direction.

## 2.6 Artefacts

EEG artefacts can be roughly divided into two categories: physiological and technical. The most common sources for non-cerebral physiological artefacts are electrocardiogram (ECG), electrooculogram (EOG) and electromyogram (EMG). Technical artefacts can be caused by loose or poorly attached electrodes, or oscillations of 50/60 Hz originated from powerline, for example.

### 2.6.1 Sources of artefacts

ECG artefacts are caused by the electrical activity of the heart. Even though ECG amplitudes, when measured via EEG electrodes positioned on the scalp, are usually quite low in comparison to EEG, they may appear as considerable disturbance at certain EEG channels. But as heartbeat is a rhythmic and regularly occurring pattern, ECG artefacts can be recognized and removed quite reliably from EEG. However, spiky ECG artefacts can be incorrectly interpreted as epileptic activity in cases where ECG is hardly visible in EEG. ECG artefact removal is potentially easier and more reliable if ECG is measured, as a reference signal, simultaneously with the EEG measurement.

EOG artefacts are elicited by eye and eyelid movements. Eye movements can be rather easily confused with slow EEG oscillations, whereas eyelid movements, or blinks, are having higher frequency components (Sörnmo and Laguna 2005). For reference signal based artefact removal methods it is beneficial to have separate electrodes placed close to the eyes to measure EOG.

EMG reflects muscular activity. These artefacts spread over the frequency range from 0 to 500 Hz and are most dominant in 50-150 Hz range (Luca 2002). EMG artefacts are particularly challenging to be isolated, but on the other hand they may not pose that severe issue, as voluntary muscular activity is fairly controllable, at least in such experiments where the subjects are requested to remain steady and relaxed.

Technical, or external, sources of EEG artefacts are various. One common source is the movement of electrodes that may cause "electrode-pop" artefact that is visible in EEG as temporarily changed baseline level. Other possible cause of technical artefacts is insufficiently shielded cabling between the electrodes and the jackbox that can cause 50/60 Hz powerline interference in the EEG. As a common rule, if an artefact is visible in a single electrode it may indicate issues with the electrode or the related cabling, or issues with the plug-in channel to which the electrode is connected (Tatum 2014).

### 2.6.2 Artefact detection and removal

Artefacts in the EEG signal can be detected either manually by an expert, or automatically by a detection algorithm implemented for the purpose. The most straightforward way to process the identified artefacts is to simply discard the EEG segments containing artefacts, but with this simple approach it is possible that also essential EEG data is lost with discarded segments. Thus it is more desirable to develop and employ such artefact removal methods that attempt to retain the underlying EEG, to some extent at least, while removing the artefacts. It should be noted that the visual assessment is still needed to ensure that automated artefact removal methods perform appropriately.

Jiang, Bian and Tian (2019) present in their review a fairly comprehensive set of applicable methods of the artefact detection and removal for EEG signals. The wavelet transform is considered in Chapter 4.2, but other than that any detailed descriptions of these meth-

ods are not in the scope of this thesis. However, the methods presented in the review are listed here to give an idea of the variety of the available techniques:

- Regression based methods
- Wavelet transform (WT)
- Principal component analysis (PCA)
- Independent component analysis (ICA)
- Canonical correlation analysis (CCA)
- EEG source imaging (ESI)
- Empirical mode decomposition (EMD)
- Adaptive filtering
- Wiener filtering
- Sparse component analysis (SCA)
- EMD - Blind source selection (BSS)
- Wavelet - BSS
- BSS and support vector machine (SVM)

More detailed information on these methods can be found in the review itself and in the related references that are given in the review.

## **3 MENTAL LOAD, GAMING AND EEG**

Video games are usually associated with entertainment but they have also proven to be valuable media for the mental load related research. Video game playing provides a quite favourable environment for the EEG based mental load evaluation as a setup for such experiments is relatively simple, affordable and portable. Also the whole interactive environment can be created mostly in virtual terms that is an enabler for an easy variability, e.g. the difficulty level and detail level affecting the mental load are easily modifiable. Mental load and related indicators are discussed in the following sub-chapters, in general terms and from the gaming perspective. The cognitive development aspect of gaming is touched as well.

### **3.1 Mental workload**

Definition of mental workload is not unambiguous in the literature but various proposed definitions arbitrarily include and exclude defining variables (Acker et al. 2018). One practical attempt is to define mental workload as objective task demand imposed on a person (Dasari, Shou and Ding 2017). As stated by Radüntz (2017), the core of this concept is the relation between the demands of a task being executed and a person's cognitive capacity. Human information processing capabilities are limited by the amount of available cognitive resources, and it can be considered that the amount of task demands placed on these limited resources corresponds to mental workload (Radüntz 2017). When person's cognitive resources cannot meet the demand, that is, mental load is increasing excessively, it can result in degraded task performance at behavioural level, like increased response times and erroneous responses. It should be noted that too low mental load imposed by an actual task can degrade performance as well as it is possible that in that kind of situation the person's vigilance on the task is distracted more easily (Aricò et al. 2016).

#### **3.1.1 Cognitive load theory**

Cognitive load theory (CLT) approaches load definition from memory resources perspective. CLT is based on the cognitive architecture that is comprised of a working memory that has limited capacity for temporarily storing or processing new information, and a long-term memory that has virtually unlimited capacity (Antonenko et al. 2010; Sweller, Merrienboer and Paas 1998). Working memory limitations for dealing with new informa-

tion are a bottleneck as, according to Miller (1956), the working memory has the capacity for only  $7 \pm 2$  information elements. Furthermore, the number of the simultaneously handled information elements decreases when information also needs to be processed and not only to be remembered (Cowan 2001). Cognitive processes can increase the processing capacity with the help of earlier formatted patterns, categories and groupings that ease the handling of information elements (Miller 1956). Working memory is the key asset in multitasking (Preece, Sharp and Rogers 2015).

According to CLT, cognitive load is divided into three categories depending on its origin (Sweller 2010):

- *Intrinsic load* is induced by interaction between elements a task contains. If the number of interacting elements for the task is high, intrinsic load will be high as well. But if the task can be accomplished without the need to process elements dependently, intrinsic load will be low.
- *Extraneous load* is an unnecessary burden from the task accomplishing perspective, that is, it consumes resources without having any additional value. It may be induced by situation, environment, time pressure or other external sources.
- *Germane load* refers to cognitive resources that are required to handle intrinsic cognitive load. Germane load is placed for schema acquisition and automation. The more working memory resources are consumed by extraneous cognitive load, the less will be available to handle intrinsic cognitive load.

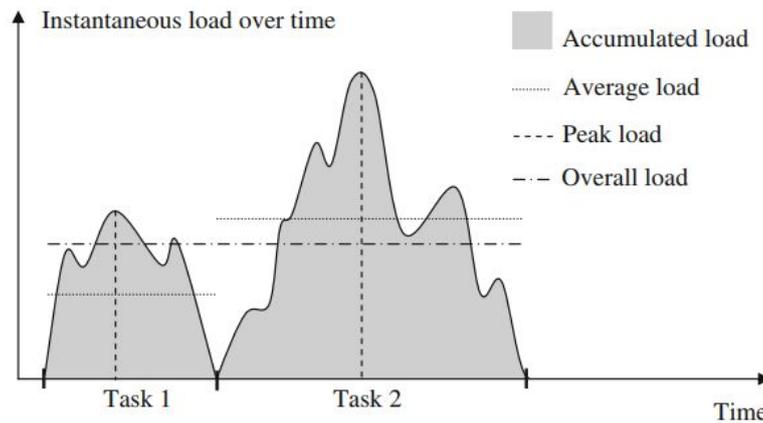
Total experienced cognitive load is a sum of intrinsic, extraneous and germane load.

In order to describe the mental workload more precisely, Xie and Salvendy (2000) proposed a framework for definitions covering the following terms: instantaneous workload, accumulated workload, average workload, peak workload and overall workload. *Instantaneous workload* represents measured or assessed load at a certain moment of time when a task is being performed. *Accumulated workload* is instantaneous workload summarized over a time period. *Average workload* is equal to the accumulated workload divided by the duration of time over which the accumulated workload was calculated. *Peak workload* is the maximum value of the instantaneous workload. *Overall workload* is a subjective experience of imposed mental workload during the task execution. The different types of mental workload are illustrated in Figure 3.1.

### 3.1.2 Mental workload measures

Mental workload measures can be divided into three main categories: subjective measures, performance measures and psychophysiological measures (Galy, Cariou and Mélan 2012).

For *subjective measures* two commonly used methods are NASA-Task load index (NASA-TLX) (Hart and Staveland 1988) and subjective workload assessment technique (SWAT) (Reid and Nygren 1988). NASA-TLX consists of six workload related factors: mental de-



**Figure 3.1.** Mental workload types (Antonenko et al. 2010).

mand, physical demand, temporal demand, frustration, effort and performance, as subjectively experienced. SWAT method is based on three subjective workload factors: time load, mental effort load and psychological stress load. NASA-TLX has been shown to be quite sensitive to small changes in intrinsic load (Collet, Averty and Dittmar 2009).

*Performance measures* are most typically based on response accuracy and response latency. There is a certain trade-off between these two indicators. When task difficulty is raised, latency for responses may increase due to increase in needed processing time. However, in most cases it could be considered that correct response is more important than low response latency if one of these needs to be sacrificed.

*Psychophysiological measures* are based on changes in human body at physiological or electrophysiological level. Heart rate and respiration frequency are examples of physiological variables associated with mental workload. Mental workload increases neural activity which in turn increases metabolic demand which is then probably the cause for increased heart rate and respiration (Hogervorst, Brouwer and Erp F. 2014). EEG is an example of a such electrophysiological variable that is reactive to mental load changes.

## 3.2 Mental workload induced by video game playing

Video game playing can be a cognitively demanding task simultaneously requiring several cognitive skills as attention, memory, decision making and alike.

### 3.2.1 Sources of mental load in gaming

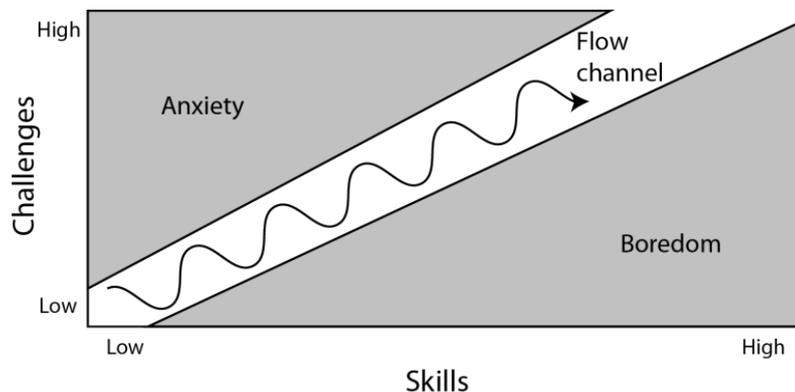
Mental load in gaming can be induced from several sources. Direct and the most loading source is typically game interactions, like player interacting with game environment by controlling in-game character or performing other activities that are directly connected to game objectives. Somewhat extraneous load can be caused by user interface that may, for example, hold information on game character's inventory and health, and other information that may or may not be relevant in real-time during game playing. User interface

related monitoring and interaction is potentially distracting player from actual playing activities. In online games social interactions with other players increase mental load (Ang, Zaphiris and Mahmood 2007). It is quite obvious that the amount of imposed mental load may depend on game genre and difficulty level. According to Allison and Polich (2008), an increase in difficulty level in a first-person shooter video game was reflected as an increase in the experienced mental workload.

Many modern commercial video games, especially first-person shooter games, are graphically highly advanced. Even though they are not quite photorealistic yet, the level of details can be remarkably high and also environmental physics are present, to some extent at least. All this increases virtual presence which has been shown to affect experienced mental load during video game playing, that is, the higher the virtual presence, the higher the mental load (Schrader and Bastiaens 2012).

### 3.2.2 Learning in games

In-game learned and developed gaming skills have been shown to have positive effect on cognitive load and on ability to cope with increased game challenges. Total cognitive load is decreased when earlier learned patterns can be utilized for processing imposed intrinsic load (Chen, Ou and Whittinghill 2015). The flow theory by Csikszentmihalyi (1975) is applicable to gaming as well. When gaming skills and game challenge are in balance, that is, cognitive resources are not overloaded nor underloaded, player is in the flow state, or within the flow channel, where motivation is retained and there is a drive for further development (Figure 3.2).



**Figure 3.2.** Flow channel. Adapted from (Csikszentmihalyi 1975)

If a player's gaming skills are not in balance with the challenges imposed by the game, i.e. the skills are significantly higher or lower than the demands, the player will drift out from the flow channel to boredom or anxiety zone, respectively, and eventually may even quit playing if this discomfort is prolonged.

Gaming may commonly be taken as activity for entertainment but gaming elements and gaming itself have been also harnessed for educational purposes. Gamified learning, or "serious gaming", is a topic of its own and as such will not be discussed in any deeper

details in this thesis. However, some quite interesting findings have been made related to mental workload and gamified learning that are worth to mention. Su (2015) showed in the structural equation modelling based study that cognitive load in gamified learning could be reduced via the proper choice of teaching strategies. Turan et al. (2016) had a more conservative outcome in an empirical study clearly showing higher cognitive load for the gamified learning group in comparison to the control group learning in the traditional way. Higher cognitive load in the latter study was considered to be caused by the competitive element related to gamification. Nevertheless, in both of these studies it was concluded that further research in the field of gamified learning is needed. It was also evident that cognitive load plays an important role in gamified learning and it is something that needs to be controlled in such learning environments. It is an intriguing thought if it would be possible to develop such gamified learning environments where the same learning results can be achieved with less cognitive load in comparison to traditional learning methods.

### **3.3 EEG indicators on mental workload**

EEG indicators, or measurable EEG metrics, for the mental load evaluation can be based on a power spectrum, event related potential or brain connectivity analysis. These indicators are discussed in the following sub-chapters and also some related studies and obtained results are shortly described.

#### **3.3.1 Power spectrum**

Power spectrum or power spectral density measures can be used to evaluate the signal energy in a given frequency band. These measures are basically the main foundation for mental load studies that assume the relation between the signal power in different EEG frequency bands and the mental load. Power spectral density as an EEG signal analysis tool is described in more detail in Chapter 4.1.

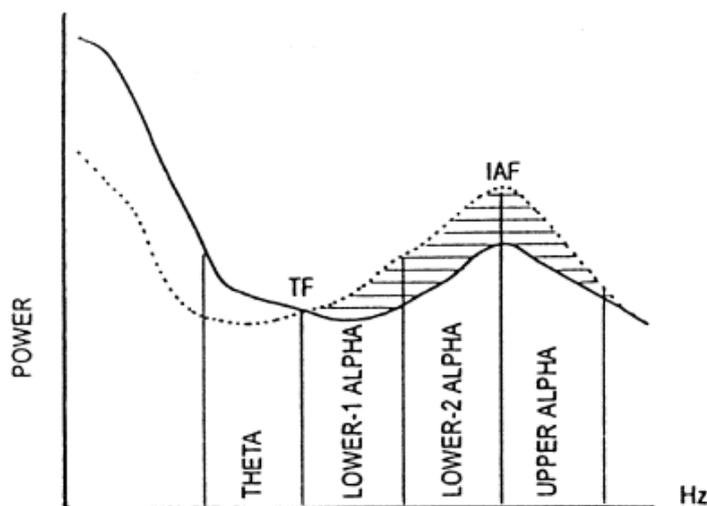
As discussed earlier, alpha waves are associated with relaxed wakefulness with eyes closed and they decrease or disappear with drowsiness, concentration, stimulation or visual fixation. This suggests that the measured signal power on alpha frequency range (8-12 Hz) could have a relation to mental workload, higher power indicating lower load. Activity on theta frequency band (4–8 Hz) is associated to activity on working memory. As load imposed on working memory is directly related to increased mental load, this suggests that theta power could reflect mental load, higher theta power indicating higher mental load.

The assumed decrease in alpha power, i.e. alpha desynchronization, and the increase in theta power, i.e. theta synchronization, due to increased mental load have been supported by several studies where mental load has been assessed or indicated with the aid of EEG power spectra, such as (Palomäki et al. 2012) and (Radüntz 2017). However,

there are also studies with partially opposite findings. Klimesch et al. (1999) found in the memory task related study that upper alpha band power increased during the task performance. This alpha synchronization was assumed to be caused by the high episodic short-term memory load that inhibited semantic long-term memory related processes that were found in other experiments to be related to upper alpha desynchronization.

Holm et al. (2009) presented in their study an EEG power spectrum based index for estimating the mental load. This index was calculated as the ratio of the absolute power of frontal theta activity to the absolute power of parietal alpha activity. It was found that the value of the index increased when task demands increased. As such, this finding is in line with the earlier findings, that alpha power decreases and theta power increases with increasing mental load. However, the particularly interesting outcome of this study was the proposed *scalar* index for mental load evaluation.

Klimesch (1999) proposed that instead of using fixed frequency bands for alpha and theta, individually adjusted frequency ranges should be used in order to avoid incorrect interpretations. Individual alpha and theta ranges are defined in a relation to the individual alpha frequency (IAF). Based on the experiments made by Klimesch it was found that IAF lies around 4 Hz above the transition frequency (TF) that is the intersection of the power spectrum measured in rest and the power spectrum measured during a task performance. Four frequency bands with the width of 2 Hz can be defined in the relation to IAF: theta [ $IAF - 6 \text{ Hz}, IAF - 4 \text{ Hz}$ ], lower-1 alpha [ $IAF - 4 \text{ Hz}, IAF - 2 \text{ Hz}$ ], lower-2 alpha [ $IAF - 2 \text{ Hz}, IAF$ ] and upper alpha [ $IAF, IAF + 2 \text{ Hz}$ ]. IAF, TF and alpha and theta bands are illustrated in Figure 3.3. It should be emphasized that these frequency ranges or offsets are not standardized which is the case for all the other major brain rhythms as well, as discussed in Chapter 2.4.



**Figure 3.3.** Individual alpha frequency (IAF), transition frequency (TF) and frequency bands. Dotted line is power spectrum measured in rest. Solid line is power spectrum measured during task performance. (Klimesch 1999)

### 3.3.2 Event related potential

ERPs (see Chapter 2.5) have been evinced as plausible indicator for mental load in studies where stimuli or mentally loading tasks are imposed in a timely manner. The scheduling and tracking of stimuli or task occurrences is of the essence here as ERPs are time-locked phenomena. It is possible to evaluate mental load of the primary task directly based on ERPs evoked by the primary task execution or indirectly based on ERPs evoked by the secondary task execution. Some studies on the ERP based mental load evaluation are shortly viewed in the following.

Watter, G. M. Geffen and L. B. Geffen (2001) showed in the n-back memory game related study that the P300 peak latency was constant during tasks, regardless of imposed mental load, but the P300 peak amplitude decreased when mental load increased due to increased memory load. Intuitively or quickly thought this decrease in the P300 peak amplitude might appear as opposite to the expectations. The decrease in the P300 amplitude was explained to be related to the dual nature of the n-back game. Memory requirements in the n-back game increase when difficulty, or the value of n, is increased whereas cognitive resources needed for the match making remain the same. Memorization and thus increase in the working memory allocation was assumed to take place already before the next stimulus onset. The P300 peak amplitude measured after the stimulus onset was reflecting cognitive load imposed mainly by the matching evaluation. This study is an example of ERP measurement used as the direct indicator of mental load of the primary task performance. The n-back memory game is discussed in more details in Chapter 3.6.

Allison and Polich (2008) applied auditory stimuli as secondary task during first-person shooter gaming sessions. Subjects were requested either to count or ignore infrequently elicited auditory probes that were external to the game. When game difficulty was increased the results showed decreased amplitudes for several ERP components (P2, N2 and P3) that were timely linked to auditory probes. It was inferred that this decrease in the amplitudes was due to the shortage of available cognitive resources as those were consumed increasingly by the more difficult primary task. The study represents an example of ERP measurement of the secondary task to evaluate mental load of the primary task.

### 3.3.3 Brain connectivity

Neurons and neural connections between them form a huge and complex network that is currently beyond any practical connectivity measurements at such microscopic level. The complexity is not only due to the huge number of neurons and connections but also that neural connections are formed and terminated dynamically (Friston 2011; Sakkalis 2011). A practical and feasible approach is to measure brain connectivity at macroscopic level between larger cerebral areas, e.g. between different cortical lobes. The brain connectivity can be divided into three subcategories: neuroanatomical, functional and

effective connectivity. According to Friston (2011) *neuroanatomical connectivity* can be considered as "fiber pathways tracking over extended regions of the brain, which are in accordance with general anatomical knowledge". However, neuroanatomical connections as such are not essential from the EEG measurement perspective. *Functional connectivity* (FC) is defined as temporal and statistically significant dependency between remote neurophysiological events (Friston 2011). *Effective connectivity* (EC) is defined as "the influence that one neural system exerts over another", and it describes the directional interactions among brain regions (Sakkalis 2011).

Cross-frequency coupling (CFC) is a phenomenon that is highly important in the EEG based brain connectivity analysis. Basically CFC can reveal temporal interaction or linkage between oscillations of different frequencies or frequency bands. From the brain connectivity analysis point of view this means in simplicity that the stronger the coupling is between the oscillations in different brain areas, the stronger these areas are connected to each other, in functional sense. Any combination of frequency, amplitude or phase can be coupled. For example, in the phase-amplitude coupling (PAC) the phase of one frequency component modulates the amplitude of another frequency component in the same or different signal. Various measures are available for evaluating the strength of the coupling, such as phase-locking value (PLV) (Lachaux et al. 1999; Vanhatalo et al. 2004), envelope to signal correlation (ESC) (Bruns and Eckhorn 2004) and mean vector length (MVL) (Canolty et al. 2006). As an example of a CFC application in the load evaluation, Gong et al. (2019) studied the relation between the mental load imposed by action real-time strategy gaming (ARSG) and the brain connectivity, using PLV as CFC measure. Their finding was that during ARSG session connections between the temporal and the central area were strengthened in comparison to the resting condition.

Coherence is also widely used and important measure for the EEG based brain connectivity analysis. It measures synchronization between two signals based on the phase difference. Coherence takes higher values when there is less variation in the phase difference (Srinivasan et al. 2007). As an example of using coherence in the connectivity based mental load evaluation, Payne and Kounios (2008) applied the wavelet transformation to calculate temporal coherence between the brain regions during altering mental load. It was found that when memory load was increased, the coherence of theta frequency band between frontal-midline and left temporal-parietal regions, as well as the coherence of the alpha frequency band between midline parietal and left temporal/parietal regions, were increased.

### **3.3.4 Adjustment of automation based on mental load assessment**

One prominent application for the aforementioned mental load indicators is adaptive automation (AA) (Rouse 1988). Basically AA means the adjustment of the automation level in human-machine systems according to the mental state of an operator, that is, when

the mental load increases excessively, the automation level is raised in order to lower or stabilize the load. Similar kind of methodology could be applied to video gaming as well, to keep a player engaged and motivated by adjusting game conditions in such manner that cognitive resources are not overloaded nor underloaded. It is obvious that the mental load for triggering the automated adaptation should be evaluated or classified in an automated way and also preferably in near real-time during the task execution.

Aricò et al. (2016) studied automated mental load classification in an experiment where air traffic controllers were performing control tasks in a simulated environment. Linear classifier algorithms were applied to conduct the binary classification of the mental load ("high" and "low" load). These separate load classes were used to trigger AA. EEG features used for classification were alpha and theta power spectral density measurements where the frequency ranges were adjusted with the IAF (see Chapter 3.3.1). The achieved mean classification accuracy in this study was approximately 75% (with 10% standard deviation).

Roy et al. (2016) used also a linear binary classifier in an experiment where subjects performed the Multi-Attribute Task Battery – II (Comstock and Arnegard 1992). In this experiment the used EEG feature was ERP evoked by an external audio probe. Similarly as in the already described study by Allison and Polich (2008), these ERPs evoked by external audio probes were modulated by cognitive load and thus ERP measurements could be used as workload indicators. The achieved mean classification accuracy was approximately 90% (with 10% standard deviation).

Both of the studies presented above promoted the feasibility of the EEG measurement based implementation for automated mental load classifiers. They could be also considered encouraging for further studies to apply and evaluate other classification methods and algorithms, e.g. in an attempt to develop classifiers with finer granularity.

### **3.4 Gaming and cognitive skills**

As discussed, video gaming is a source for mental load but there is also a cognitive development aspect in gaming that has gained some interest in the field of cognitive research. Video game playing might be intuitively thought as potentially beneficial for developing cognitive skills, that is, for increasing cognitive performance. However this is not that evident when results from various studies in this field are interpreted. Improvements in visual spatial cognition and attention can be usually more robustly related to gaming but, e.g. for memory and general cognition, results are more complicated and general conclusions cannot be made (Kühn, Gallinat and Mascherek 2019).

Even though various research results on the assumed positive influence of gaming on cognitive skills may appear controversial, the visual selective attention (VSA) performance seems quite clearly to benefit from gaming according to several studies. VSA and two related studies, as an example and to demonstrate usable methods for VSA performance assessment, are shortly described in the following.

VSA means the cognitive ability to extract essential information from ample visual inputs. The processing capacity of the human visual system would become serious bottleneck if all visual information within the field of view had to be constantly processed. This information overflow is blocked by VSA so that only a selection of visual inputs are allowed to reach high-level cortical processing. (Zhang and Lin 2013)

Matern, Westhuizen and Mostert (2019) studied VSA performance between the groups of video game players (VGP) and non-video game players (NVGP). A computerized version of Stroop task was used to measure VSA performance of both groups. Stroop task is a widely used method for attentional measures (MacLeod 1992). The study results showed that there was significant difference in VSA performance between VGPs and NVGPs, in favour of VGPs.

Green and Bavelier (2003) studied the relation between video gaming and VSA performance with various methods. Firstly, a flanker compatibility test (Lavie and Cox 2016) was applied to study the visual attention capacity. Secondly, an enumeration task was performed to evaluate the capacity based on how many briefly flashed items on a display a subject could numerate without actually counting them one by one. Thirdly, the visual attention was tested over different viewing ranges. Lastly, an attentional blink test (Raymond, Shapiro and Arnell 1992) was performed. This study was quite comprehensive from the test repertoire point of view and all of the performed tests showed better VSA performance for VGP group in comparison to NVGP group.

### **3.5 EEG in virtual reality environment**

Virtual reality (VR) gaming is a rapidly expanding market. In 2019 it was valued at USD 7.7 billion and it is predicted to reach USD 42.50 billion by 2025 (Research and Markets 2020). VR gaming has also gained interest in the field of EEG related research. The sensation of immersion and presence is stronger in VR compared to traditional presentation methods like a conventional 2D display (Buttussi and Chittaro 2018). The increased sense of immersion may be measured objectively by EEG (Tauscher et al. 2019).

#### **3.5.1 EEG measuring during VR session**

VR gaming requires special equipment compared to the traditional gaming where a 2D display provides visual interface and a gamepad, mouse and keyboard are the most typical control devices. A head-mounted display (HMD) that is strapped to a person's head is a common consumer grade display equipment for VR setups. A hemispherical video projection system, known as a dome, is a VR setup where no additional wearable display is needed. It is a much higher scale setup compared to the HMD setup as the diameter of the dome installation can be several meters.

HMD usage poses some potential impediments when EEG measurements are to be performed during a VR session. HMD and EEG electrodes, or EEG electrode cap, may

obstruct each other which can cause e.g. cable twisting and displacement of the electrodes which in turn may distort EEG measurements. Tauscher et al. (2019) studied the signal quality of EEG measurements when an EEG electrode cap and HMD were mounted simultaneously on subjects. Their finding was that the combination of HMD and EEG electrode cap for performing EEG measurements is feasible but without any modifications this combined equipment was experienced as non-comfortable by some subjects and also EEG signal distortions were introduced due to pressure and cable twisting. With the custom modification of the HMD strap, both of these inconveniences could be mitigated to some extent. The conclusion was that further study is needed for the proper mounting of HMD and EEG cap simultaneously. An ideal solution could be that both the HMD and EEG electrodes are integrated into one single device.

### **3.5.2 Mental load evaluation in VR environment**

With VR technology it is possible to enhance computerized environments for experiments to be more lifelike or at least to add lifelike elements, such as 3D visuals that are reactive to head movements, which increase the presence and immersion. This certainly makes a VR environment interesting, and potentially different compared to a non-VR (NVR) environment, from the mental load evaluation point of view, as demonstrated by two examples of VR related mental load studies described in the following.

Tremmel et al. (2019) conducted a study where mental load was evaluated in a VR environment. HMD and EEG electrode cap were mounted to participants who performed n-back memory game tasks at different difficulty levels. In a virtual version of the n-back game, different coloured balls were used as stimuli. Power spectrum analysis showed that the discrimination of several mental load levels was possible based on the EEG measurements in the VR environment. Interestingly, fronto-parietal alpha and theta activations were not that constantly seen as they usually are in the corresponding mental load studies performed in a NVR environment. Also, beta and gamma activity was more prominent than in the related studies in NVR environments. So this study showed not only that the mental load evaluation is feasible in a VR environment, but also that the frequency band activity is different compared to the usual results obtained in a NVR environment. This difference was assumed to result from body movements during the test and, somewhat more interestingly, also from rich visual input in the VR environment.

Dan and Reiner (2017) applied cognitive load index (CLI, Holm et al. (2009)) in their study to compare mental workloads in VR and NVR environments. The content in both environments was basically the same, a person giving instructions for a paper-folding task. In the NVR environment the video of the instructor was shown on a 2D display whereas in the VR environment the realistic digital avatar of the instructor was shown in 3D. Based on the CLI measurements, the mental load was found higher in the NVR environment than in the VR environment. One might assume quite the opposite that the mental load would be lower in a 2D NVR environment due to less visual cues, like depth,

to be processed in comparison to a 3D VR environment. However, if the reduced visual cues are informative, the cognitive processes must be able to perform with increased ambiguity which increases the mental load.

### **3.6 N-back memory game and EEG**

The n-back memory game was introduced by Kirchner (1958) and it has been a widely used tool since then in numerous cognitive performance related studies, especially in those related to working memory performance and mental load.

In the n-back memory game a person is presented sequential stimuli that are perceptually identifiable, e.g. visual stimuli consisting of the set of alphanumeric characters or auditory stimuli consisting of the set of auditory probes. The person playing the game is requested to respond, e.g. by pressing a button, when the current stimulus being presented was presented also n items back. The matching stimulus is called the target whereas a non-matching stimulus is called a non-target. The outcome that reflects the task performance, like the number of correct (targets) and incorrect (non-targets) responses, and the response delay, can be then used by an experimenter to further analyse the cognitive performance. The game difficulty is controlled with the value of n, i.e. the higher the value is, the more difficult the game becomes as it imposes higher load on the working memory. When the value of n is zero, the stimulus used as a target is predefined.

One common variation of the n-back game is the dual n-back where two different sequences of stimuli are presented simultaneously. The dual task is considered an effective tool to study the central executive system (CES). CES is considered to be responsible for the allocation and the coordination of attentional resources and in the dual task attentional resources are divided to different simultaneous control processes (Jaeggi et al. 2003).

In addition to its value as a tool for cognition related studies, the n-back game has been also promoted to have positive impacts on working memory performance. Pergher et al. (2018) made a study where they showed, based on EEG measurements for P300 component of event related potential, that n-back training improved working memory performance and it also improved attention and fluid intelligence.

While n-back task has strong face validity, that is, in various experiments it has subjectively seemed to measure what it is meant for, its construct validity has been argued (Kane et al. 2007). Gajewski et al. (2018) suggested that the effect of distinct cognitive functions related to n-back performance varies with aging. However, from the mental load point of view, in case it is not relevant how the cognitive load is distributed over different cognitive functions, these possible nuances may not be that meaningful in that context.

From EEG and the mental load measurement perspective the n-back game is a highly convenient tool for experiments as mental loading can be increased or decreased easily and in a controllable manner (by increasing or decreasing the value of n), stimulus can be invoked temporally accurate and, in general, it takes quite low effort to be employed

and learned. As a such right tool for the job, the n-back memory game has been used in several EEG based mental load studies (e.g., Watter, G. M. Geffen and L. B. Geffen (2001); Pesonen, Hämäläinen and Krause (2007); Palomäki et al. (2012); Tremmel et al. (2019)) of which some have been already discussed in the earlier chapters dealing with EEG indicators for the mental load.

## 4 EEG ANALYSIS METHODS

In the following sub-chapters some methods and techniques used commonly in EEG analysis are introduced. Even though this list is very far from being exhaustive, it includes methods that have been successfully applied to the EEG analysis in the field of cognitive performance and mental load related research, as discussed earlier, and which were partially also seen applicable to the experimental part of this thesis (Chapter 5).

### 4.1 Power Spectral Density

Power spectral density (PSD) is a measure that describes how the signal power is distributed over the frequency. For PSD estimation there are two approaches, a parametric and a non-parametric. The parametric approach is based on the assumption that a signal is generated as an output of a linear system that is fed with white noise. If there is enough information available on the measured signal, this assumed system can be modeled via parameterization. An autoregressive (AR) model is quite commonly used for the parametric PSD estimation. Non-parametric methods have no assumptions on the underlying model but they are based on the discrete-time Fourier transform (DTFT) of the sampled signal.

In the following a short presentation regarding non-parametric methods is given (parametric methods are not considered here). First some backgrounds are briefly described, followed by a basic introduction of non-parametric PSD methods. It should be noted that any proofs for the presented equations are out of the scope of this thesis. The interested reader may refer, e.g. to Stoica and Moses (2005) for further information, including parametric methods as well.

Let us assume that a deterministic discrete-time sequence  $\{y(t) \mid t \in \mathbb{Z}\}$  has finite energy, that is:

$$\sum_{t=-\infty}^{\infty} |y(t)|^2 < \infty. \quad (4.1)$$

Then, a DTFT is defined for the sequence as follows:

$$Y(e^{j\omega}) = \sum_{t=-\infty}^{\infty} y(t)e^{-j\omega t}, \quad (4.2)$$

where  $\omega$  is the angular frequency measured in radians per sampling interval.

Energy spectral density for this deterministic discrete-time sequence is given as:

$$S(e^{j\omega}) = |Y(e^{j\omega})|^2. \quad (4.3)$$

However, most real-world signals, like EEG, are non-deterministic by nature which means it cannot be determined exactly how the signal varies outside the measured period(s). In the discrete-time domain these non-deterministic signals, or random signals, can be considered as sequences of random variables. For the realizations of discrete-time random signals the inequality (4.1) does not hold due to infinite duration and the fact that, in general, random signals do not decay when time goes to the infinity, and thus such signals do not possess DTFT either. Even though random signals do not have finite signal energy, they usually do have finite average signal power.

PSD for discrete-time signals can be defined as the DTFT of the autocorrelation sequence (ACS). The ACS for  $y(t)$  is defined as follows:

$$r(k) = E[y(t)y^*(t-k)]. \quad (4.4)$$

In the above equation  $E[\cdot]$  denotes an expectation operator and  $y^*(\cdot)$  denotes a complex conjugate of  $y(\cdot)$ .

The PSD is then defined as the DTFT of the ACS:

$$\Phi(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r(k)e^{-j\omega k}. \quad (4.5)$$

Another definition for PSD, that is based on the assumption that the ACS converges, is as follows:

$$\Phi(e^{j\omega}) = \lim_{N \rightarrow \infty} E \left[ \frac{1}{N} \left| \sum_{t=1}^N y(t)e^{-j\omega t} \right|^2 \right], \quad (4.6)$$

where  $N$  is the number of signal samples.

Theoretically PSD can be obtained by either of the above definitions (4.5 or 4.6) but in practice the observed number of discrete-time samples for the measured signal is limited and thus only an estimate for PSD can be given. Periodogram and correlogram are two common non-parametric PSD estimators derived from (4.6) and (4.5), respectively.

Periodogram is defined as follows:

$$\hat{\Phi}_p(e^{j\omega}) = \frac{1}{N} \left| \sum_{t=1}^N y(t)e^{-j\omega t} \right|^2. \quad (4.7)$$

Correlogram is defined as follows:

$$\hat{\Phi}_c(e^{j\omega}) = \sum_{k=-(N-1)}^{N-1} \hat{r}(k)e^{-j\omega k}. \quad (4.8)$$

The term  $\hat{r}(k)$  in (4.8) denotes an ACS estimate, for which the standard biased version is usually used:

$$\hat{r}(k) = \frac{1}{N} \sum_{t=k+1}^N y(t)y^*(t-k), \quad 0 \leq k \leq N-1. \quad (4.9)$$

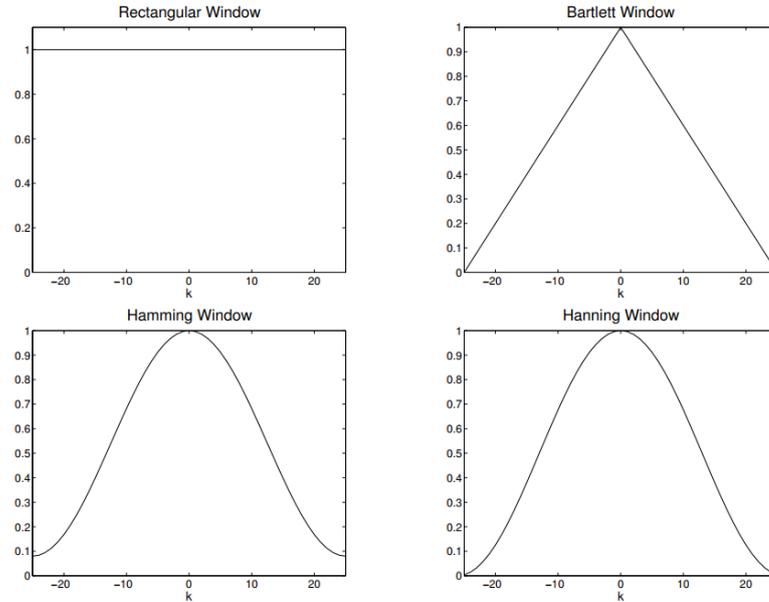
Both the periodogram and correlogram provide high frequency resolution but the problem is that the variance for the estimated PSD is relatively high, and it remains high even if the number of the samples is increased. This high and non-controllable variance makes the periodogram and correlogram poor PSD estimators as such, but they form a solid base for improved non-parametric PSD estimators that have smaller variance, at the cost of lower frequency resolution.

Windowing is a characteristic feature for the periodogram based PSD estimators. The windowed periodogram is defined as follows:

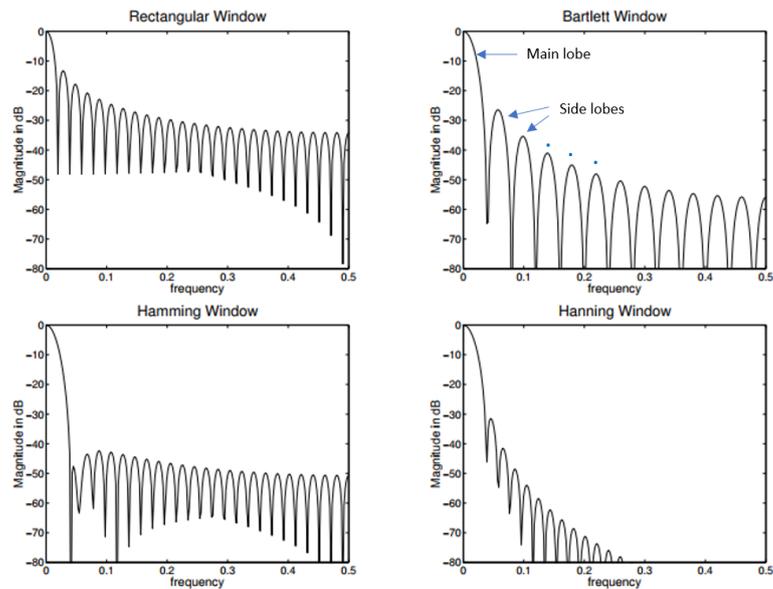
$$\hat{\Phi}_W(e^{j\omega}) = \frac{1}{N} \left| \sum_{t=1}^N v(t)y(t)e^{-j\omega t} \right|^2, \quad (4.10)$$

where the weighting sequence  $v(t)$  may be called a *temporal window*, or alternatively, a taper.

Some common window functions and their respective frequency responses are illustrated in Figures 4.1 and 4.2. Two major concerns, when selecting a window function to be applied, are the width of the main lobe and the magnitude of the side lobes in the frequency response of the window. The impact of the main lobe is that it smooths, or smears, the estimated spectrum. Depending on the width of the main lobe it may be that peaks close to each other in the power spectrum are estimated as one broader peak, that is, the wider the main lobe, the lower the spectral resolution. The effect of the side lobes is that they transfer, or leak, power from the frequency bands to adjacent bands that may contain less, or no power at all.



**Figure 4.1.** Window functions. Adapted from (Stoica and Moses 2005).



**Figure 4.2.** Frequency responses of window functions. Adapted from (Stoica and Moses 2005).

The Welch method is a commonly used non-parametric PSD estimate for EEG signals. The procedure for computing the Welch estimate is as follows:

1. The signal of  $N$  samples is divided into segments of  $M$  samples. The segments are overlapped by  $M - K$  samples ( $0 < K \leq M$ ). The total number of the segments is then

$$S = \left\lceil \frac{N - M}{K} \right\rceil + 1$$

The  $i$ th segment  $y_i(k)$  is obtained as:

$$y_i(k) = y((i - 1)K + k), \quad (4.11)$$

where  $i = 1..S$ ,  $k = 1..M$ .

2. The windowed periodogram (see (4.10)) for  $y_i(t)$  is calculated as:

$$\hat{\Phi}_i(e^{j\omega}) = \frac{1}{MP} \left| \sum_{t=1}^M v(t)y_i(t)e^{-j\omega t} \right|^2, \quad (4.12)$$

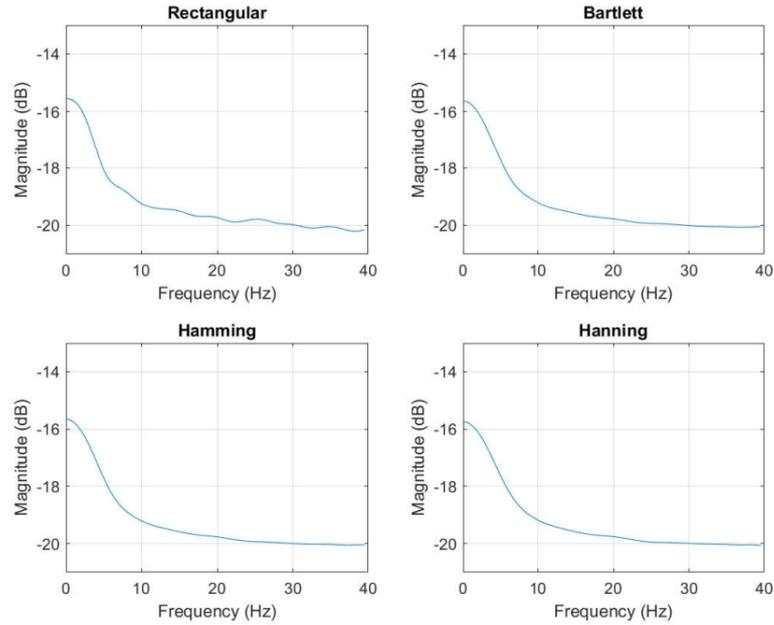
where  $P$  is defined as the power of the window  $v(t)$ :

$$P = \frac{1}{M} |v(t)|^2. \quad (4.13)$$

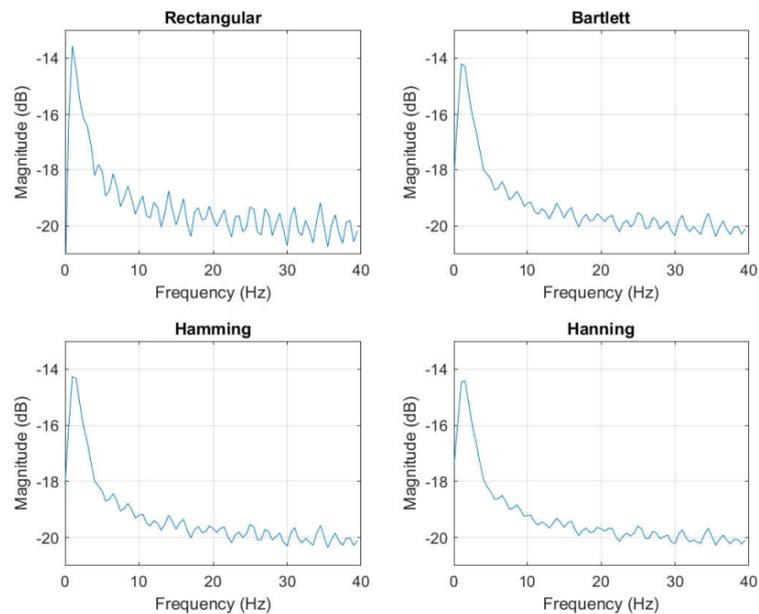
3. The Welch PSD estimate is then calculated as an average of the windowed periodograms:

$$\hat{\Phi}_W(e^{j\omega}) = \frac{1}{S} \sum_{i=1}^S \hat{\phi}_i(e^{j\omega}). \quad (4.14)$$

As an example of the effect of applying different window functions and different window length for PSD estimates, Welch PSD estimates for a measured EEG channel (C3) are shown in Figures 4.3 and 4.4. The sampling rate for the EEG signal is 500 Hz and the length of the sampled signal is 420500 samples. It can be quite clearly observed that the PSD estimate calculated with the rectangular window makes a notable difference to the PSD estimates calculated with the other window functions. This difference is most notable in Figure 4.3 where the PSD estimates with the other window functions appear to have less fluctuation. This is as expected due to the narrower main lobe of the frequency response of the rectangular window function, resulting in the better spectral resolution, in comparison to the other window functions. It is also clearly visible that the spectral resolution increases in the PSD estimates with the increasing window size.



**Figure 4.3.** Welch PSD estimates for the same EEG signal. Different window functions are applied. The length of the sampled signal is 420500 samples. The windows are 128 samples wide and overlapping by 50%.



**Figure 4.4.** Welch PSD estimates for the same EEG signal. Different window functions are applied. The length of the sampled signal is 420500 samples. The windows are 512 samples wide and overlapping by 50%.

## 4.2 Wavelet transform

In the following, an overview of wavelets and the wavelet transform (WT) is given, with some more general background information. The discrete wavelet transform (DWT) is considered in a relation to the multiresolution analysis (MRA) as from this viewpoint DWT

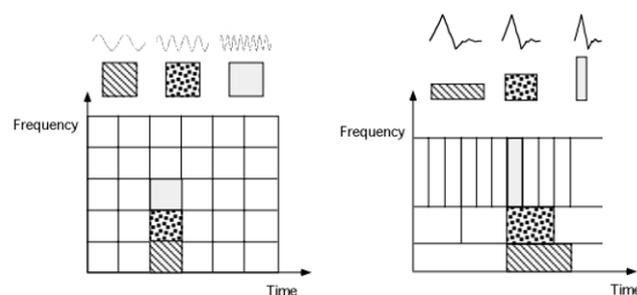
provides a very prominent tool for the EEG analysis. For the more detailed information and for the rigorous mathematical details and proofs, the interested reader may refer to e.g. Haddad and Akansu (2000), S. Mallat and Peyre (2008) and S. G. Mallat (1989).

While regular Fourier transform (FT) has an excellent frequency resolution, it totally lacks the temporal resolution. Thus FT is an inappropriate tool for analysing non-stationary signals, if it is of importance to timely localize the frequencies in the signal. Short time Fourier transform (STFT) applies a windowing technique that effectively divides the signal in shorter segments and thus enables time-frequency localization with the resolution depending on the window length. However, one shortcoming of STFT is that the window has fixed length, that is, the resolution is the same for all frequencies. WT provides an alternative for STFT. The basis functions of WT, called wavelets, have finite support, in contrast to the basis functions of FT, sine and cosine functions, that have infinite support. The fundamental difference of WT in comparison to STFT is that WT uses windows, or wavelets, of variable length, i.e. short windows for high frequencies and long windows for low frequencies.

Heisenberg inequality states an universal physical dependency between the time and frequency resolution:

$$\Delta t \Delta f \geq \frac{1}{4\pi}. \quad (4.15)$$

That is, it is not possible to have an arbitrarily high resolution simultaneously for both the time and frequency, but the higher the time resolution is, the lower the frequency resolution, and vice versa. Time-frequency resolution for STFT and WT is illustrated in Figure 4.5. As it can be seen, for STFT the resolution remains constant for both the frequency and time, whereas for WT the resolution is adaptive, that is, higher frequency resolution and lower time resolution for low frequencies, and lower frequency resolution and higher time resolution for high frequencies. This kind of adaptive nature of WT is best suitable for analysing signals that have short bursts of high frequencies and low frequencies of longer duration, which actually quite often is the case for practical signals, like EEG.



**Figure 4.5.** Time-frequency resolution for STFT (left) and WT (right). Adapted from (Graps 1995).

Wavelet family is defined as follows:

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right), \quad (4.16)$$

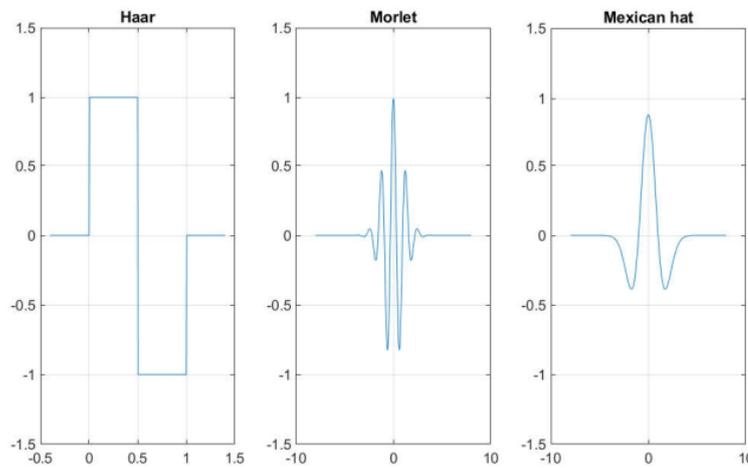
where  $\psi(\cdot)$  is called the *mother wavelet* from which child wavelets are generated by scaling (by parameter  $a$ ) and shifting (by parameter  $b$ ).

Wavelet functions have to satisfy the admissibility condition, that is:

$$\int_0^\infty \frac{|\Psi(\Omega)|^2}{\Omega} d\Omega < \infty, \quad (4.17)$$

where  $\Psi(\Omega)$  is the FT of the wavelet function  $\psi(t)$ .

Some common wavelet functions are illustrated in Figure 4.6.



**Figure 4.6.** Examples of some common wavelet functions (Haar, Morlet and Mexican hat wavelet).

Continuous wavelet transform (CWT) for a continuous function, or signal,  $x(t)$  is then defined as follows:

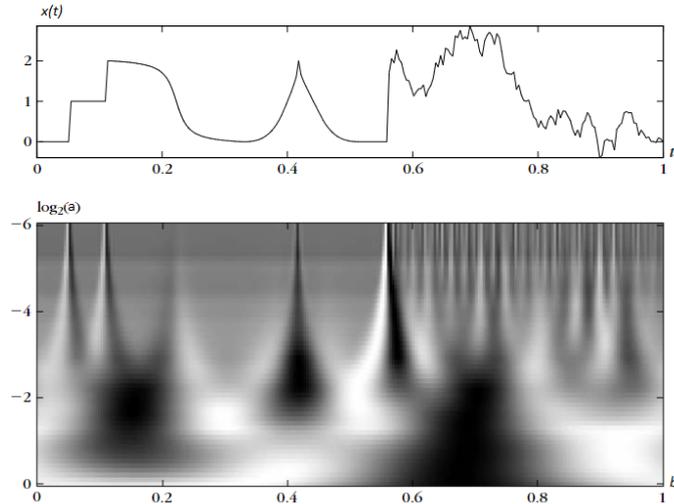
$$W(a, b) = \int_{-\infty}^{\infty} \psi_{ab}(t)x^*(t)dt. \quad (4.18)$$

Alternatively the more compact notation can be used:

$$W(a, b) = \langle \psi_{ab}, x \rangle. \quad (4.19)$$

By varying the scaling and shifting parameters throughout their respective domains, CWT yields WT coefficients to fill the whole time-frequency plane. An illustrated example of CWT is given in Figure 4.7.

However, CWT is not that practical signal analysis tool as it is highly redundant as both the scaling and shifting parameters are continuous and thus the wavelets at the same



**Figure 4.7.** Wavelet transform  $W(a, b) = \langle \psi_{ab}, x \rangle$  computed with a Mexican hat wavelet. Black, grey and white points correspond, respectively, to positive, zero and negative wavelet coefficients. Adapted from (S. Mallat and Peyre 2008).

scale and between the scales are significantly overlapping.

DWT provides a powerful tool for the MRA of signals. MRA is based on the signal approximations at different resolutions. The approximation of the signal  $x(t)$  at resolution  $2^{-m}$  is defined as an orthogonal projection of  $x(t)$  on the subspace  $\mathbf{V}_m \subset \mathbf{L}^2(\mathbb{R})$ . These subspaces  $\{\mathbf{V}_m \mid m \in \mathbb{Z}\}$  have to satisfy the following properties:

$$\mathbf{V}_\infty \subset \dots \subset \mathbf{V}_1 \subset \mathbf{V}_0 \subset \mathbf{V}_{-1} \subset \dots \subset \mathbf{V}_{-\infty} \quad (4.20)$$

$$\bigcap_{m \in \mathbb{Z}} \mathbf{V}_m = \emptyset, \quad \bigcup_{m \in \mathbb{Z}} \mathbf{V}_m = \mathbf{L}^2(\mathbb{R}) \quad (4.21)$$

$$x(t) \in \mathbf{V}_m \Leftrightarrow x(t - 2^m k) \in \mathbf{V}_m, \quad m, k \in \mathbb{Z} \quad (4.22)$$

$$x(t) \in \mathbf{V}_m \Leftrightarrow x(2t) \in \mathbf{V}_{m-1}, \quad m \in \mathbb{Z} \quad (4.23)$$

and that there also exists scaling function  $\varphi(t) \in \mathbf{V}_0$ , such that

$$\{\varphi_{mn}(t) = 2^{-m/2} \varphi(2^{-m}t - n)\}, \quad m, n \in \mathbb{Z} \quad (4.24)$$

is an orthonormal basis for  $\mathbf{V}_m$ , that is:

$$\langle \varphi_{mk}, \varphi_{ml} \rangle = \delta(k - l), \quad (4.25)$$

where  $\delta(\cdot)$  is Kronecker delta function.

Based on (4.20), (4.23) and (4.24) the following holds for the scaling function  $\varphi(t)$ :

$$\varphi(t) = 2 \sum_n h_0(n) \varphi(2t - n), \quad (4.26)$$

where the coefficients  $h_0(n)$ , known as the interscale basis coefficients, define the scaling function  $\varphi(t)$ . The equation (4.26) can be called either the refinement equation, the dilation equation or the multiresolution analysis equation.

Let  $\mathbf{W}_m$  be a complementary subspace for  $\mathbf{V}_m$ , such that:

$$\begin{aligned} \mathbf{V}_{m-1} &= \mathbf{V}_m \oplus \mathbf{W}_m \\ \mathbf{V}_m &\perp \mathbf{W}_m \end{aligned} \quad (4.27)$$

From (4.27) and (4.21) it follows that:

$$\bigcup_{m \in \mathbb{Z}} \mathbf{W}_m = \mathbf{L}^2(\mathbb{R}) \quad (4.28)$$

It can be assumed that there exists function  $\psi(t) \in \mathbf{W}_0$ , such that

$$\{\psi_{mn}(t) = 2^{-m/2} \psi(2^{-m}t - n)\}, \quad m, n \in \mathbb{Z} \quad (4.29)$$

is an orthonormal wavelet basis for  $\mathbf{W}_m$ .

The wavelet function can be expressed as linear combination of the translates of  $\varphi(2t)$ , in the similar manner as the equation (4.26) for the scaling function:

$$\psi(t) = 2 \sum_n h_1(n) \varphi(2t - n). \quad (4.30)$$

Orthogonal wavelet transform for  $x(t)$  is now defined as:

$$d(m, n) = \langle x, \psi_{mn} \rangle. \quad (4.31)$$

and orthogonal scaling transform for  $x(t)$  is defined as:

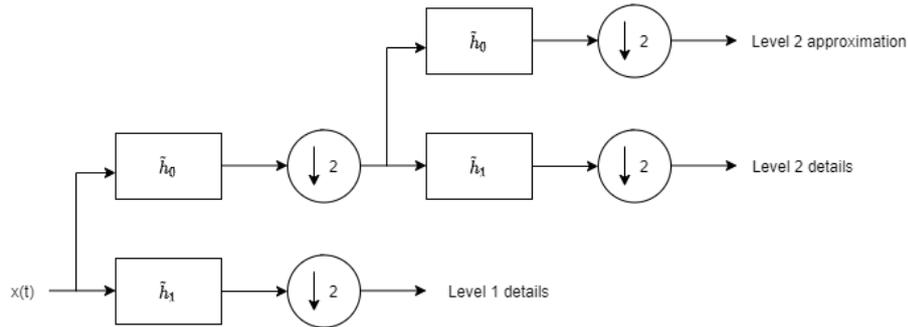
$$c(m, n) = \langle x, \varphi_{mn} \rangle. \quad (4.32)$$

Signal  $x(t) \in \mathbf{V}_0$  can be presented, or reconstructed, as linear combination of its orthogonal projections:

$$x(t) = \sum_{n=-\infty}^{\infty} c(L, n) 2^{-L/2} \varphi\left(\frac{t}{2^L} - n\right) + \sum_{m=1}^L \sum_{n=-\infty}^{\infty} d(m, n) 2^{-m/2} \psi\left(\frac{t}{2^m} - n\right). \quad (4.33)$$

That is,  $x(t)$  is presented as approximation at scale  $L$  complemented with the sum of  $L$  detail components at different resolutions.

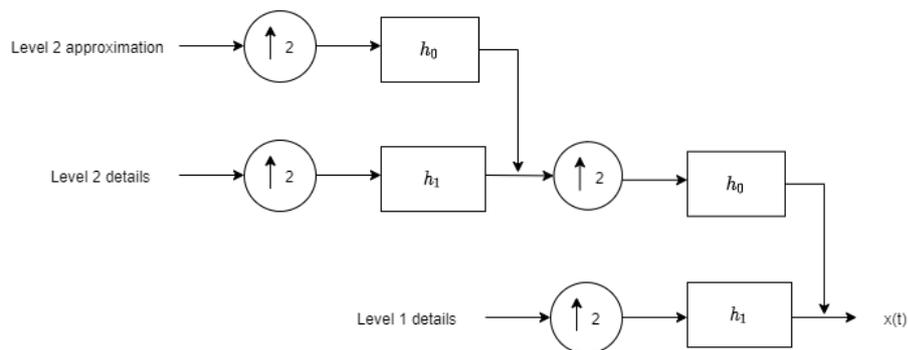
Detail coefficients (4.31) and approximation coefficients (4.32) at different scales, or resolutions, can be obtained via dyadic half-band filter bank implementation, with downsampling at each stage. In MRA terms, when a sampled signal  $x(t)$  is fed to the filter bank, a coarser approximation and respective details of the signal are obtained at each level, that is, effectively a signal decomposition, or MRA, is carried out. The filter bank implementation for MRA is illustrated in Figure 4.8.



**Figure 4.8.** Dyadic analysis filter bank implementation of DWT.

The filter coefficients  $\tilde{h}_0$  for the low-pass filter, and  $\tilde{h}_1$  for the high-pass filter are derived from those coefficients that define the scaling function and the wavelet function in (4.26) and (4.30), respectively, such that  $\tilde{h}_i(n) \hat{=} h_i(-n)$ . The latter holds as analysis filters in this construction are anticausal.

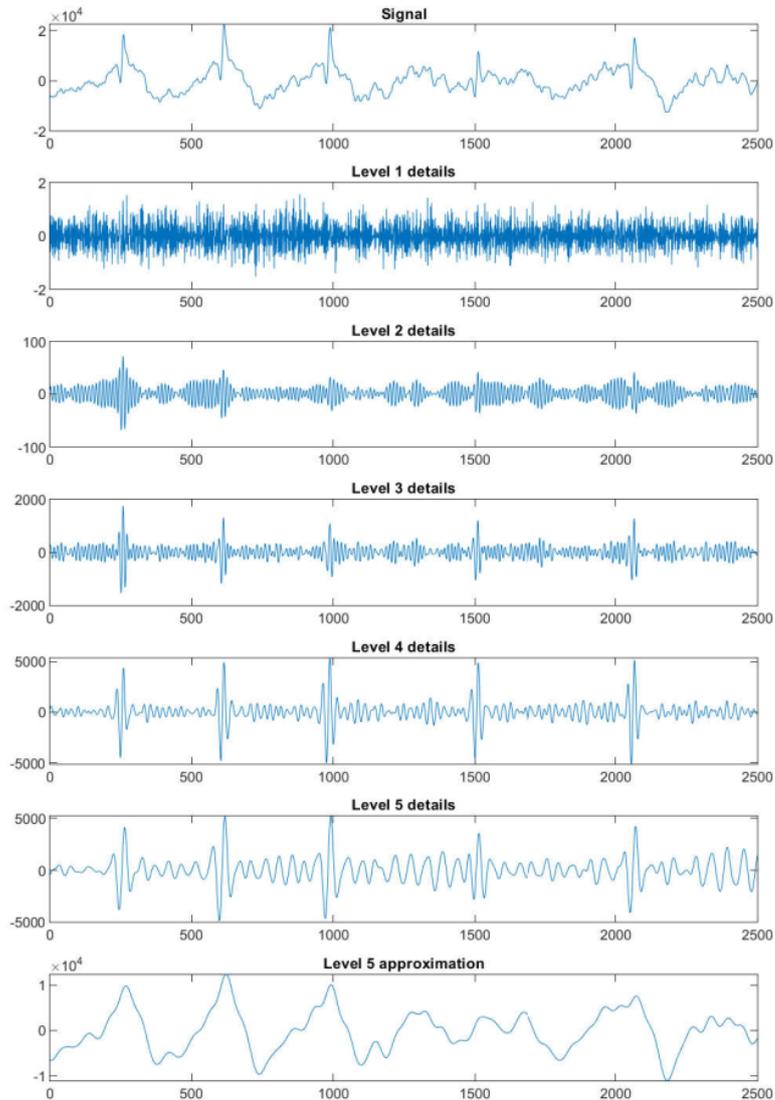
The signal can be reconstructed from the signal decompositions via dyadic synthesis filter bank with upsampling at each stage, as illustrated in Figure 4.9. The filter coefficients  $h_0$  for the low-pass filter, and  $h_1$  for the high-pass filter are those that define the scaling function and the wavelet function in (4.26) and (4.30), respectively.



**Figure 4.9.** Dyadic synthesis filter bank implementation.

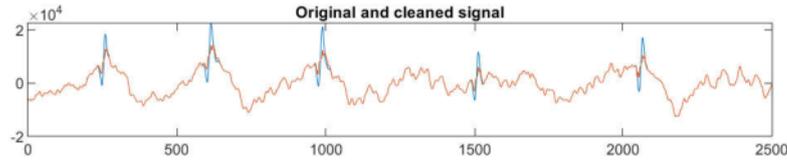
The more detailed description for the filter bank implementation can be found in Haddad and Akansu (2000).

As an example of the DWT based MRA, the signal decomposition for a segment of the EEG signal measured from the C3 channel is illustrated in Figure 4.10. The sampling rate for the EEG signal is 500 Hz.



**Figure 4.10.** DWT based decomposition for EEG signal (with Haar wavelet).

In the EEG signal, presented in Figure 4.10, there are clearly visible spikes, or artefacts, that are caused by ECG. MRA provides tools for cleaning these artefacts efficiently and, importantly, without removing the underlying EEG signal. This can be accomplished by thresholding the detail components and then reconstructing the signal from the thresholded details and the approximation, as per equation (4.33). The EEG signal before and after the ECG artefact removal is shown in Figure 4.11. Artefact removal is one of the most prominent applications of MRA in the EEG analysis.



**Figure 4.11.** The original EEG signal (blue) and the cleaned and reconstructed EEG signal (red).

### 4.3 Coherence

Coherence is a statistic that indicates whether two signals are linearly related. As discussed already earlier, it basically measures the synchronization between the signals based on the phase difference.

The magnitude squared coherence (MSC) is defined as follows:

$$\gamma_{xy}^2(\omega) = \frac{|S_{xy}(\omega)|^2}{S_{xx}(\omega)S_{yy}(\omega)}, \quad (4.34)$$

where  $\omega$  is the angular frequency,  $S_{xx}(\omega)$  and  $S_{yy}(\omega)$  are the autospectra of the signals  $x$  and  $y$ , respectively, and that are calculated similarly as (4.5).  $S_{xy}(\omega)$  is the cross-spectrum between the signals  $x$  and  $y$ , defined in the similar manner as (4.5):

$$S_{xy}(\omega) = \sum_{k=-\infty}^{\infty} r_{xy}(k)e^{-j\omega k}, \quad (4.35)$$

where  $r_{xy}$  is cross-correlation function, given as follows:

$$r_{xy}(k) = E[x(t)y^*(t-k)]. \quad (4.36)$$

MSC gives a value between 0 and 1. The value close to 1 indicates that the signals are strongly synchronized, whereas the value close to 0 indicates that the signals are desynchronized.

### 4.4 Phase-amplitude coupling and phase locking value

As already shortly discussed in Chapter 3.3.3, CFC phenomenon provides a quite important and useful viewpoint for the EEG based brain connectivity analysis. Methods for evaluating phase-phase and phase-amplitude couplings are presented in the following.

PLV is a measure that evaluates phase-locking between two signals, and it is calculated according to the following procedure (Lachaux et al. 1999):

1. Two signals  $x$  and  $y$  are filtered via narrow band-pass filters (center frequencies set to the frequencies of interest,  $f_1$  for  $x$  and  $f_2$  for  $y$ ) to obtain the frequency

components  $x_{f_1}$  and  $y_{f_2}$

2. For both  $x_{f_1}$  and  $y_{f_2}$  their instantaneous phases are obtained using the Hilbert transform (alternatively the Gabor wavelet transform could be used in this step, as originally suggested by Lachaux et al. (1999)):  $\phi(x_{f_1}) = \phi(\mathbf{H}[x_{f_1}])$  and  $\phi(y_{f_2}) = \phi(\mathbf{H}[y_{f_2}])$
3. PLV between  $x_{f_1}$  and  $y_{f_2}$ , denoted as  $\text{PLV}_{xy}$ , is then calculated as:

$$\text{PLV}_{xy} = \frac{1}{N} \left| \sum_{n=0}^{N-1} e^{j(\phi(x_{f_1}(n)) - \phi(y_{f_2}(n)))} \right|, \quad (4.37)$$

where  $N$  is the number of the samples in the signal.

The original concept of PLV can be extended and adapted to give a similar measure for PAC, that is, to measure the strength of the locking between the phase and amplitude of the signals at the frequencies of interest. Typically, PAC is calculated between the phase of a lower frequency component and the phase of the amplitude envelope of a higher frequency component, within the same signal or between different signals (different EEG channels, for example). The procedure to measure PAC is given as follows (Lipping et al. 2018):

1. Modulating signal  $x$  is filtered via low-pass filter to obtain the modulating frequency component  $x_{low}$
2. The instantaneous phase of the signal of the modulating frequency is obtained using the Hilbert transform:  $\phi(x_{low}) = \phi(\mathbf{H}[x_{low}])$
3. Modulated signal  $y$  is filtered via high-pass filter to obtain the higher (modulated) frequency component  $y_{high}$
4. The amplitude envelope (denoted by  $\langle \cdot \rangle$ ) of the  $y_{high}$  is obtained by calculating the norm of the Hilbert transform of the result of step 3:  $\langle y_{high} \rangle = |\mathbf{H}[y_{high}]|$
5. The result of step 4 is filtered via low-pass filter to obtain the low frequency component of the amplitude envelope  $\langle y_{high} \rangle_{low}$
6. The instantaneous phase of the low-pass filtered amplitude envelope of step 5 is obtained using the Hilbert transform:  $\phi(\langle y_{high} \rangle_{low}) = \phi(\mathbf{H}[\langle y_{high} \rangle_{low}])$
7. Then the locking value for PAC, denoted as  $\text{PAC}_{xy}$ , is obtained as follows:

$$\text{PAC}_{xy} = \frac{1}{N} \left| \sum_{n=0}^{N-1} e^{j(\phi(x_{low}(n)) - \phi(\langle y_{high} \rangle_{low}(n)))} \right|, \quad (4.38)$$

where  $N$  is the number of the samples in the signal.

PLV or PAC gives a value between 0 and 1. The value close to 0 indicates that there is

no locking between the signals, whereas the value close to 1 indicates that the signals are strongly locked, at the selected frequencies.

## 5 EXPERIMENT ON N-BACK RECORDINGS

The practical part of this thesis is described in the following sub-chapters. Some earlier discussed methods were applied to process EEG recordings that were acquired during n-back gaming sessions, and to calculate EEG metrics. The processing and calculation were performed with the software developed for the purpose.

### 5.1 EEG validator and calculator

One goal of this thesis was to implement a tool for filtering, cleaning and validating EEG recordings and also for calculating EEG metrics (for the mental load evaluation) from the processed recordings. The source code for the tool can be found in the GitHub repository (Ahonen 2020).

#### 5.1.1 Required functionality for the tool

EDF (European Data Format; Kemp, Värri et al. (1992)) and EDF+ (Kemp and Olivan 2003) are de-facto standards for the exchange and storage of polygraphic signals in a digital format. EDF+ is an enhanced version of EDF supporting, for example, annotations. The only incompatibility is that EDF+ supports interrupted recordings which are not supported by basic EDF. The EDF+ header and data structure are presented in Table 5.1. The very fundamental requirement for the developed tool was that it shall support *EDF+ file format*.

For the convenient processing of several recordings in one run, it was decided that the tool shall support *reading EDF file names and EEG channels to be processed from a Microsoft Excel sheet* that has been filled in by the user.

The core functionalities of the tool are the preprocessing of EEG recordings and the calculation of EEG metrics that could be applied to evaluate mental load. These two aspects are actually the main reasons why this tool was developed. The preprocessing of EEG recordings consists of the filtering, cleaning and validation of the EEG channels included in the recording. For the sake of flexibility and extensibility the tool shall support *filtering, cleaning, validation and calculation algorithms defined and implemented by the user* and provide the means to inject them with ease. The cleaning covers the *detection and removal of artefacts* which, depending on the implemented cleaning algorithm(s), could mean the zeroing of artefactual signal segments or some more sophisticated method that

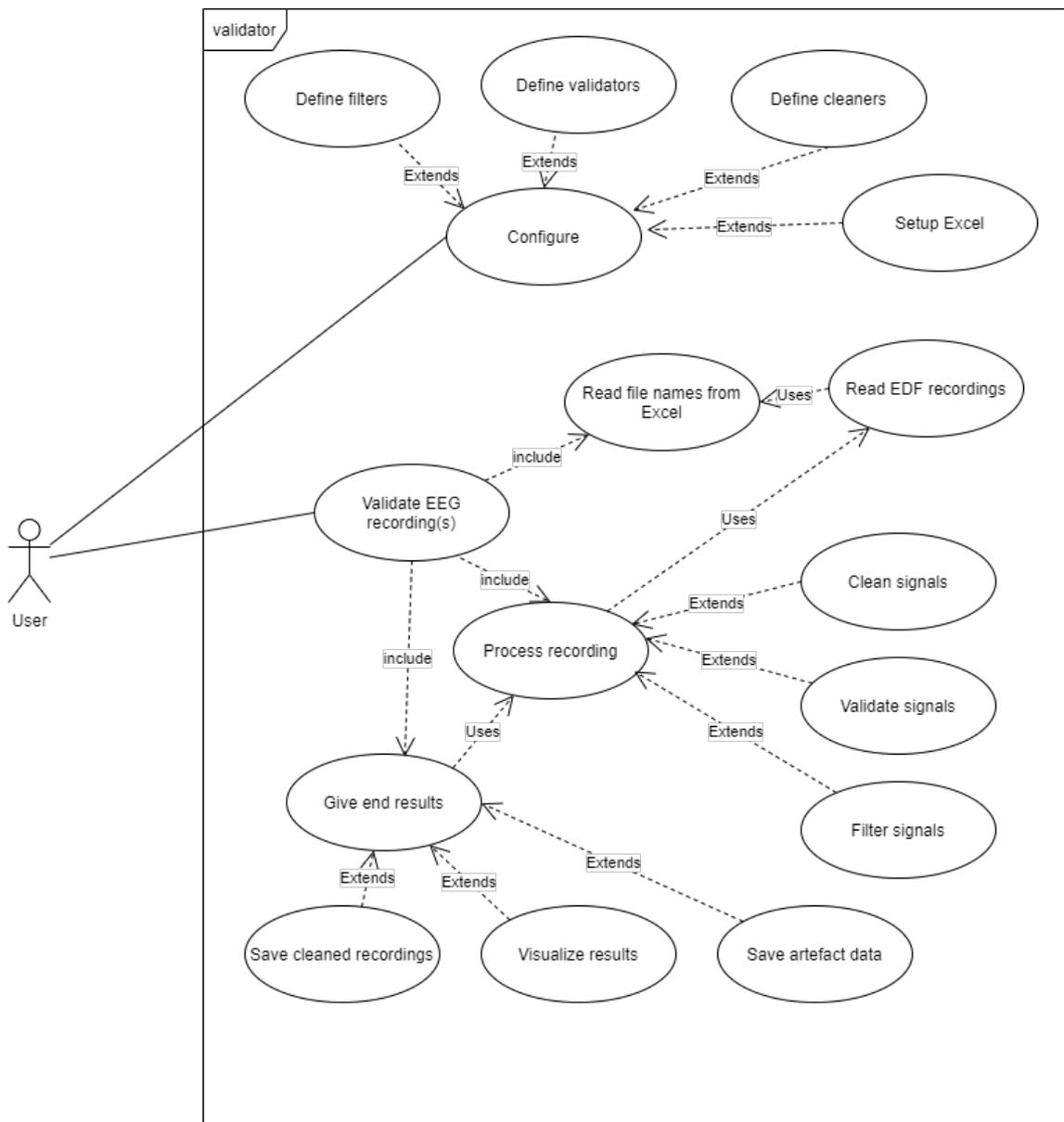
attempts to retain the underlying "real" EEG within the artefactual segments. The validation covers the *detection of artefactual segments and storing of the related information*, according to the defined validation rule(s) or algorithm(s). The validation shall take place after the signals have been cleaned. For an overall picture on the goodness or validity of EEG channels the tool shall provide the *visualization in the time domain indicating the valid and artefactual segments*. The tool shall *save the cleaned recording in EDF format* to be used for the calculation of EEG metrics.

**Table 5.1.** EDF header and data record structure (Kemp, Värri et al. 1992).

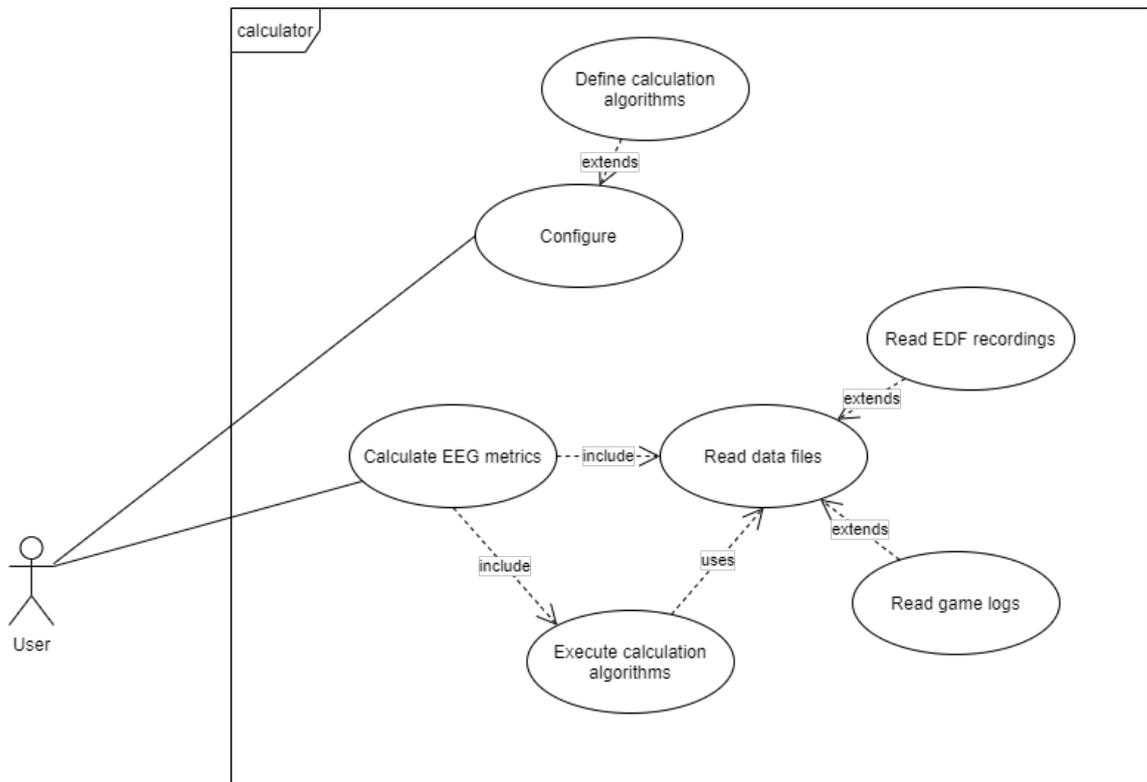
HEADER RECORD	
8	ascii : version of this data format (0)
80	ascii : local patient identification
80	ascii : local recording identification
8	ascii : startdate of recording (dd.mm.yy)
8	ascii : starttime of recording (hh.mm.ss)
8	ascii : number of bytes in header record
44	ascii : reserved
8	ascii : number of data records (-1 if unknown)
8	ascii : duration of a data record, in seconds
4	ascii : number of signals (ns) in data record
ns * 16	ascii : ns * label (e.g. EEG Fpz-Cz or Body temp)
ns * 80	ascii : ns * transducer type (e.g. AgAgCl electrode)
ns * 8	ascii : ns * physical dimension (e.g. uV or degreeC)
ns * 8	ascii : ns * physical minimum (e.g. -500 or 34)
ns * 8	ascii : ns * physical maximum (e.g. 500 or 40)
ns * 8	ascii : ns * digital minimum (e.g. -2048)
ns * 8	ascii : ns * digital maximum (e.g. 2047)
ns * 80	ascii : ns * prefiltering (e.g. HP:0.1Hz LP:75Hz)
ns * 8	ascii : ns * nr of samples in each data record
ns * 32	ascii : ns * reserved
DATA RECORD	
nr of samples[1]	* integer : first signal in the data record
nr of samples[2]	* integer : second signal
..	
..	
nr of samples[ns]	* integer : last signal

The primary objective that was defined for the EEG metrics calculation was to *support mental load evaluation during n-back memory game sessions, based on the logged game events*. Log files that were recorded during the gaming sessions, along with EEG recordings, included timestamps for the game events assumed to evoke cognitive activity and thus to increase the mental load. It was considered that annotations supported by EDF+ could be useful, at least for the manual analysis, and thus the tool shall *construct EDF+ compatible annotation files for timestamped game events*.

The tool can be divided into two main level logical entities: *the validator* covering the preprocessing of EEG recordings and *the calculator* covering the EEG metrics calculation from the preprocessed EEG recordings. This division into two entities and their related functional requirements in the form of use cases are depicted in the use case diagrams (Figures 5.1 and 5.2).



**Figure 5.1.** The validator use case diagram.



**Figure 5.2.** The calculator use case diagram.

## 5.1.2 Programming language and runtime environment

The tool was implemented with Matlab (ML; version R2019a) as scripts that are run in the Matlab environment. The choice of ML for a programming language and environment was quite straightforward as, just to name a few reasons, ML provides an extensive set of built-in functions suitable for signal processing and statistical analysis, it has been comprehensively documented with examples, it has an active user community, and in general, ML has proven its capabilities in a myriad of different research projects.

## 5.1.3 Tool architecture

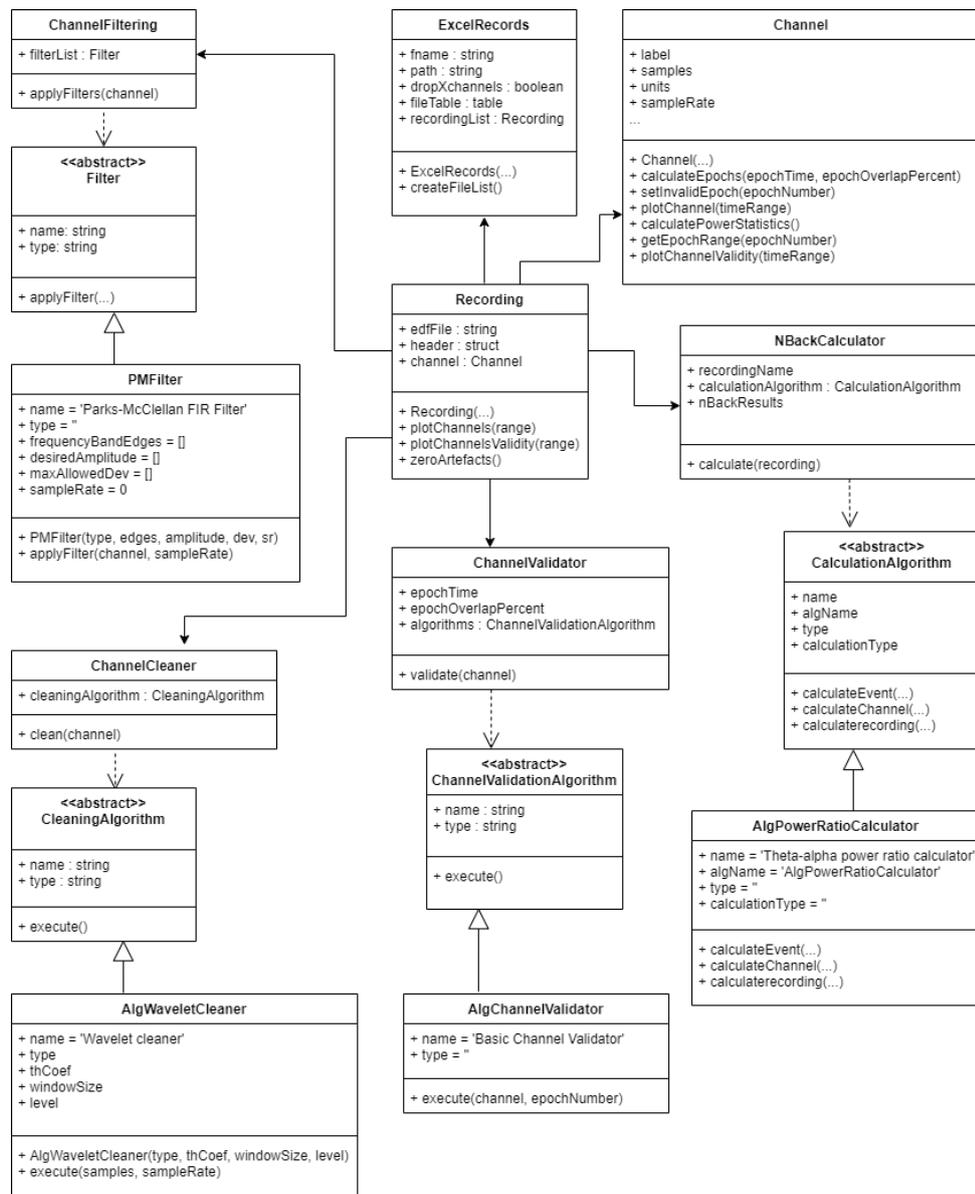
Accordingly to the logical division into two main entities, the validator and the calculator, the tool is also physically divided into two separately runnable Matlab scripts (`validator.m` and `calculator.m`). These two scripts can be considered as the "main executables" that gather and take control over the required classes and functions to compose a complete end-to-end application for both the validator and calculator entities.

The implementation is partly object oriented, deploying classes, and partly procedural, deploying direct ML function calls. Object oriented implementation supports modularity by nature, which was seen as important property for the tool. While this mixed approach may not be according to any commonly acknowledged good programming practices, it was taken as it enables fast and fluent prototyping in the ML environment. For some other

choice of the programming language and environment this approach might not have been feasible.

The only third-party function used in the implementation is slightly modified *edfread*<sup>4</sup> for reading EDF files, and it was also taken as basis for a new function *edfsave* that converts and saves internal data structure in EDF format.

The class diagram for the tool is shown in Figure 5.3. It should be noted that the classes implementing the abstract classes in the diagram are those that are deployed in the practical experiment (see Chapter 5.2), but they could be anything else as well, as defined and implemented by the user, on the condition that they implement the abstract classes as required.



**Figure 5.3.** The class diagram for the tool.

<sup>4</sup><https://se.mathworks.com/matlabcentral/fileexchange/31900-edfread>

The basic process flow for the EEG signal validation and metrics calculation is illustrated in Figure 5.4. As it can be seen, the flow is very lean and straightforward from end to end with a quite minimal, if any, user intervention needed during the execution. The activities are considered in more details in the sub-chapters 5.1.5 and 5.1.6.

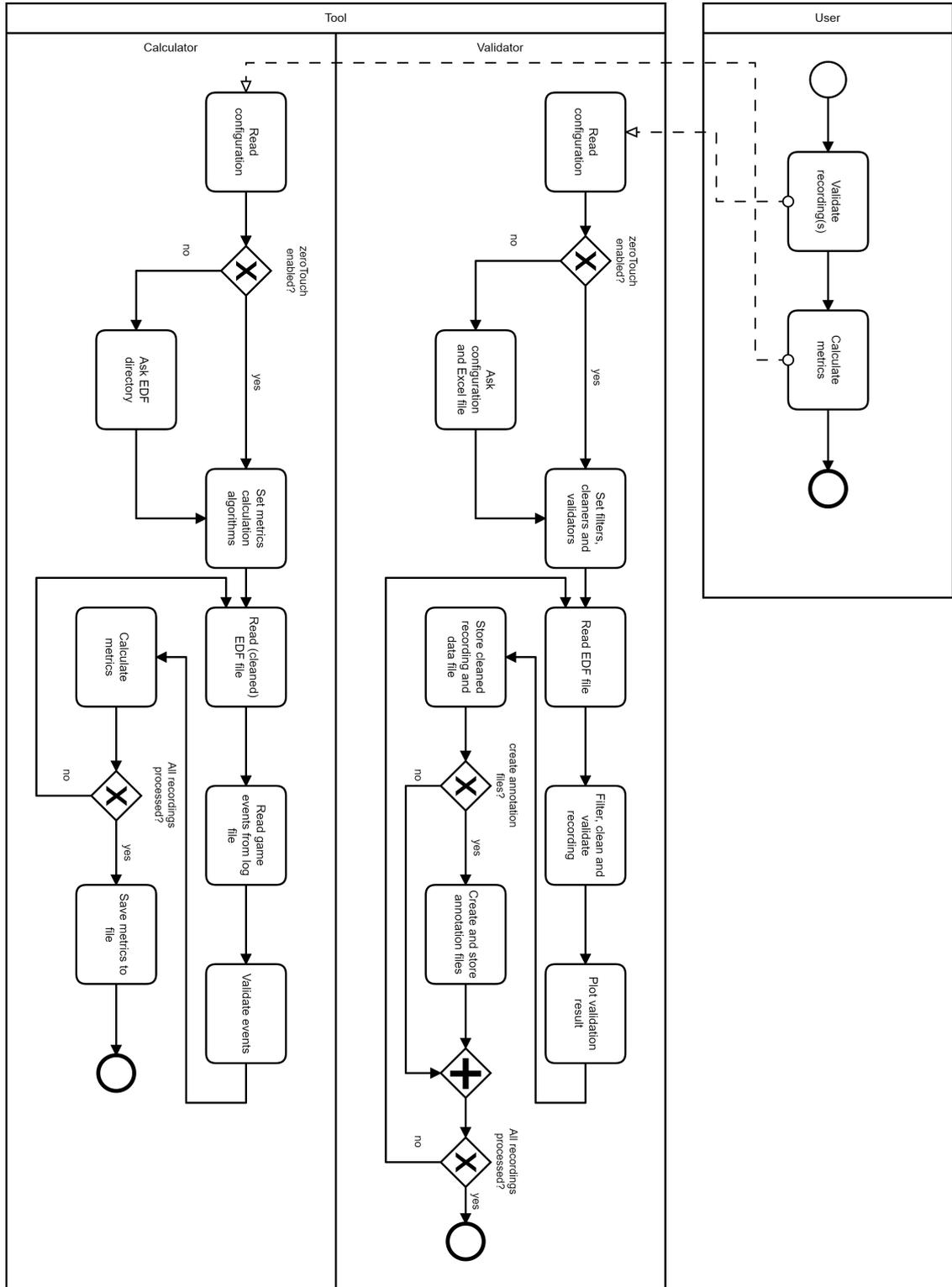


Figure 5.4. The process flow for the validation and calculation.

### 5.1.4 Tool configuration

User configurable parameters, including information on algorithms to be applied, are provided via the configuration files written in JavaScript Object Notation (JSON) syntax. For example, the user can configure the duration and overlap percentage values, according to which the EEG signals are divided into epochs that are then validated one by one. As these files are containing plain ASCII code, they can be edited with any text editor. It is possible to have several pre-defined configurations that can be deployed easily by changing or updating the configuration file that is read by the tool. The separate configuration files provide some convenience to the user and allow to keep the actual ML code intact, as there is no need to make any modifications directly to the tool itself due to configuration changes.

The user can implement new algorithms for the filtering, cleaning, validation and EEG metrics calculation. The algorithms are implemented as separate classes inherited from their respective abstract parent classes, and they are deployed via the configuration files. As an example, key-value pairs related to a band-pass filter deployment are shown in the code snippet taken from the JSON configuration file for the validator (Code 5.1). The snippet from the related ML code, including the constructor method *PMFilter* and the method *applyFilter* that performs the actual filtering, is shown in Code 5.2.

```
1 "filters" :
2 [
3   {
4     "filter" : "PMFilter",
5     "parameters" : {
6       "type" : "'Band Pass'",
7       "frequencyBandEdges" : "[0, 1, 44, 45]",
8       "desiredAmplitude" : "[0, 1, 0]",
9       "maxAllowedDev" : "[0.05, 0.05, 0.05]",
10      "sampleRate" : "0"
11    }
12  }
13 ]
```

**Code 5.1.** Filter deployment via the JSON configuration file.

```

1  methods
2      % constructor
3      function this = PMFilter(type, edges, amplitude, dev, sr)
4          this.type = type;
5          this.frequencyBandEdges = edges;
6          this.desiredAmplitude = amplitude;
7          this.maxAllowedDev = dev;
8          this.sampleRate = sr;
9      end
10
11     function filteredChannel = applyFilter(self, channel, sampleRate)
12         ...
13         [n,fo,ao,w] = firpmord(self.frequencyBandEdges,
14             self.desiredAmplitude, self.maxAllowedDev, sampleRate);
15         B = firpm(n,fo,ao,w);
16         filteredChannel = filtfilt(B, 1, channel);
17     end

```

**Code 5.2.** Code snippet from the *PMFilter* class implementation.

The tool reads the EDF files and processes the channels according to the information given in the Microsoft Excel sheet that is created and filled in by the user and stored in the same directory with the tool scripts. The information that the user needs to fill in is, for example, the filenames for the EDF files and character strings indicating how the channels in a specific EDF file are treated, that is, whether a channel contains EEG or ECG measurements, or is to be discarded. A screenshot showing the contents of one such configuration sheet is presented in Figure 5.5.

	A	B	C	D	E	F	G	H
1	filename	game	edf	easy	info	log	channels	comments
2	2019_G1	n-back	x	x	x	x	eeeeecx	e= EEG, c=ECG, x='discard'
3	2019_G2	n-back	x	x	x	x	eeeeecx	
4	2019_G3	n-back	x	x	x	x	eeeeecx	

**Figure 5.5.** Screenshot from an Excel sheet containing information on EDF files.

### 5.1.5 Functional split for the validator

In the following the main functionalities of the validator entity are described.

#### Reading the configuration

At the startup the validator reads the respective JSON configuration file. Based on the value of the control parameter *zeroTouch*, the validator either continues with the default settings or asks the user for the alternative configuration file and the location of the Excel sheet containing information on EDF files.

#### Reading EDF recordings

The validator creates an instance of the class *ExcelRecords* that forms the recording

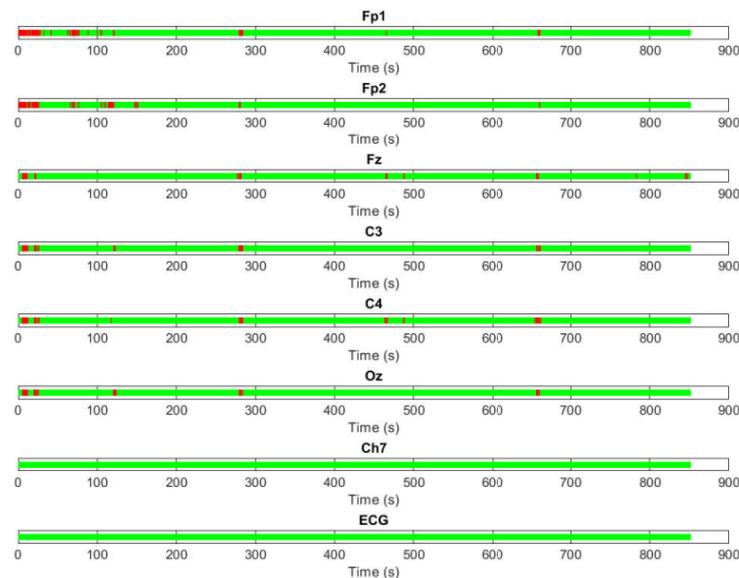
list containing the EDF file names and the channel specific information, as defined in the given Excel sheet. By following this recording list, the validator reads the EDF files and creates the *Recording* object, associated with the *Channel* objects, for each EEG recording.

### Filtering, cleaning and validation

The algorithms for filtering, cleaning and validating the EEG recordings are designed and implemented by the user, and they are deployed via the configuration file. The validator creates *ChannelFilter*, *ChannelCleaner* and *ChannelValidator* objects that are associated to their respective instances of the algorithm classes, as configured. EEG channels in each *Recording* object are then filtered, cleaned and validated by executing the algorithms via these objects.

### Visualization of cleaned channels

After the channels in the recording have been filtered, cleaned and validated, the validator gives the user a visual overview, in the time domain, of the validity of the channels. Those segments that are considered as valid (non-artefactual) are presented in green colour, and those segments that are considered as non-valid (artefactual) are presented in red colour. An example of this visualization is shown in Figure 5.6.



**Figure 5.6.** The visualization of the validity of the channels. Valid epochs are marked with green colour, and artefactual epochs with red (channels 'Ch7' and 'ECG' are not processed and thus they appear valid as such).

### Saving cleaned recordings and artefact data

The validator converts the filtered and cleaned recordings into EDF format and stores them as EDF files to the disk. In addition, a data file containing information on artefactual segments and other channel specific information is stored to be used for the EEG metrics calculation.

## Creating annotation files

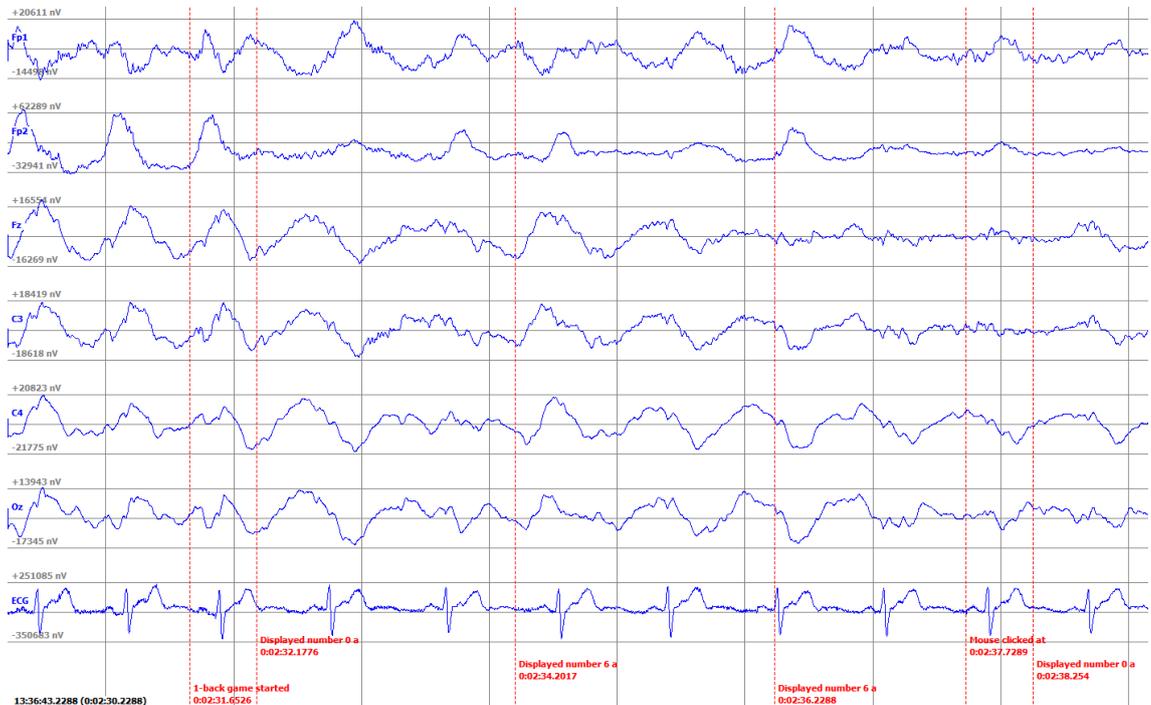
EEG recordings stored in the EDF format can be visually observed with the *EDFbrowser* application that supports annotations to be shown on the top of the visualized signals. For this purpose the validator creates annotation files that indicate game events derived from log files recorded during gaming sessions. Annotation files follow the XML format supported by EDFbrowser. A snippet of an XML annotation file created by the validator is shown in Code 5.3, and in Figure 5.7 it is shown how the imported XML annotations appear in EDFbrowser.

```

1  <?xml version="1.0" encoding="UTF-8"?>
2  <annotationlist>
3    <recording_start_time>2017-04-07T13:34:13.0000000</recording_start_time>
4    <annotation>
5      <onset>2017-04-07T13:36:49.228895</onset>
6      <duration></duration>
7      <description>Displayed number 6 at </description>
8    </annotation>
9    <annotation>
10     <onset>2017-04-07T13:36:50.728980</onset>
11     <duration></duration>
12     <description>Mouse clicked at </description>
13   </annotation>
14 </annotationlist>

```

**Code 5.3.** A snippet of an XML annotation file.



**Figure 5.7.** Annotations shown in EDFbrowser.

## 5.1.6 Functional split for the calculator

In the following the main functionalities of the calculator entity are described.

### Reading the configuration

This part is quite similar to the validator. At the startup the calculator reads the respective JSON configuration file. Based on the value of the control parameter *zeroTouch*, the calculator either continues with the default settings or asks the user for the alternative configuration file, and the location of the EDF files containing the EEG recordings for which the EEG metrics are calculated.

### Reading cleaned EDF recording and data files

The calculator lists all the EDF files found in the given directory (and in its all sub-directories) and creates *Recording* object for each recording. The created objects are complemented with the channel specific data that is read from the related data files constructed and stored by the validator.

### Reading the game events

The calculator reads the log files that have been stored during n-back game sessions and derives the following information to be stored into an internal data structure:

- the current value of  $n$
- timestamp when a number was displayed
- whether the mouse was clicked
- the delay between the number displayed and the mouse clicked
- whether the displayed number was the target or a non-target.

A snippet of an n-back game log file is shown in Code 5.4.

```
Recording started at 2017-04-07 13:34:16.975186
Deep breathing started at 2017-04-07 13:34:37.073336
1-back game started at 2017-04-07 13:36:44.652633
Displayed number 0 at 2017-04-07 13:36:45.177663
Displayed number 6 at 2017-04-07 13:36:47.201779
Displayed number 6 at 2017-04-07 13:36:49.228895
Mouse clicked at 2017-04-07 13:36:50.728980
```

**Code 5.4.** A snippet of an n-back game log file.

The stored game session specific data, that is related to the certain EEG recording, can be then used for calculating the game event related EEG metrics.

### Validating the game events

By comparing the timestamps of the artefactual EEG segments and the game events, the calculator discards the events that are overlapping with the artefactual segments, that is,

those events are not considered as valid for the EEG metrics calculation.

### **EEG metrics calculation**

The algorithms for calculating the EEG metrics are designed and implemented by the user, and they are deployed via the configuration file. These calculation algorithms have to implement the methods inherited from the abstract parent class, that is, each child class has to have the methods to calculate the metrics for the single events, for the channels and for the whole recording. Depending on the case, it may be that it is not reasonable, or feasible, to calculate the metrics for all the different levels, and in that case "dummy" implementation(s) might be needed.

The calculator creates the *NBackCalculator* object that is associated to the instances of the algorithm classes for the EEG metrics calculation, as configured. The EEG metrics, related to the logged game events, are then calculated for each *Recording* object (at the recording, channel or event level) by executing the algorithms via the *NBackCalculator* object. The results are stored into the internal data structure that is also saved to the disk.

## **5.2 The n-back experiment**

In the following, the experimental part of this thesis, that is, the employment of the developed tool for the EEG signals processing and mental load evaluation in n-back gaming sessions, and the analysis of the results, is described. The EEG measurements and stimuli related data used in this experiment were acquired by groups of Tampere University students and staff, as a part of course projects during the years 2019 and 2020.

### **5.2.1 The setup and procedure**

The test setup for course projects was comprised of the following components:

- a laptop
- n-back game and EEG recording software running on the laptop
- an EEG electrode cap with a measurement unit connected wirelessly to the laptop.

The n-back game involved simple visual stimuli as single digits (0-9) displayed on the computer screen in 2-3 second intervals. The subjects were supposed to click the mouse button when a displayed number was the target. The gaming sessions were always started with 1-back, proceeding up to 4-back, with a short break between the levels. The duration for each level was a few minutes. All game events during the sessions, i.e. n-back game level changes, displayed numbers and mouse clicks, were logged and stored to the disk in the log file format presented in Chapter 5.1.6.

The EEG during the n-back gaming sessions was measured and recorded with Enobio<sup>®</sup> (Neuroelectronics, Spain) EEG system. The version of the measurement device that was

used in the course projects carried out in 2019 supported up to 8 channels, and the version used in the projects in 2020 supported up to 20 channels.

In total, the projects covered 20 n-back gaming sessions, for which the EEG measurements and game events related data were recorded. A setup with 6 connected EEG channels was used in 18 of these sessions, and a higher density setup with 19 connected EEG channels in two of the sessions. The EEG signal sampling rate was 500 Hz for all recordings, and in all cases the ECG was recorded simultaneously with the EEG. The connected EEG channels in both the 6-channel and 19-channel setups are shown in Table 5.2. The channels in the 6-channel setup are a subset of the channels in the 19-channel setup.

**Table 5.2.** 6-channel and 19-channel setups.

Setup	Connected EEG channels
6-channel	Fp1, Fp2, C3, C4, O1, O2
19-channel	Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, O2

## 5.2.2 The EEG processing

The provided EEG recordings were preprocessed with the *validator* tool, that is, all EEG channels in all EEG recordings were filtered, cleaned and validated, in this order, by the respective algorithms that are described in the following.

Built-in ML functions *firpmord* and *firpm* are applied in the implemented filtering algorithm (*PMFilter*) to design a bandpass filter. The bandpass filter was deployed with the lower cut-off frequency set at 1 Hz and the upper cut-off frequency at 44 Hz.

The implemented artefact cleaning algorithm (*AlgWaveletCleaner*) employs DWT based MRA, by the application of ML functions *wfilters* and *modwt*. The cleaning algorithm was deployed with the low-pass and high-pass decomposition and reconstruction filters (see Chapter 4.2) associated with the ML pre-defined wavelet "db2" from the Daubechies wavelet family. With this setting, the *AlgWaveletCleaner* algorithm executes MRA decomposition for the EEG signals, and thresholding for the decomposed detail components. The thresholding is carried out in the manner that samples of the detail components are zeroed if they have an amplitude higher than 2.5 times the standard deviation calculated over the sample window of duration of 10 seconds, and which is moved in the 5 second intervals. This kind of less frequently moved window gives a remarkable performance gain in comparison to a window that is moved for each sample, without excessively losing the accuracy in this case. After the thresholding, the inverse DWT, via the ML function *imodwt*, is performed to reconstruct the (cleaned) EEG signal.

The implemented validation algorithm (*AlgChannelValidator*) loops through the channel specific EEG signal segments, or epochs, and compares the average power of each epoch to the average power of the whole channel. If the average power of an epoch is

more than two standard deviations away from the average channel power, the epoch is marked as invalid and will be omitted in the EEG metrics calculation. The epoch length was set to one second (500 samples at 500 Hz sampling rate), with 50% overlap between adjacent epochs.

### 5.2.3 The EEG metrics calculation

The EEG metrics were calculated with the *calculator* tool, after the recordings were first preprocessed with the *validator* tool, as described in the previous section. The selected EEG metrics for mental load evaluation, and the implemented algorithm to perform the needed computations are described in the following.

Different EEG indicators for mental load evaluation were discussed in Chapter 3.3. In this experiment the power spectrum based approach was chosen, by following the CLI method introduced by Holm et al. (2009), with such adaptation that instead of calculating the ratio of theta and alpha power between the EEG channels, the ratio was calculated within each EEG channel. According to its definition, CLI is calculated between the Fz and Pz channels, which suggests that these channels could be feasible candidates for intra-channel power ratio calculations as well. However, in this experiment the parietal channels were not measured, except in two cases, and the power ratio was calculated for all the measured channels that were common for all recordings (Fp1, Fp2, C3, C4, O1 and O2).

The implemented EEG metrics calculation algorithm (*AlgPowerRatioCalculator*) computes the theta and alpha power ratio (TAPR) for EEG signal segments that are time-locked to n-back game events. The measurement interval, or the event segment, over which TAPR is calculated, was configured to start 100 ms prior to a number is displayed, with the total duration of 1000 ms (500 samples at 500 Hz sampling rate). The *AlgPowerRatioCalculator* algorithm calculates an event segment specific theta (4-8 Hz) and alpha (8-12 Hz) power by the application of the Welch's PSD estimate using the ML function *pwelch*. TAPR is then obtained as the ratio of the PSD estimates for the theta and alpha frequency ranges summed over their respective frequency bins. The algorithm discards those event segments that overlap with invalid EEG signal segments, as indicated by the *validator* tool. The Hamming window of length 128 samples, with the overlap of 64 samples, was applied in the PSD estimates.

### 5.2.4 Results and analysis

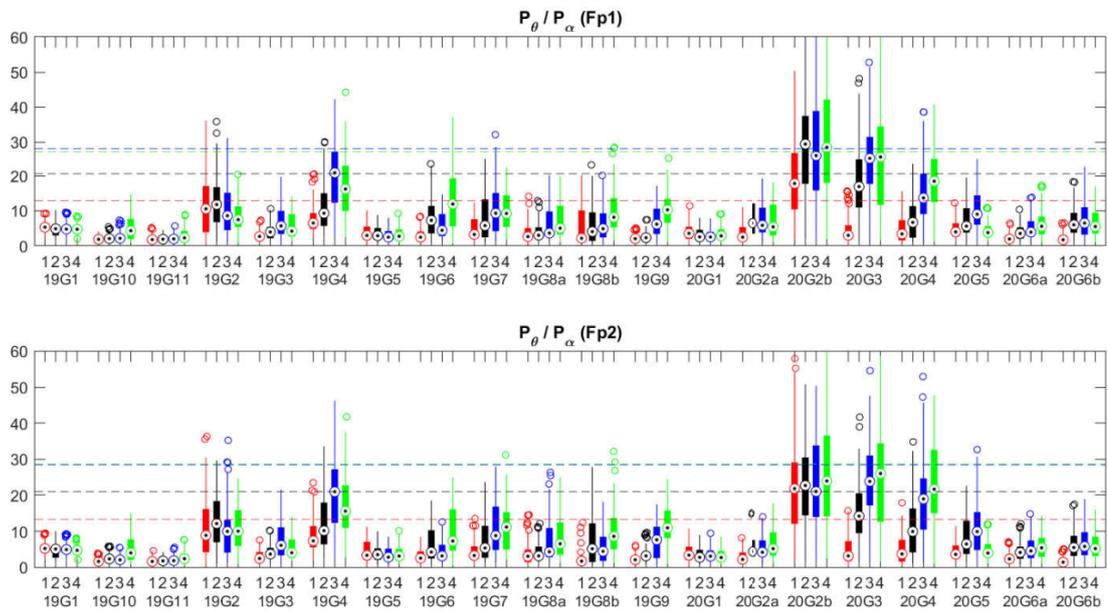
TAPR values, related to n-back game events, were calculated per n-back game level for each channel in each recording. Outliers in these channel and recording specific n-back level sets were removed by employing the median absolute deviation (MAD) method (Leys et al. 2013). MAD value is defined as the median of the deviations from the median,

adjusted with a scale factor:

$$\text{MAD} = b \cdot \text{Median}(|X_i - \tilde{X}|), \quad (5.1)$$

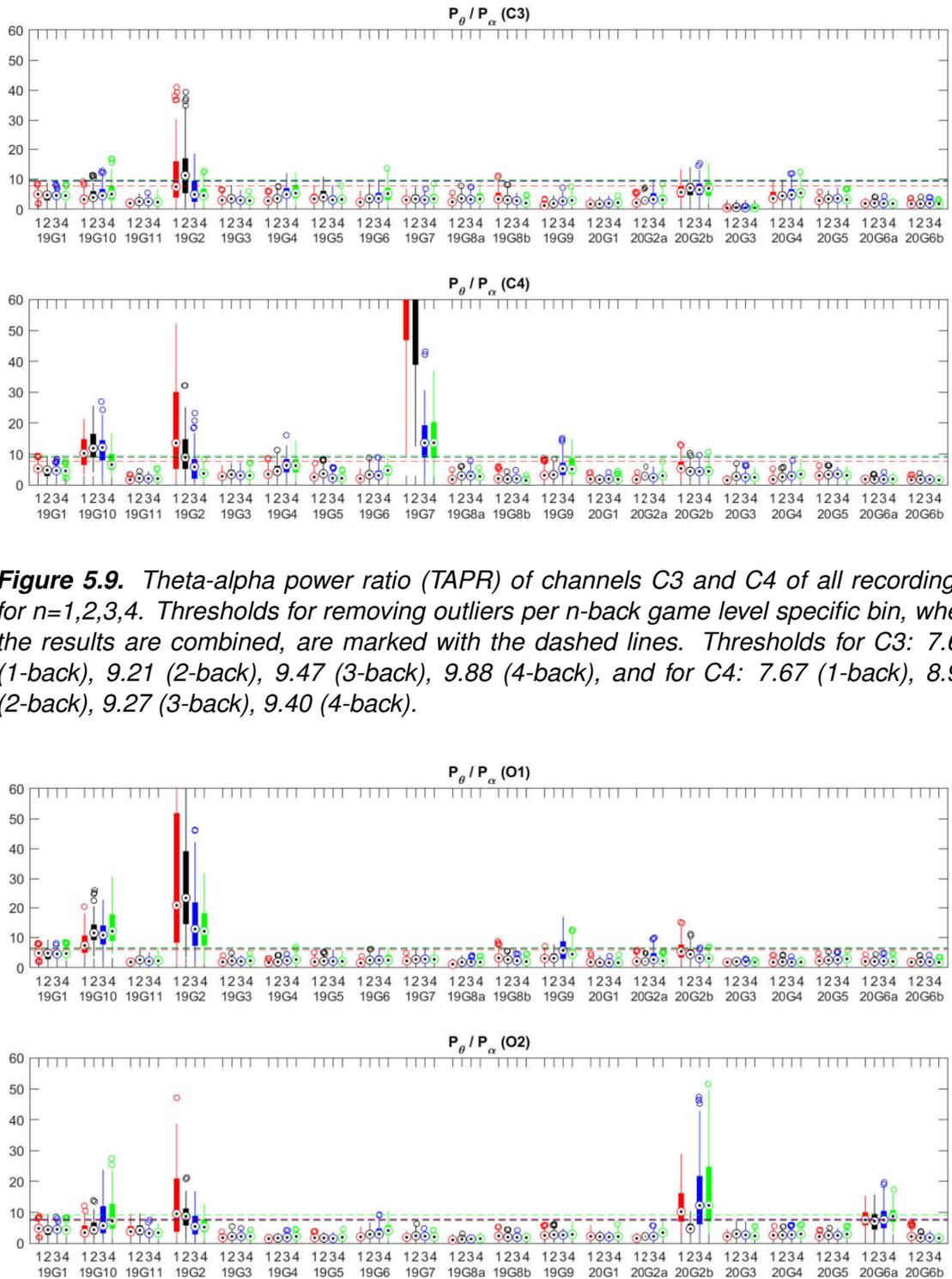
where  $\tilde{X}$  is the median of the data set  $X$ ,  $X_i$  represents individual values within this set and  $b$  is the scale factor (of value 1.4826, see (Leys et al. 2013) for details).

TAPR values more than three MAD values away from the median were considered as outliers. The results for each recording are shown in Figures 5.8, 5.9 and 5.10, where the recordings have been named after the identifiers of their respective project groups. The dashed lines mark the n-back game level specific thresholds for removing outliers (values above the threshold), when TAPR values for all recordings are combined in the channel specific n-back game level bins, as discussed later. The 19-channel setup was used in the recordings 20G6a and 20G6b, and the 6-channel setup in the rest.



**Figure 5.8.** Theta-alpha power ratio (TAPR) of channels Fp1 and Fp2 of all recordings, for  $n=1,2,3,4$ . Thresholds for removing outliers per n-back game level specific bins, when the results are combined, are marked with the dashed lines. Thresholds for Fp1: 12.84 (1-back), 20.59 (2-back), 28.11 (3-back), 27.30 (4-back), and for Fp2: 13.24 (1-back), 20.97 (2-back), 28.45 (3-back), 28.67 (4-back).

Based on the boxplot visualizations (Figures 5.8, 5.9 and 5.10), some recordings, e.g. 20G2b and 20G3 for the Fp2 channel, hold apparently higher TAPR values in comparison to other recordings. This could be due to individual divergence either in the theta or alpha activity, or in the both. Other reason could be a lack of concentration on the task execution, as these course project related measurements were not performed in strictly controlled circumstances. Artefacts may also contribute to these deviations, as the prototyped artefact removal algorithm was probably not capable of removing all the significant artefacts.



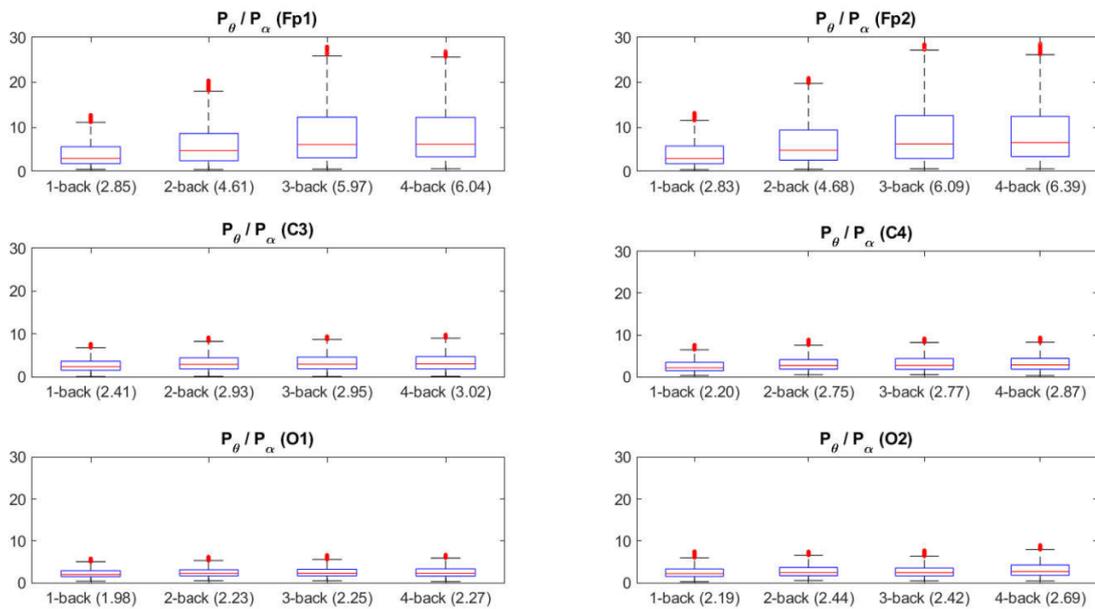
As discussed earlier, the higher the value of  $n$  is in the n-back game, the higher the imposed mental load, and the increase in the mental load should be reflected in the EEG as increased theta power and decreased alpha power. The Wilcoxon rank-sum test was conducted to test this hypothesis from TAPR perspective. The null hypothesis ( $H_0$ ) and alternative hypothesis ( $H_1$ ) were formulated as follows:

$$H_0 : \tilde{\mu}_x = \tilde{\mu}_y$$

$$H_1 : \tilde{\mu}_x < \tilde{\mu}_y ,$$

where  $\tilde{\mu}_x$  is the median of TAPR values calculated from EEG measurements during an n-back game session at level  $x$ ,  $\tilde{\mu}_y$  is the median of TAPR values calculated from EEG measurements during an n-back game session at level  $y$ , and  $x < y$ .

The channel specific TAPR values for the same n-back game level were combined from all recordings, and the MAD method was applied to remove outliers from these n-back specific bins. The combined TAPR values are shown in Figure 5.11. The thresholds, above which the TAPR values were removed as outliers, are illustrated with dashed lines in Figures 5.8, 5.9 and 5.10.



**Figure 5.11.** Theta-alpha power ratio (TAPR) of all channels of combined recordings, for  $n=1,2,3,4$ . The median of TAPR for each n-back level is given in parenthesis.

The left-sided Wilcoxon rank-sum test was separately performed for each pair of the n-back levels of the combined TAPR values. The p-values corresponding to the obtained rank-sum statistics are listed in Table 5.3. For the reference, the p-values for the rank-sum tests performed separately for each recording are given in Appendix A.

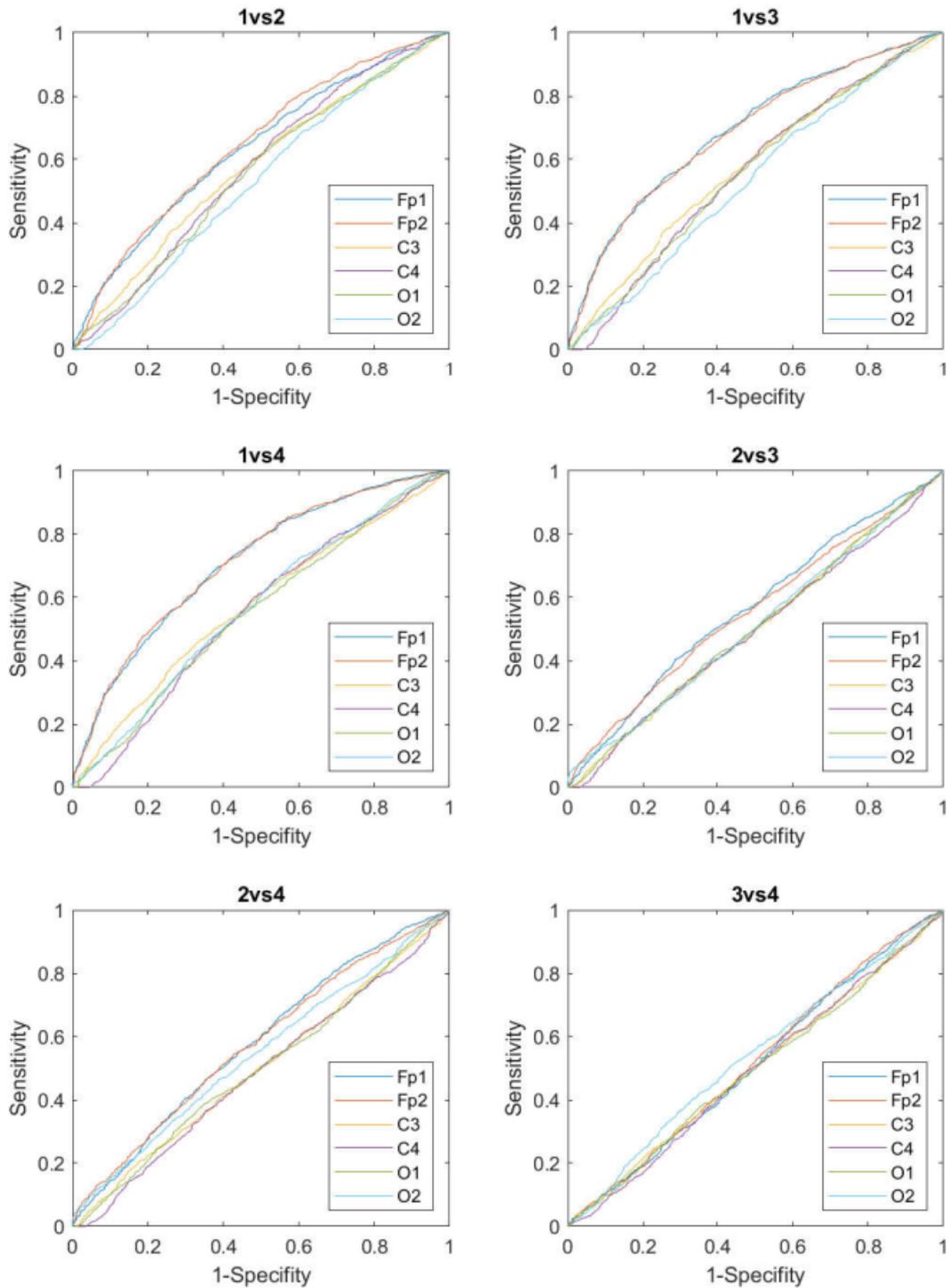
**Table 5.3.** *P-values for TAPR comparison between different n-back levels (left-sided Wilcoxon rank-sum test).*

<b>Channel</b>	<b>1vs2</b>	<b>1vs3</b>	<b>1vs4</b>	<b>2vs3</b>	<b>2vs4</b>	<b>3vs4</b>
Fp1	<0.001	<0.001	<0.001	<0.001	<0.001	0.177
Fp2	<0.001	<0.001	<0.001	<0.001	<0.001	0.055
C3	<0.001	<0.001	<0.001	0.203	0.217	0.520
C4	<0.001	<0.001	<0.001	0.451	0.825	0.884
O1	<0.001	<0.001	<0.001	0.108	0.404	0.839
O2	0.048	0.001	<0.001	0.070	<0.001	<0.001

It is visible in the boxplot presentations (Figure 5.11) that the differences between the medians of TAPR values for different n-back levels are highest in the Fp1 and Fp2 channels, whereas the medians in all the other channels reside on an almost flat line. This better TAPR performance for the Fp1 and Fp2 channels is evident by statistically significant p-values (Table 5.3), at the significance level 0.001, for TAPR comparison between each pair of n-back levels, except between the 3-back and 4-back. Although there are significant p-values for n-back comparisons in the other channels as well, they are not as dominant as in the Fp1 and Fp2 channels. TAPR values calculated for 4-back game events may not be that consistent in comparison to the results obtained from the lower n-back levels, as the n-back game at level 4 becomes so demanding, that it may start to cause concentration issues that disturb the evaluation.

ROC curves for TAPR performance in the different channels, and between the different n-back levels, give another visualization that also supports the better performance of the Fp1 and Fp2 channels (Figure 5.12).

To shortly summarize the presented observations and analysis, TAPR measure for the frontal EEG channels Fp1 and Fp2 seems to provide a plausible EEG indicator for mental load evaluation and comparison. This is aligned with the earlier discussed studies that indicate the theta synchronization and alpha desynchronization with increasing mental load. However, in an attempt to obtain more consistent and less divergent series of TAPR results between different recordings, more measurements should be performed in a controlled environment, and also other variations and methods for the artefact removal should be studied and experimented, including the visual assessment by an expert.



**Figure 5.12.** Channel specific ROC curves for TAPR measures between the different  $n$ -back levels.

## 6 CONCLUSIONS

Three main objectives were set for this thesis. The first one was to give a literature based overview on the feasibility of the EEG analysis in evaluating the mental workload of gaming. To start with, the fundamental questions to be covered were how the mental workload is defined and what are the mental workload measure types. In simple terms, mental workload can be considered as an objective task demand imposed on a person's cognitive resources. CLT provides a more profound theoretical framework, that is based on the cognitive architecture comprised of a working memory and a longterm memory, and that specifies mental workload according to its origins (intrinsic, extraneous or germane workload). The measures for mental workload can be divided into the categories of subjective measures, performance measures and psychophysiological measures. The EEG represents an indicator belonging to the last category. The reviewed studies indicate that gaming is a diverse source of mental workload, as one might intuitively assume, and mental workload imposed by game playing is usually fairly controllable, and the setups as such are simple and affordable. These aspects promote gaming as a favourable setting for mental load related studies. Furthermore, based on the reviewed studies, it is evident that the application of EEG analysis in evaluating mental load is highly feasible, and provides consistent and reproducible results that are in accordance to the expectations. One of the most important observations, that emerged in multiple studies, is the decrease in alpha activity and the increase in theta activity, with increasing mental load. This phenomenon was also an essential driver for the experimental part of this thesis. While in most of the reviewed studies mental workload evaluation was performed in NVR environments, some recent studies in VR environments have shown promising results on the feasibility of the EEG based mental workload analysis also in such more advanced environments.

The second objective was to develop a tool for analysing multichannel biosignal recordings of gaming sessions. The working prototype was implemented in ML environment. The chosen programming paradigm was a mixture of object-oriented (OO) and procedural programming. The modularity is an important property for the developed software, and OO approach is a natural enabler for modularity via the class based implementation. On the other hand, ML provides plenty of built-in functions that can be called directly, which justifies a partly procedural approach for fast prototyping. The main reason behind the modular design was to provide a user a convenient means to deploy new algorithms for the EEG processing and metrics calculation. The implemented algorithms for filtering,

cleaning and validating EEG signals, as well as for calculating the EEG metrics, were described in Chapter 5.2. According to the conducted experiments, the tool processes EEG signals and calculates EEG metrics as expected, and thus clearly fulfills the defined functional requirements. For future improvements, the program code could be investigated for further optimizations, as execution times might become an issue, if the tool needs to be run frequently, and for several recordings in one go. With the current system used in the experiment (CPU: Intel Core i5-7200U 2.50 GHz, memory: 8 GB DDR-4 2133 MHz, operating system: Windows 10 Professional 64-bit), and with the currently implemented algorithms, the validation for 20 EEG recordings, of the average length of roughly 15 minutes, took approximately 18 minutes, and the metrics calculation (TAPR) took approximately 10 minutes. Other improvements could be a graphical user interface, including parameter configuration, a built-in analysis section for providing statistics and graphs derived from the calculated metrics, and configurable game events, or character strings, to define triggering points for metrics calculation, also other than those based on n-back game log files.

The third main objective was to analyze a set of existing recordings for the detection of changes in the EEG during an n-back memory game. The setup, procedure, metrics calculation and analysis of results were described in details in Chapter 5.2. As a summary, the calculated TAPR metrics, for the frontal channels Fp1 and Fp2, seemed to provide meaningful results in order to evaluate and compare the mental workload imposed by the n-back game at different difficulty levels. However, for some recordings, remarkably high divergences in TAPR values were observed, in comparison to other recordings. This finding raises an obvious need for further acquisition of measurements, under strictly controlled conditions, and also for designing and experimenting efficient artefact cleaning algorithms. In addition, other EEG metrics than TAPR could be evaluated, e.g. those discussed in Chapter 3.3. A version of calculation algorithm for PAC metrics was preliminary tested. However, the initial results did not reflect the changes in mental workload, at least not as expected, but this is something that should be studied and experimented further.

## REFERENCES

- Acker, B. V., Parmentier, D. D., Vlerick, P. and Saldien, J. (2018). Understanding mental workload: from a clarifying concept analysis toward an implementable framework. *Cognition, technology & work* 20.3, pp. 351–365.
- Ahonen, V. (2020). *EEG-validator-calculator source code*. URL: <https://github.com/vahonen/eeg-validator-calculator>.
- Allison, B. Z. and Polich, J. (2008). Workload assessment of computer gaming using a single-stimulus event-related potential paradigm. *Biological psychology; Biol Psychol* 77.3, pp. 277–283.
- Ang, C. S., Zaphiris, P. and Mahmood, S. (2007). A model of cognitive loads in massively multiplayer online role playing games. *Interacting with Computers* 19.2, pp. 167–179.
- Antonenko, P., Paas, F., Grabner, R. and Gog, T. van (2010). Using Electroencephalography to Measure Cognitive Load. *Educational psychology review* 22.4, pp. 425–438.
- Aricò, P., Borghini, G., Flumeri, G. D., Colosimo, A., Bonelli, S., Golfetti, A., Pozzi, S., Imbert, J.-P., Granger, G., Benhacene, R. and Babiloni, F. (2016). Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain-Computer Interface Application in Realistic Air Traffic Control Environment. *Frontiers in human neuroscience; Front Hum Neurosci* 10, p. 539.
- Beer, N. A. de, Hooff, J. C. van, Brunia, C. H., Cluitmans, P. J., Korsten, H. H. and Beneken, J. E. (1996). Midlatency auditory evoked potentials as indicators of perceptual processing during general anaesthesia. *British journal of anaesthesia : BJA; Br J Anaesth* 77.5, pp. 617–624.
- Bruns, A. and Eckhorn, R. (2004). Task-related coupling from high- to low-frequency signals among visual cortical areas in human subdural recordings. *International Journal of Psychophysiology; Int J Psychophysiol* 51.2, pp. 97–116.
- Buttussi, F. and Chittaro, L. (2018). Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario. *IEEE Transactions on Visualization and Computer Graphics; IEEE Trans Vis Comput Graph* 24.2, pp. 1063–1076.
- Byrne, J. H. and Roberts, J. L. (2004). *From molecules to networks an introduction to cellular and molecular neuroscience*. Amsterdam ; Elsevier Academic Press.
- Canolty, R. T., Edwards, E., Dalal, S. S., Soltani, M., Nagarajan, S. S., Kirsch, H. E., Berger, M. S., Barbaro, N. M. and Knight, R. T. (2006). High Gamma Power Is Phase-

- Locked to Theta Oscillations in Human Neocortex. *Science (American Association for the Advancement of Science)*; *Science* 313.5793, pp. 1626–1628.
- Chen, Y., Ou, J. and Whittinghill, D. M. (2015). Cognitive Load in Real-Time Strategy Gaming: Human Opponent Versus AI Opponent. *The Computer Games Journal* 4.1, pp. 19–30.
- Collet, C., Averty, P. and Dittmar, A. (2009). Autonomic nervous system and subjective ratings of strain in air-traffic control. *Applied Ergonomics; Appl Ergon* 40.1, pp. 23–32.
- Comstock, J. R. and Arnegard, R. J. (1992). *The multi-attribute task battery for human operator workload and strategic behavior research*.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *The Behavioral and brain sciences; Behav Brain Sci* 24.1, pp. 87–114.
- Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco: Jossey-Bass.
- Dan, A. and Reiner, M. (2017). EEG-based cognitive load of processing events in 3D virtual worlds is lower than processing events in 2D displays. *International journal of psychophysiology; Int J Psychophysiol* 122, pp. 75–84.
- Dasari, D., Shou, G. and Ding, L. (2017). ICA-Derived EEG Correlates to Mental Fatigue, Effort, and Workload in a Realistically Simulated Air Traffic Control Task. *Frontiers in neuroscience; Front Neurosci* 11, p. 297.
- Ebersole, J. S., Husain, A. M. and Nordli, D. R. (2015). *Current Practice of Clinical Electroencephalography*. English.
- Friston, K. J. (2011). Functional and Effective Connectivity: A Review. *Brain connectivity; Brain Connect* 1.1, pp. 13–36.
- Gajewski, P. D., Hanisch, E., Falkenstein, M., Thönes, S. and Wascher, E. (2018). What Does the n-Back Task Measure as We Get Older? Relations Between Working-Memory Measures and Other Cognitive Functions Across the Lifespan. *Frontiers in psychology; Front Psychol* 9, p. 2208.
- Galy, E., Cariou, M. and Mélan, C. (2012). What is the relationship between mental workload factors and cognitive load types?: *International journal of psychophysiology; Int J Psychophysiol* 83.3, pp. 269–275.
- Gong, D., Li, Y., Yan, Y., Yao, Y., Gao, Y., Liu, T., Ma, W. and Yao, D. (2019). The high-working load states induced by action real-time strategy gaming: An EEG power spectrum and network study. *Neuropsychologia* 131, pp. 42–52.
- Graps, A. (1995). An introduction to wavelets. *IEEE Computational Science & Engineering* 2.2, pp. 50–61.

- Green, C. S. and Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature (London)*; *Nature* 423.6939, pp. 534–537.
- Haddad, P. A. and Akansu, A. N. (2000). *Multiresolution signal decomposition: transforms, subbands, and wavelets*. Academic Press.
- Hart, S. G. and Staveland, L. E. (1988). *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*.
- Hogervorst, M. A., Brouwer, A.-M. and Erp F., J. B. van (2014). Combining and comparing EEG, peripheral physiology and eye-related measures for the assessment of mental workload. *Frontiers in neuroscience*; *Front Neurosci* 8, p. 322.
- Holm, A., Lukander, K., Korpela, J., Sallinen, M. and M.I., K. M. (2009). Estimating brain load from the EEG. *TheScientificWorld*; *ScientificWorldJournal* 9, pp. 639–651.
- Jaeggi, S. M., Seewer, R., Nirkko, A. C., Eckstein, D., Schroth, G., Groner, R. and Gutbrod, K. (2003). Does excessive memory load attenuate activation in the prefrontal cortex? Load-dependent processing in single and dual tasks: functional magnetic resonance imaging study. *NeuroImage*; *NeuroImage* 19.2, pp. 210–225.
- Jiang, X., Bian, G.-B. and Tian, Z. (2019). Removal of Artifacts from EEG Signals: A Review. *Sensors (Basel, Switzerland)*; *Sensors (Basel)* 19.5, p. 987.
- Jones, E. G. (2010). *Cerebral Cortex*. Elsevier Ltd, pp. 769–773.
- Kandel, E. R. (2013). *Principles of neural science*. 5th ed. New York: McGraw Hill.
- Kane, M. J., Conway, A. R. A., Miura, T. K. and Colflesh, G. J. H. (2007). Working Memory, Attention Control, and the N-Back Task: A Question of Construct Validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*; *J Exp Psychol Learn Mem Cogn* 33.3, pp. 615–622.
- Kemp, B. and Olivan, J. (2003). European data format ‘plus’ (EDF+), an EDF alike standard format for the exchange of physiological data. *Clinical Neurophysiology*; *Clin Neurophysiol* 114.9, pp. 1755–1761.
- Kemp, B., Värri, A., Rosa, A. C., Nielsen, K. D. and Gade, J. (1992). A simple format for exchange of digitized polygraphic recordings. *Electroencephalography and clinical neurophysiology*; *Electroencephalogr Clin Neurophysiol* 82.5, pp. 391–393.
- Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of experimental psychology*; *J Exp Psychol* 55.4, pp. 352–358.
- Klem, G. H., Lüders, H. O., Jasper, H., Elger, C. et al. (1999). The ten-twenty electrode system of the International Federation. *Electroencephalogr Clin Neurophysiol* 52.3, pp. 3–6.

- Klimesch, W., Doppelmayr, M., Schwaiger, J., Auinger, P. and Winkler, T. (1999). Paradoxical' alpha synchronization in a memory task. *Brain research.Cognitive brain research; Brain Res Cogn Brain Res* 7.4, pp. 493–501.
- Klimesch, W. (1999). EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews* 29.2, pp. 169–195.
- Koeppen, B. M. and Stanton, B. A. (2018). *Berne & Levy physiology*. Philadelphia, PA: Elsevier.
- Kühn, S., Gallinat, J. and Mascherek, A. (2019). Effects of computer gaming on cognition, brain structure, and function: a critical reflection on existing literature. *Dialogues in clinical neuroscience; Dialogues Clin Neurosci* 21.3, pp. 319–330.
- Lachaux, J., Rodriguez, E., Martinerie, J. and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human brain mapping* 8.4, pp. 194–208.
- Lavie, N. and Cox, S. (2016). On the Efficiency of Visual Selective Attention: Efficient Visual Search Leads to Inefficient Distractor Rejection. *Psychological science; Psychol Sci* 8.5, pp. 395–396.
- Leys, C., Ley, C., Klein, O., Bernard, P. and Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of experimental social psychology* 49.4, pp. 764–766.
- Lipping, T., Erkintalo, N., Särkelä, M., Takala, R., Katila, A., Frantzén, J., Posti, J., Müller, M. and Tenovuo, O. (June 2018). Connectivity Analysis of Full Montage EEG in Traumatic Brain Injury Patients in the ICU. pp. 97–100.
- Luca, C. D. (2002). Surface electromyography: Detection and recording.
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. Cambridge, Massachusetts: The MIT Press.
- MacLeod, C. M. (1992). The Stroop task: The "gold standard" of attentional measures. *Journal of experimental psychology.General* 121.1, pp. 12–14.
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11.7, pp. 674–693.
- Mallat, S. and Peyre, G. (2008). *A Wavelet Tour of Signal Processing: The Sparse Way*. San Diego: Elsevier Science & Technology.
- Matern, M. F., Westhuizen, A. van der and Mostert, S. N. (2019). The effects of video gaming on visual selective attention. *South African Journal of Psychology*, pp. 183–194.

- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review; Psychol Rev* 63.2, pp. 81–97.
- Millet, D. (2002). *The origins of EEG. Seventh Annual Meeting of the International Society for the History of the Neurosciences (ISHN)*. URL: <http://www.bri.ucla.edu/nha/ishn/ab24-2002.htm>.
- Mulert, C. and Lemieux, L. (2010). *EEG - fMRI Physiological Basis, Technique, and Applications*. 1st ed. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Nayak, C. and Anilkumar, A. (2020). *EEG Normal Waveforms*. URL: <https://www.ncbi.nlm.nih.gov/books/NBK539805/>.
- Niedermeyer, E., Schomer, D. L. and Lopes da Silva, F. H. (2011). *Niedermeyer's electroencephalography : basic principles, clinical applications, and related fields*. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins Health.
- Olejniczak, P. (2006). Neurophysiologic Basis of EEG. *Journal of Clinical Neurophysiology* 23.3, pp. 186–189.
- OpenStax (2016). *Anatomy and Physiology*. URL: <http://cnx.org/contents/14fb4ad7-39a1-4eee-ab6e-3ef2482e3e22@8.24>.
- Palomäki, J., Kivikangas, M., Alafuzoff, A., Hakala, T. and Krause, C. M. (2012). Brain oscillatory 4–35Hz EEG responses during an n-back task with complex visual stimuli. *Neuroscience letters* 516.1, pp. 141–145.
- Paulraj, M. P., Subramaniam, K., Yaccob, S. B., Adom, A. H. B. and Hema, C. R. (2015). Auditory Evoked Potential Response and Hearing Loss: A Review. *The open biomedical engineering journal; Open Biomed Eng J* 9.1, pp. 17–24.
- Payne, L. and Kounios, J. (2008). Coherent oscillatory networks supporting short-term memory retention. *Brain research; Brain Res* 1247, pp. 126–132.
- Pergher, V., Wittevrongel, B., Tournoy, J., Schoenmakers, B. and M., M. V. H. (2018). N-back training and transfer effects revealed by behavioral responses and EEG. *Brain and behavior; Brain Behav* 8.11, p. e01136.
- Pesonen, M., Hämäläinen, H. and Krause, C. M. (2007). Brain oscillatory 4–30 Hz responses during a visual n-back memory task with varying memory load. *Brain research; Brain Res* 1138, pp. 171–177.
- Preece, J., Sharp, H. and Rogers, Y. (2015). *INTERACTION DESIGN: BEYOND HUMAN-COMPUTER INTERACTION*. Chichester: Wiley.
- Radüntz, T. (2017). Dual Frequency Head Maps: A New Method for Indexing Mental Workload Continuously during Execution of Cognitive Tasks. *Frontiers in physiology; Front Physiol* 8, p. 1019.

- Raymond, J. E., Shapiro, K. L. and Arnell, K. M. (1992). Temporary Suppression of Visual Processing in an RSVP Task: An Attentional Blink?: *Journal of Experimental Psychology: Human Perception and Performance* 18.3, pp. 849–860.
- Reid, G. B. and Nygren, T. E. (1988). *The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload*.
- Research and Markets (July 2020). *Global Virtual Reality In Gaming Market (2020 to 2025) - Growth, Trends, and Forecast*.
- Rouse, W. B. (1988). Adaptive aiding for human/computer control. *Human factors* 30.4, pp. 431–443.
- Roy, R. N., Bonnet, S., Charbonnier, S. and Campagne, A. (2016). Efficient Workload Classification based on Ignored Auditory Probes: A Proof of Concept. *Frontiers in human neuroscience; Front Hum Neurosci* 10, p. 519.
- Sakkalis, V. (2011). Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG. *Computers in biology and medicine* 41.12, pp. 1110–1117.
- Schrader, C. and Bastiaens, T. J. (2012). The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior* 28.2, pp. 648–658.
- Seeck, M., Koessler, L., Bast, T., Leijten, F., Michel, C., Baumgartner, C., He, B. and Beniczky, S. (2017). The standardized EEG electrode array of the IFCN. *Clinical Neurophysiology* 128.10, pp. 2070–2077.
- Sörnmo, L. and Laguna, P. (2005). *EEG Signal Processing*. Elsevier.
- Srinivasan, R., Winter, W. R., Ding, J. and Nunez, P. L. (2007). EEG and MEG coherence: Measures of functional connectivity at distinct spatial scales of neocortical dynamics. *Journal of neuroscience methods; J Neurosci Methods* 166.1, pp. 41–52.
- Stern, J. M. and Engel, J. (2013). *Atlas of EEG patterns*.
- Stoica, P. and Moses, R. L. (2005). *Spectral analysis of signals*. Upper Saddle River, NJ: Pearson Education.
- Su, C.-H. (2015). The effects of students' motivation, cognitive load and learning anxiety in gamification software engineering education: a structural equation modeling study. *Multimedia Tools and Applications* 75.16, pp. 10013–10036.
- Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational psychology review* 22.2, pp. 123–138.
- Sweller, J., Merrienboer, J. J. G. van and Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational psychology review* 10.3, pp. 251–296.

- Tatum, W. (2014). *Handbook of EEG Interpretation*. English. New York: Demos Medical Publishing.
- Tauscher, J., Schottky, F. W., Grogorick, S., Bittner, P. M., Mustafa, M. and Magnor, M. (2019). Immersive EEG: Evaluating Electroencephalography in Virtual Reality. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1794–1800.
- Teplan, M. (2002). Fundamental of EEG Measurement. *MEASUREMENT SCIENCE REVIEW* 2.
- Tremmel, C., Herff, C., Sato, T., Rechowicz, K., Yamani, Y. and Krusienski, D. J. (2019). Estimating Cognitive Workload in an Interactive Virtual Reality Environment Using EEG. *Frontiers in human neuroscience* 13, pp. 1–12.
- Turan, Z., Avinc, Z., Kara, K. and Goktas, Y. (2016). Gamification and Education: Achievements, Cognitive Loads, and Views of Students. *International journal of emerging technologies in learning* 11.7, p. 64.
- Vanhatalo, S., Palva, J. M., Holmes, M. D., Miller, J. W., Voipio, J., Kaila, K. and Raichle, M. E. (2004). Infralow Oscillations Modulate Excitability and Interictal Epileptic Activity in the Human Cortex during Sleep. *Proceedings of the National Academy of Sciences - PNAS; Proc Natl Acad Sci U S A* 101.14, pp. 5053–5057.
- Watson, C., Kirkcaldie, M. and Paxinos, G. (2010). *The Brain : An Introduction to Functional Neuroanatomy*. English. Saint Louis: Elsevier Science & Technology.
- Watter, S., Geffen, G. M. and Geffen, L. B. (2001). The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology; Psychophysiology* 38.6, pp. 998–1003.
- Xie, B. and Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. *Work and Stress* 14.1, pp. 74–99.
- Zhang, L. and Lin, W. (2013). *Selective visual attention : computational models and applications*. Singapore: John Wiley & Sons Inc.

## A P-VALUES SEPARATELY FOR EACH RECORDING AND EACH CHANNEL

Table A.1 lists the p-values corresponding to the left-sided Wilcoxon rank-sum statistics of TAPR comparison between the different n-back levels, separately for each recording and each channel.

**Table A.1.** P-values related to TAPR comparison between the different n-back levels

<b>19G1</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	0.05	0.13	0.01	0.64	0.6	0.42
Fp2	0.12	0.32	0.05	0.51	0.84	0.37
C3	0.2	0.04	0.03	0.64	0.69	0.86
C4	0.04	0.02	< 0.01	0.84	0.46	0.47
O1	0.37	0.14	0.25	0.73	0.85	0.7
O2	0.07	0.07	0.13	0.9	0.59	0.76
<b>19G10</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	0.12	< 0.01	< 0.001	0.15	< 0.001	< 0.001
Fp2	< 0.001	< 0.01	< 0.001	0.4	< 0.001	< 0.001
C3	0.01	< 0.001	< 0.001	0.19	0.1	0.61
C4	0.02	0.31	< 0.001	0.28	< 0.001	< 0.001
O1	< 0.001	< 0.001	< 0.001	0.17	0.42	0.03
O2	0.15	< 0.001	< 0.001	< 0.01	< 0.001	0.17
<b>19G11</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	0.23	0.25	< 0.01	0.98	0.03	0.04
Fp2	0.37	0.37	< 0.01	0.92	0.04	0.06
C3	< 0.01	< 0.01	< 0.01	0.98	0.84	0.96
C4	0.01	< 0.01	0.1	0.56	0.68	0.51
O1	< 0.001	0.07	< 0.01	0.01	0.65	0.16
O2	0.6	< 0.01	0.11	< 0.01	0.03	0.25

*Continues on the next page*

Table A.1 – Continued from the previous page

<b>19G2</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	0.31	0.63	0.11	0.07	< 0.001	0.14
Fp2	0.07	0.87	0.45	0.05	0.14	0.44
C3	0.25	< 0.01	< 0.001	< 0.001	< 0.001	0.59
C4	0.04	< 0.001	< 0.001	< 0.001	< 0.001	0.01
O1	0.7	< 0.01	< 0.01	< 0.001	< 0.001	0.43
O2	0.33	< 0.01	< 0.01	< 0.01	< 0.001	0.43
<b>19G3</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	< 0.01	< 0.001	< 0.001	< 0.001	0.05	0.1
Fp2	0.01	< 0.001	< 0.001	< 0.001	0.07	0.05
C3	0.09	0.71	0.66	0.11	0.01	0.31
C4	0.02	0.56	0.31	0.11	0.17	0.87
O1	< 0.01	0.05	< 0.01	0.55	0.95	0.51
O2	< 0.01	< 0.01	< 0.01	0.9	0.85	0.73
<b>19G4</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	< 0.01	< 0.001	< 0.001	< 0.001	< 0.001	0.08
Fp2	< 0.01	< 0.001	< 0.001	< 0.001	< 0.01	0.03
C3	< 0.01	< 0.001	< 0.001	< 0.001	< 0.001	0.73
C4	0.02	< 0.001	< 0.001	< 0.001	< 0.01	0.57
O1	0.03	< 0.001	< 0.001	< 0.01	< 0.001	0.05
O2	0.01	< 0.001	< 0.001	0.09	< 0.01	0.21
<b>19G5</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	0.65	0.04	0.21	0.08	0.43	0.4
Fp2	0.34	0.06	0.23	0.37	0.77	0.5
C3	0.02	0.3	0.37	< 0.01	< 0.01	0.77
C4	< 0.01	0.03	< 0.01	< 0.001	< 0.001	0.45
O1	0.34	0.4	0.33	0.95	0.95	1
O2	0.96	0.16	0.45	0.17	0.51	0.07
<b>19G6</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	< 0.001	< 0.001	< 0.001	0.04	< 0.01	< 0.001
Fp2	< 0.001	0.03	< 0.001	0.02	< 0.01	< 0.001
C3	< 0.001	< 0.001	< 0.001	0.44	< 0.01	0.02

Continues on the next page

Table A.1 – Continued from the previous page

C4	< 0.001	< 0.001	< 0.001	0.75	< 0.01	< 0.01
O1	< 0.001	< 0.001	< 0.001	0.25	0.32	0.95
O2	< 0.001	< 0.001	< 0.001	0.73	< 0.01	0.01
<b>19G7</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	0.02	< 0.001	< 0.001	0.04	0.02	0.87
Fp2	< 0.01	< 0.001	< 0.001	0.01	< 0.01	0.63
C3	0.02	0.47	0.02	0.1	0.69	0.04
C4	0.85	< 0.001	< 0.001	< 0.001	< 0.001	0.78
O1	< 0.01	< 0.001	< 0.001	0.64	0.74	0.93
O2	0.01	0.03	0.22	0.48	0.15	0.27
<b>19G8a</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	0.08	< 0.001	< 0.001	0.02	< 0.001	0.12
Fp2	0.07	< 0.001	< 0.001	< 0.01	< 0.001	0.29
C3	< 0.001	< 0.001	< 0.001	0.34	0.82	0.6
C4	< 0.001	< 0.001	< 0.001	0.26	0.53	0.13
O1	< 0.001	< 0.001	< 0.001	0.69	0.42	0.72
O2	< 0.001	< 0.001	< 0.001	0.8	0.08	0.1
<b>19G8b</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	0.54	0.08	< 0.001	0.35	< 0.001	< 0.001
Fp2	< 0.01	< 0.001	< 0.001	0.64	< 0.01	< 0.001
C3	0.43	0.02	< 0.001	0.09	< 0.001	< 0.01
C4	0.05	0.27	< 0.001	0.28	0.06	< 0.01
O1	0.17	< 0.01	< 0.001	0.01	< 0.001	0.06
O2	0.15	0.07	< 0.01	0.88	0.09	0.1
<b>19G9</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	0.07	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
Fp2	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001
C3	< 0.001	< 0.001	< 0.001	< 0.01	< 0.001	0.21
C4	0.62	< 0.001	< 0.001	< 0.001	< 0.001	0.61
O1	0.8	< 0.001	< 0.001	< 0.001	< 0.001	0.06
O2	0.44	0.76	0.18	0.63	0.49	0.24
<b>20G1</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4

Continues on the next page

Table A.1 – Continued from the previous page

Fp1	0.09	0.03	0.29	0.84	0.38	0.21
Fp2	0.25	0.22	0.19	0.97	0.95	0.86
C3	0.91	0.37	0.16	0.26	0.13	0.72
C4	0.61	0.5	0.61	0.17	0.95	0.25
O1	0.55	0.24	0.46	0.71	0.91	0.74
O2	0.27	0.28	0.84	0.92	0.45	0.46
<b>20G2a</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	< 0.001	< 0.001	< 0.001	0.58	0.27	0.59
Fp2	< 0.001	< 0.001	< 0.001	0.95	0.27	0.22
C3	< 0.001	< 0.001	< 0.001	0.31	0.28	0.78
C4	< 0.001	< 0.001	< 0.001	0.27	0.5	0.1
O1	0.47	< 0.001	0.13	< 0.001	0.34	< 0.01
O2	< 0.001	< 0.001	< 0.001	0.78	< 0.001	< 0.001
<b>20G2b</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	< 0.001	< 0.001	< 0.001	0.81	0.46	0.31
Fp2	0.18	0.14	0.03	0.95	0.44	0.47
C3	0.02	0.1	0.03	0.48	0.93	0.49
C4	< 0.001	< 0.001	< 0.01	0.54	0.9	0.48
O1	0.01	< 0.001	< 0.001	< 0.001	< 0.001	0.87
O2	< 0.001	0.31	0.09	< 0.001	< 0.001	0.36
<b>20G3</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	< 0.001	< 0.001	< 0.001	< 0.001	< 0.01	0.87
Fp2	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.47
C3	< 0.001	< 0.01	0.03	0.04	< 0.01	0.59
C4	< 0.001	< 0.001	< 0.001	0.12	0.08	0.87
O1	< 0.01	0.01	0.22	0.38	< 0.001	< 0.001
O2	< 0.001	< 0.01	0.28	0.16	< 0.001	0.04
<b>20G4</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	< 0.01
Fp2	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.06
C3	0.02	0.01	< 0.001	0.46	< 0.01	0.08
C4	< 0.001	< 0.001	< 0.001	0.06	< 0.001	0.15

Continues on the next page

Table A.1 – Continued from the previous page

O1	0.05	0.64	0.02	0.09	0.87	0.06
O2	0.94	0.22	< 0.01	0.25	< 0.01	0.05
<b>20G5</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
Fp1	< 0.001	< 0.001	0.65	< 0.01	< 0.001	< 0.001
Fp2	< 0.001	< 0.001	0.37	0.01	< 0.001	< 0.001
C3	< 0.01	< 0.001	0.02	0.18	0.47	0.05
C4	0.03	< 0.01	0.08	0.38	0.57	0.11
O1	0.02	0.01	< 0.001	0.8	0.07	0.08
O2	< 0.01	< 0.01	< 0.001	0.98	0.1	0.11
<b>20G6a</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
P7	0.01	0.93	0.99	0.02	0.03	0.94
P4	0.88	0.2	< 0.001	0.11	< 0.01	0.26
Cz	< 0.01	0.57	0.12	0.04	0.23	0.37
Pz	0.21	0.03	< 0.001	0.28	0.01	0.12
P3	0.6	0.97	0.45	0.64	0.18	0.43
P8	< 0.001	< 0.001	< 0.001	0.19	0.73	0.29
O1	0.77	0.41	0.96	0.34	0.81	0.45
O2	0.1	0.58	0.01	0.28	< 0.001	< 0.01
T8	0.81	0.88	0.71	0.79	0.89	0.69
F8	0.03	0.25	0.19	< 0.01	< 0.01	0.99
C4	0.96	0.95	0.77	0.88	0.81	0.76
F4	0.62	0.05	0.09	0.12	0.01	< 0.001
Fp2	< 0.001	< 0.001	< 0.001	0.17	< 0.01	0.22
Fz	0.25	< 0.001	< 0.001	< 0.001	< 0.01	0.39
C3	0.51	0.67	0.44	0.8	0.17	0.25
F3	0.33	< 0.01	< 0.01	0.08	0.12	0.87
Fp1	< 0.001	< 0.001	< 0.001	< 0.01	< 0.001	0.02
T7	0.89	0.04	0.46	0.02	0.37	0.22
F7	< 0.01	< 0.001	< 0.001	0.43	< 0.01	0.02
<b>20G6b</b>	1vs2	1vs3	1vs4	2vs3	2vs4	3vs4
P7	0.12	< 0.001	< 0.01	< 0.01	0.16	0.16
P4	< 0.001	< 0.001	< 0.001	0.2	0.41	0.03

Continues on the next page

Table A.1 – Continued from the previous page

Cz	0.35	< 0.001	< 0.01	< 0.01	0.02	0.55
Pz	0.25	< 0.01	0.25	0.11	0.98	0.11
P3	0.71	0.61	< 0.01	0.38	< 0.01	< 0.01
P8	0.1	0.49	< 0.01	0.36	0.17	0.03
O1	0.41	0.16	0.63	0.49	0.19	0.05
O2	< 0.001	< 0.01	< 0.001	0.29	0.3	0.04
T8	0.3	< 0.001	0.01	0.01	0.13	0.4
F8	< 0.001	< 0.001	< 0.001	0.04	0.68	0.01
C4	0.22	0.06	0.95	0.6	0.27	0.09
F4	< 0.001	< 0.001	0.8	0.47	< 0.001	< 0.001
Fp2	< 0.001	< 0.001	< 0.001	0.92	0.86	0.72
Fz	0.73	0.02	0.2	< 0.01	0.09	0.25
C3	0.43	0.72	0.26	0.23	0.72	0.1
F3	0.29	0.01	< 0.001	< 0.001	< 0.001	0.13
Fp1	< 0.001	< 0.001	< 0.001	0.56	0.79	0.48
T7	0.95	< 0.01	0.37	< 0.01	0.38	0.09
F7	< 0.001	< 0.001	< 0.001	0.94	0.79	0.77