Objectivation of Visual Perception

Simon Stock

Objectivation of Visual Perception

Zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

von der KIT-Fakultät für
Elektrotechnik und Informationstechnik
des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

M.Sc. Simon Stock

geb. in Aalen

Tag der mündlichen Prüfung: 24.01.2022

Hauptreferent: Prof. Dr. rer. nat. Wilhelm Stork
Korreferent: Prof. Dr. med. Gerd U. Auffarth

Preamble

> What does it mean, to see? The plain man's answer (and Aristotle's, too) would be, to know what is where by looking. In other words, vision the process of discovering from images what is present in the world, and where it is.
>
> *David Marr*
> *Vision [Mar10]*

This dissertation is revolving around visual perception in humans and computers and the link between both.

Passive voice is partially avoided in this book for a more pleasant reading flow. Therefore, "we" is often used, but it is noted that it expresses the author and the reader's world of thoughts. Unless otherwise stated or highlighted, the author of the thoughts is always the author of this dissertation.

Two citation styles are used in this work. For external sources, the alphabetic style finds application. Otherwise, for own sources like supervised theses and own publications, the IEEE style is used.

Abstract

The sense of sight enables a detailed perception of the world. Virtual Reality (VR), Brain-Computer Interfaces (BCI), and Deep Learning are technologies that provide opportunities for visual perception research. This dissertation introduces a system for ophthalmology that can simulate eye diseases in VR and provide an objective classification of visual field defects by adding BCI and AI. To better understand this work, a general four-step model is introduced to compare human vision with computer vision. Within the model, interfaces between the biological-real and the technological-virtual world are evaluated.

Nowadays, if a patient is suspected of having a visual field defect (scotoma), ophthalmological devices such as the perimeter are used to assess the visual field. However, the state-of-the-art procedure is based on subjective feedback from the patient. Accordingly, learning effects can influence the examination. This dissertation successfully implemented and evaluated an objective perimetry system based on VR, BCI, and Deep Learning. The new system may also be used in people with disabilities, small children, and animals. Initially, the pathology of glaucoma patients is assessed and mapped virtually using VR technologies in a simulation environment. Real-world individual glaucoma progression is mapped to an immersive VR experience. The resulting VR application was evaluated with healthcare professionals and glaucoma patients at the Heidelberg University eye clinic (N=22). It was shown that VR is a suitable measure for simulating visual conditions and can contribute to understanding the patient's condition. Based on this simulation environment, additional software and hardware modules were added. Generated steady-state visual stimuli were used to detect (simulated) visual field defects through an electroencephalography (EEG)-based BCI. The complete system has been tested and validated in an international laboratory study (N=15) in collaboration with the Massachusetts Institute of Technology. The collected data suggest that the system can provide good classification results for the central (88% accuracy per 2.5 seconds of EEG data) and peripheral visual field (63-81% accuracy),

while limitations (50-57% accuracy) reside for far peripheral positions due to technology sensitivity. Accordingly, the system could be used for detecting scotomas in central vision or entire quadrants. Due to the need for a better ambulatory EEG measurement setup, modular, cross-platform software implementations and novel patent-pending EEG electrodes are presented. The new electrodes provide better signal quality than state-of-art dry electrodes (1.35 dB improvement), are fast to set up, reusable, and leave little to no unwanted residue in the patient's hair. This dissertation lays the foundation for an ambulatory VR, BCI, and AI-based perimetry measurement system that could be used in the outpatient setting or for patients with limitations.

Zusammenfassung

Der Sehsinn ermöglicht eine detailgenaue Wahrnehmung der Welt. Virtual Reality (VR), Brain-Computer Interfaces (BCI) und Deep Learning sind neue Technologien, die uns hierbei neue Möglichkeiten für die Erforschung der visuellen Wahrnehmung geben. In dieser Dissertation wird ein System für die Augenheilkunde vorgestellt, das Augenkrankheiten in VR simulieren kann und durch Hinzufügen von BCI und KI eine objektive Diagnostik von Gesichtsfeldausfällen ermöglicht. Für ein besseres Verständnis der Arbeit wird das menschliche Sehen mit Modellen der Computer Vision verglichen und basierend hierauf ein allgemeines vierstufiges Seh-Modell eingeführt. Innerhalb des Modells werden Schnittstellen zwischen der biologisch-realen und der technologisch-virtuellen Welt evaluiert.

Besteht heutzutage bei einem Patienten der Verdacht auf einen Gesichtsfeldausfall (Skotom), so werden ophthalmologische Geräte wie das Perimeter zur Ausmessung des Gesichtsfeldes eingesetzt. Das dem Stand der Technik entsprechende Verfahren liegt dem subjektiven Feedback des Patienten zugrunde. Entsprechend können Lerneffekte beim Patienten das Ergebnis nicht unwesentlich beeinflussen. Um diese Problematiken zu umgehen, wurde in dieser Dissertation ein objektives Perimetriesystem auf Basis von VR, BCI und Deep Learning erfolgreich implementiert und evaluiert. Ein weiterer Vorteil des neuen Systems ist die Möglichkeit zur Einsetzung bei Menschen mit Schwerbehinderung, Kindern und Tieren. Der Lösungsansatz dieser Dissertation ist die Simulation (pathologischer/eingeschränkter) Sehzustände. Hierfür wurde der Zustand von Glaukompatienten mit Hilfe von VR-Technologien virtuell abgebildet. Die resultierende VR-Anwendung bildet individuelle Glaukomverläufe immersiv in VR ab. Evaluiert wurde die Simulationsumgebung mit medizinischem Fachpersonal und Glaukompatienten an der Augenklinik des Universitätsklinikums Heidelberg (N=22). Hierbei wurde gezeigt, dass VR eine geeignete Maßnahme zur Simulation von Sehbedingungen ist und zum Verständnis des Patientenzustandes einen Beitrag leisten kann. Aus-

gehend von dieser Simulationsumgebung wurden weitere Software- und Hardwaremodule hinzugefügt. Erzeugte stationäre visuelle Stimuli wurden hierbei eingesetzt, um (simulierte) Sehfehler durch ein Elektroenzephalographie (EEG)-basiertes BCI zu erkennen. Das System wurde in einer internationalen Laborstudie (*N*=15) in Zusammenarbeit mit dem Massachusetts Institute of Technology getestet und validiert. Die gesammelten Daten deuten darauf hin, dass das System für die Klassifizierung des zentralen (88% Genauigkeit pro 2,5 Sekunden EEG-Daten) und peripheren Gesichtsfeldes (63-81% Genauigkeit) geeignet ist, während es für periphere Positionen aufgrund der Technologiesensitivität zu Einschränkungen (50-57% Genauigkeit) kommt. Entsprechend sollte das System für Skotome eingesetzt werden, sofern der Sehausfall das zentrale Sehen oder ganze Quadranten des Gesichtsfelds betrifft. Aufgrund der Notwendigkeit für einen besseren ambulanten EEG-Messaufbau werden modulare, plattformübergreifende Softwareimplementierungen und neuartige, zum Patent angemeldete, EEG-Elektroden vorgestellt. Die neuartigen Elektroden bieten ein besseres Signal-zu-Rausch-Verhältnis als herkömmliche Trockenelektroden (1.35 dB Verbesserung), sind schnell anzulegen, wiederverwendbar und hinterlassen kaum bis keine unerwünschten Rückstände im Haar des Patienten. Diese Dissertation legt den Grundstein für ein VR, BCI und KI-basiertes Perimetrie-Messsystem, welches insbesondere im ambulanten Setting oder bei Patienten mit Einschränkungen zum Einsatz kommen könnte.

Contents

1 Introduction

> The most beautiful thing we can experience is the mysterious. It is the source of all true art and science.
>
> *Albert Einstein*

This work aims to travel together into the realm of visual perception. We already know a lot about light, vision, and how we perceive images. However, we are nowhere near a complete understanding. In the preamble, the quote of David Marr reads that vision is the ability to know **what is where** from viewing the world.

We will have a look at vision and visual perception from various points of view. On the one hand, we will analyze the biological vision as it occurs in humans and other animals. On the other hand, we will also examine vision in computers. We will learn about similarities between visual perception in computers and vision in humans. Furthermore, by utilizing interfaces like Virtual Reality (VR) and Brain-Computer Interfaces (BCI), we can exchange information from biological to technological vision and vice versa.

The visual perception system can be broken down into four main concepts: The environment, optics, sensory, and perception. These general steps are analog in both biological and technological vision. This general principle is presented in Figure 1.1.

In the first step, the environment is something existing outside the visual system. This is essential because if there is nothing around, there would not be anything to observe, making vision obsolete. Let us first have a look at the biological, or in other words, the real-world side of things. The environment is made from matter. Furthermore, we need a light source which is illuminating this environment. Light is reflected from all surfaces and is scattered. Depending on the environment's material and the properties of the light itself,
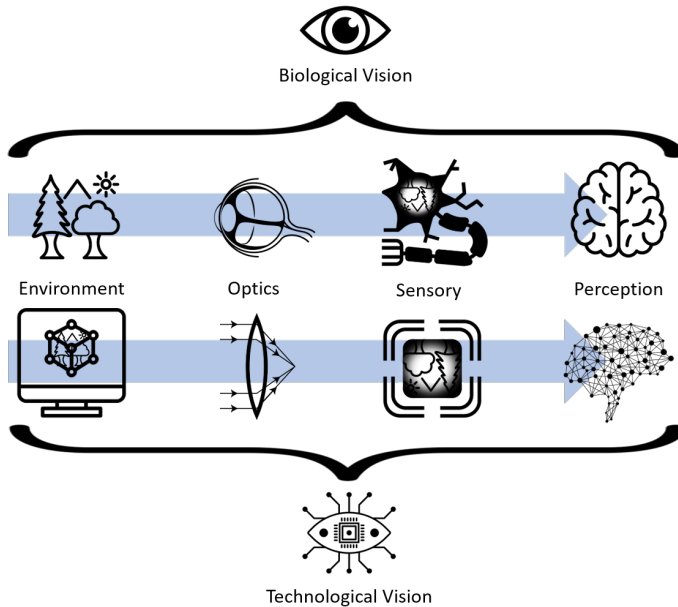
Figure 1.1: A general concept of vision in biology and technology, broken down to four successive steps.

an image is transported through space. When looking on the technological side, there is also a representation of the world. For a computer-generated 3D world, it is usually in vertexes and polygons - basically, the computational representation of the world. Computer-generated light sources can also be part of this representation. The easiest way to visualize this by an example is the not yet rendered (but existing within memory) 3D world of a computer game or a Computer Aided Design (CAD) software file.

The next step is optics. In all animals with some form of vision that is further developed than just distinguishing light and dark, some optics always play an essential role in visual perception. In mammals, these optics are the lens and the cornea of the eye. Furthermore, the form and curvature of the eye itself play an important role. The light which enters the eye is focused on the retina and its photosensitive receptors. Without this form of physical preprocessing, image

creation is hardly possible. In technological vision, this can be performed by ray-tracing in virtual worlds. Another example is a manufactured lens in front of a camera. As a result of this, the camera itself can be defined as an interface between real-world environment and technological vision.

The third stage is the sensing of light or information. In the retina, this is performed by multiple layers of photosensitive receptor cells. They physically transform the energy of the photons into electrical pulses. Bipolar, amacrine, and ganglion cells already perform information fusion and preprocessing and can be regarded as part of the final perception step. In technological vision, this can be represented by either an algorithm that manipulates the incoming rays, for example, shaders in 3D software. Another example of a technological pendant is the charge-coupled device (CCD) image sensor of a camera.

The final step which creates the understanding of seen stimuli is the perception stage. In biological vision, this is performed by the brain. The initial processing is happening in the visual cortex, which is located at the occiput for humans. Other parts of the brain use this processed information for the understanding of an image. We only know little about the actual inner workings of these processes at this time, but science is ever-evolving and improving. Computer vision perception, just like human perception, aims to make sense of the received information. With processed information, computers can detect, interact or even control objects in their environment. In the last decades, more and more sophisticated algorithms in computer vision arose. Modern computer vision is inspired by biological vision. In artificial intelligence (AI), the so-called convolutional neural networks try to mimic the processing steps of the visual cortex. They achieved classification accuracies on image recognition challenges [VDDP18].

1.1 Motivation

Vision is an essential human sense. We rely on it during most tasks in everyday settings. Right now, you are reading these words with the help of vision. Your brain extracts information from the letters and numbers presented in this book and is finally creating an understanding. Even slight visual impairments cause significant inconveniences. Wearer of glasses know that feeling when they put down their spectacles. Just imagine how your daily life would look like if you

were partially or entirely blind, everyday tasks like cooking, driving, shopping, and going outside would become significant challenges.

The **World Report on Vision** by the WHO from 2019 [WHO19] states that globally at least 2.2 billion people have some form of visual impairment. However, most of these are myopia or presbyopia, which can be easily treated with glasses, contact lenses, or even surgery. Neurodegenerative diseases like glaucoma, Age-Related Macular Degeneration (AMD), or Diabetic Retinopathy can lead to partial or complete blindness when not treated early. Like with all neurodegenerative diseases, the earlier the treatment, the better.

According to the WHO, if neglecting untreated Cataract, which is easily corrected by surgically implanting an inter-ocular lens, glaucoma is the leading cause of blindness [WHO19]. Therefore we will have a closer look at glaucoma to understand vision in general better. Glaucoma usually begins in the periphery and, if untreated, progresses towards the central vision. The perception of the world for patients affected by glaucoma is very different from the healthy population. There are texts, images, and even videos explaining the visual experience of an affected person. Today, immersive VR experiences are possible, enhancing the understanding and even treatment of the disease.

As mentioned, beginning neurodegenerative processes must be detected as early as possible. In the last decades, technological progress in Opthalmology advanced immensely from improved slit-lamps to Optical Coherence Tomography (OCT) and Standard Automated Perimetry (SAP).

Especially SAP is an important indicator for the diagnosis and is an essential part of tracking the progression of neuro visual impairments. Through the introduction of ever-improving algorithms, SAP got faster and more reliable. However, most perimetry is performed subjectively and not objectively. Therefore, it is hard to perform perimetry on mentally impaired, small children, or animals. The advances of VR technology, Brain-Computer Interfaces, and AI might introduce a new kind of device to clinical decision-making in ophthalmology. Combining these emerging technologies and applying them to real-world problems in visual perception and ophthalmology is the motivation of this book.

1.2   Definition of Research Questions

The motivation of the last section will be condensed into a set of research questions that will lead through this dissertation. The indicative and fundamental questions are 'how does human vision work?' and 'how can we interact with vision using technology?'.

As already mentioned, the investigated trending topics are Virtual Reality (VR), Brain-Computer Interfaces (BCI), and Artificial Intelligence (AI). The question guiding through this thesis regarding those technologies is 'Can we use VR, BCI, and AI technologies in ophthalmology?'. As with most rather new technologies, there is much potential to improve the current status quo. As an engineer, one is an inventor who is always asking the question 'What are the pain points?', with the intent to overcome them by transferring and inventing. To do this, we need to understand what we are working with and the starting conditions. Therefore, the initial research question is:

Research Question 1: How can we use technology to bridge different visual perception stages from the biological to technological perspective and vice versa?

Because in order to advance research in vision, we need to understand visual perception and its steps. There is the biological side and the technological perspective towards vision. Furthermore, we want to discover how we can use VR, AI, and BCIs to interface between humans and computers.

To fully understand a condition and a situation, a computer-designed simulation can help to broaden our understanding. Therefore, the second research question is:

Research Question 2: How can we simulate a visual perception experience with the help of computer technology?

By using VR technologies, immersive visual impressions can be presented to a user. We want to use suitable tools to create content that gives the impression of looking through another person's eyes. Such simulations will provide valuable

insights into vision in diseased cases and will help gather information on the general topic of visual perception.

After we learned how to re-create visual experiences, we will have to look at how visual understanding of data or pattern recognition is performed in artificial intelligence. This field has provided much innovation in recent years. This book wants to utilize it to understand further how vision works in computers and how we can benefit from computer-aided pattern recognition. Therefore the third research question is:

Research Question 3: How can Deep Learning/AI technologies be utilized in pattern recognition for vision experiments?

The focus of the research question is to facilitate available computer vision and artificial intelligence methodologies to fit the needs of a classification system.

Ultimately, the last question will combine the findings of all previous questions into a complete system. We can define the research question as the following:

Research Question 4: How can we build an objective system that can perform perimetry using a Brain-Computer Interface? What are the possibilities and limitations of the technology?

Therefore, this research question focuses again on the fundamental question of the detection of 'what is where', hence perimetry. The empathy of the research question is that we combine the research to an objective system, meaning that it can perform perimetry without conscious interactions from the patient.

2 Fundamentals and State-of-the-art

> Start by doing what's necessary; then do what's possible; and suddenly you are doing the impossible.
>
> *Francis of Assisi*

This chapter provides an introduction to the fundamental technologies and concepts on which this dissertation is based. First, an overview of the biological and physiological structure of the human visual system is provided. Visual processing starts in the eye and ends with the processing in the brain. This thesis will focus mainly on the visual field. Cues about this parameter can be measured using brain-computer interfaces like an electroencephalogram (EEG). The most significant influence on EEG experiments has the acquired signal quality. Signal quality in passive, non-invasive technologies is strongly affected by the analog-digital converter as well as its amplifier and the used electrodes. In this book, new electrodes that can be additively manufactured (3D printed) will be introduced. Therefore, this chapter will provide a quick introduction to rapid prototyping, as well. As this thesis revolves around visual information processing in the brain, state-of-the-art brain-computer interfaces are discussed. Lastly, the state-of-the-art in virtual reality for vision loss is examined, as well as typical methodologies for brain signal analysis.

2.1 The Human Visual System

The following sections give an introduction to the human visual system. First, the anatomical structure of the human eye, its anterior and posterior segment, is presented. Then, we will focus on the structure of the retina and its information processing capabilities. We will discuss how information is transported through the visual pathway and finally processed in the brain's occipital lobe.
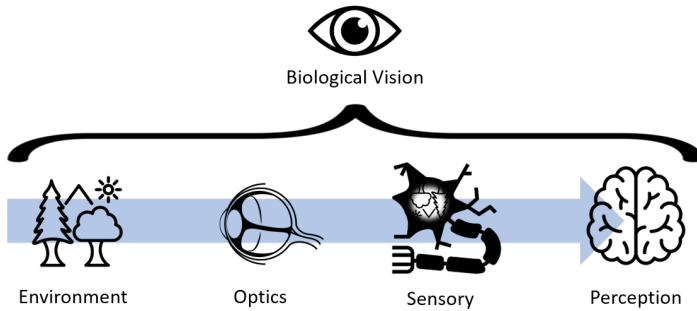
Figure 2.1: The figure shows the structure of 'biological vision', which is the general concept of visual perception in a living being. The optics of the eye refracts light from the environment. Then it interacts with the retina and its rods and cones. Finally, the information is first further processed in the visual cortex and later in other brain areas.

2.1.1 Structure of the Human Eye

In ophthalmology, the eye is often subdivided into its anterior and posterior segments. The principal role of the anterior segment is to manipulate incident light so that it is focused on the retina. There are two main protagonists for bending light: The cornea and the lens, which have an optical power ($D = \frac{1}{f}$) of about 40 and 20 dpt (diopters, $1\text{dpt} = \frac{1}{m}$), respectively [Oys99]. A healthy human eye in a relaxed state has its focus point (f) in infinity, which means that parallel light beams that are entering the eye are focused on the retina. By tension of the ciliary muscles, the zonular fibers will be less tensed, leading to an increase in the refractive power of the lens. This changes the focus point from infinity towards a closer point. This process is called accommodation. With the help of accommodation, we are able to perceive close-by objects sharply. A more detailed insight into the refractive features of the eye will be presented in Section 2.1.2. Another influencing factor to vision is the pupil size defined by the iris diaphragm. Depending on the light intensity in the environment, the iris is adapting. This is protecting the retina from high light intensities, which might damage it. Furthermore, it generally improves visual acuity by reducing the aperture diameter. [Oys99, Hub95]
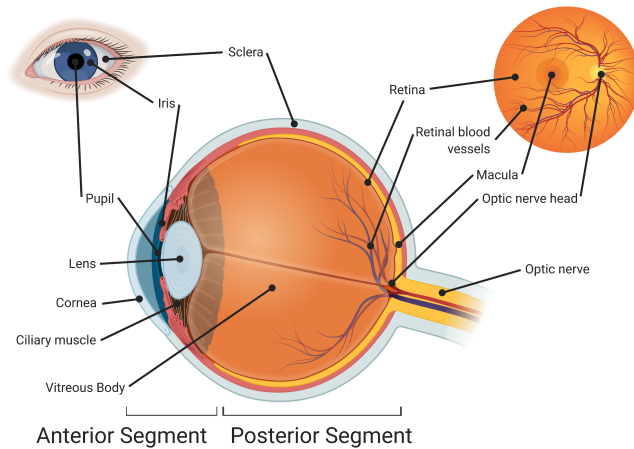
Figure 2.2: The figure shows the anatomical structure of the human eye. Incident light is focused through cornea and lens onto the retina. The excited retina transfers information through the optical nerve to the visual cortex.

Source: *Created by the author. Created with BioRender.com; Inspired by [Oys99].*

2.1.2 Optics of the Human Eye and Visual Acuity

Visual acuity is a property of vision. It is the measure what level of detail a person can perceive visually. In the example of a viewed image, high visual acuity is the ability to discern small features. In contrast, poor visual acuity means that it is hard to distinguish details. There are several ways to denote visual acuity. One is the unit of visual angular resolution in arc minutes. The average human perception is at an optical resolution of one arc minute, which corresponds to $\frac{1}{60}$° [Oys99]. In a nutshell, visual acuity is the quantitative performance variable of central vision. In the following, the most common notations for visual acuity will be introduced. The **Snellen Acuity** or Snellen Fraction named after its inventor Herman Snellen a Dutch ophthalmologist

who studied visual acuity in the 19th century. The Snellen Fraction (SF) is defined by

$$SF = \frac{\text{maximum distance at which letter was identified}}{\text{maximum distance an average subject can identify the same letter}}.$$
(2.1)

Visual acuity is usually tested at a distance of 6 meters or roughly corresponding 20 feet (6.096 meters). Therefore a SF of 1 is often referred to as 6/6 or 20/20 vision. Meaning that the subject correctly identified letters from a distance of 6 meters, just like a normative subject would. Analog a subject that scores 3/6 is only able to discern the letters when moving to a distance of three meters in comparison to the normative 6 meters.

When not referring to the Snellen Fraction as a fraction but as the calculated quotient, it is called **Decimal Acuity** (decVA). With healthy vision at a value of $\frac{6}{6} = 1.0$, a higher decVA corresponds to a better vision, whereas a lower value indicates worse visual performance.

However, one of the most used units in optometry publications is the logarithmically scaled **LogMAR Acuity** where MAR stands for the minimum angle of resolution in arc minutes. Therefore LogMAR can be calculated as

$$LogMAR = log(MAR) = log\left(\frac{1}{decVA}\right).$$
(2.2)

The question arises of what the maximal resolution of the human eye is. Therefore, the minimal angular resolution (**MAR**). When considering the general concept of vision presented in Figure 3.8, many requirements need to be met to get an estimate of the MAR. First, the environment needs to provide the level of detail. On a second step, there are optical limitations. Then the sensor has to have small enough individual perception elements (e.g., photosensitive receptor cells). Finally, the perception needs to be able to process the data delivered by the sensory input. We assume a detailed enough environment for this thought experiment because even a not detailed environment is getting detailed by increasing the distance to it.

The lens of the eye is a dynamic system that can perform accommodation by contracting the ciliary muscles. When the eye is relaxed, the lens has lower refractive power, leading to objects from infinity being focused on the retina.

If this works correctly, we are talking about the **emmetropic** eye. Most people, however, are not emmetropic. If incoming light rays are focused behind or in front of the retina, it is called **ametropia**. **Myopia** or commonly referred to as nearsightedness, is the result of light being focussed in front of the retina. Therefore, objects far away seem to be blurred, whereas closer objects can be seen sharply by accomodating. Causes can be the eye-lens, cornea (refractive myopia), or just a too-long eye (axial myopia). Generally, myopia can be treated by using a concave lens which adds some divergence of light rays which again moves the focus point back to the retina. The opposite condition is that light is focused behind the retina. That is called **hyperopia** or farsightedness. It can be due to the eye-lens and cornea (refractive hyperopia) or a too-short eye (axial hyperopia). It can be corrected using a convex lens. As the visual focal point for hyperopic people is not in infinity but beyond (divergent), it is possible for the affected to accommodate to perfect vision for far objects. This, however, is straining to the ciliary muscles and can lead to headaches and other complications.

The third major eye condition results in not perceiving close-by objects sharply, and it is called **presbyopia**. Usually, people at age 40 and over start to notice it, and eventually, everyone gets it. Therefore, presbyopia is a problem not with the optical power but with accommodation. The scientific consensus is that the crystalline lens is getting more rigid with age, influencing the accommodation process. To understand this, we need to have a closer look at the accommodation process itself. Herman von Helmholtz [VO55] described in 1855 that in a resting state, the lens is accommodated due to relaxed zonular fibers. If they contract, the lens is pulled, leading to a decrease in lens thickness and moving the focus point further. Today, there are diverse views on the deformation of the lens, with some models even stating an opposite effect of fiber-tension [SHH93]. However, according to Burd et al. [BJF99] the basic concept remains as formulated by Helmholtz: By activating fibers, a reduction in refractive power is achieved. Just like hyperopia, presbyopia can be corrected using a lens with positive power. However, if the patient was emmetrope in infinity and has to wear glasses for the near vision, he will not perceive far objects sharply anymore while wearing the glasses. Therefore, myopic people with presbyopia need two pairs of glasses or varifocals.

Myopia, hyperopia, and presbyopia have all in common that they can be corrected using a defocusing element in the glasses. Another form of refractive

error is **astigmatism** often originating from irregularities of the cornea. In such cases, the cornea is not exactly circular-shaped but toric, leading to two meridians orthogonally to each other. Subjects with astigmatism have blurry vision at all distances, which can only be corrected using spherocylindrical lenses. It is important to notice that none of the mentioned eye conditions are considered diseases as long as they are in non-extreme forms.

Now that we covered the general function of the human optical system, we can also simulate it in a computer environment. This fundamental piece of research was already published by the author in [11] and will be presented in short here. Figure 2.3 shows the basic principles of light ray tracing in the simulation of a computer-generated human eye. It is shown how an illuminated point in space is reflecting light towards an eye model (a). The eye model is refracting the incoming rays on each surface, beginning with the cornea, then the lens entry and lens exit plane (b). The rays are collected on the retina in a third calculation step (c). In a final step (d), through rasterization, an image on the retina is created. Now we simulate the pass-through for an individual point onto the retina. The next step would be to perform the same for a point cloud, hence an image. Figure 2.4 shows an exemplary computer simulation model incorporating the USAF 1951 test pattern[1] as an example. The eye model used in the simulation is based on the Navarro eye [NSB85]. The Figures 2.5 and 2.6 are displaying the simulated effects of the optics of the eye to the projected image on the retina. For this simulation, multiple test images were placed in front of the computer-simulated Navarro eye with an angle from 0° to up to 35°. It is especially interesting to notice the effect that pupil-dilation has on collected light and the sharpness of the image. Furthermore, the more peripheral the image gets, the blurrier. More computer simulations of eye models can be found in the publication [11] by the author of this book.

2.1.3 The Retina

The retina is a multi-layered neural membrane that lines the inside of the eye. Regarding vision, it is a combination of a sensor and an initial computational unit. The retina transforms incoming light into electrical impulses. This is

---

[1] A microscopic optical resolution test pattern with different resolution test targets.
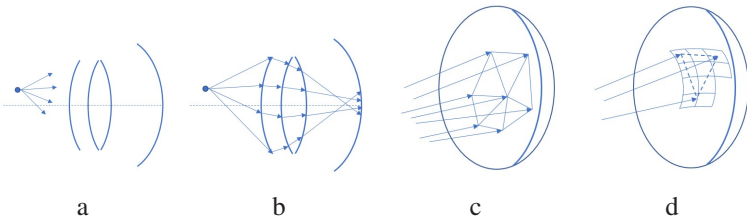
Figure 2.3: The figure shows a physical ray-tracing model: In (a), rays in all directions are produced from point sources of lights of the image. In (b), the optical surface model of a human eye is presented. Those refracted rays are hitting the retina (c). Through rasterizing, final pixel values can be obtained and, therefore, the intensity of light hitting the retina in specific locations.
Source: *Previously published by the author in [11].*

done by the photoreceptors located on the highly pigmented retinal pigment epithelium (RPE). There are two main types of photoreceptors: Rods and cones. The approximately 5×10^6 cones provide high-resolution color vision and are located primarily in the foveal region of the retina. The about 10^8 rods are responsible for dim-light vision. These sensing cells are connected to bipolar, amacrine, and horizontal cells before they are passed to the brain by approximately 10^6 retinal ganglion cells [Zha14]. From an engineer's perspective, one notices that the retina's wiring is suboptimal as the wiring to the light-sensitive photoreceptors lies on top of the RPE. Therefore the nutrition transport to the cells is producing some levels of occlusion. However, this seems not to be an evolutionary disadvantage or suboptimal, as this design works well for all vertebrates. In cephalopods, the wiring is not on the photoreceptors but beneath them. However, it is hard to compare the vision of underwater cephalopods with that of vertebrates. Overall, nature's design does not need to be logical from the engineers' point of view, but it has to deliver evolutional fitness [Ayo96]. Meaning that this wiring might even provide light-guiding properties like a reduced scattering of light within the eye and optimization of light intensity at the receiving receptors, ultimately increasing the signal-to-noise ratio [AJM+11]. Moreover, even if retinal blood vessels, compare Figure 2.2, are occluding vision in some parts of the retina, it remains unnoticed as human vision is insensitive to static (not changing) inputs. This is due to the processing of the bipolar cells in the retina. They connect multiple

13
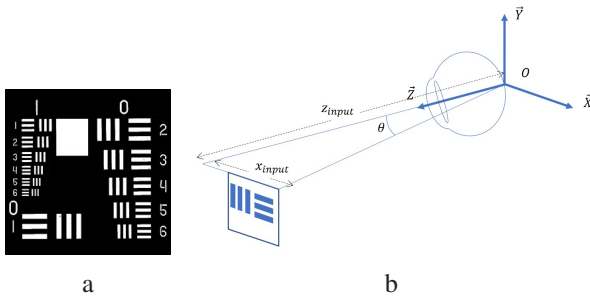
Figure 2.4: A modified image file of the USAF 1951 test pattern was used as an input stimulus
(a). The size of the point cloud image is adjusted so that the angular resolution of the
smallest line pair is at one arc minute. (b) shows the orientation in the world coordinate
system of the human eye and the input stimulus.

Source: *Previously published by the author in [11].*
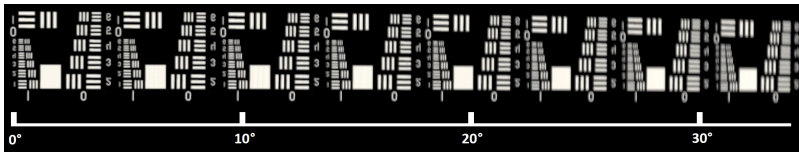
Figure 2.5: The figure shows the results for a simulated accommodated Navarro eye. Several
test-stimuli were aligned next to each other to cover up to 35°of the visual field.

Source: *Previously published by the author in [11].*

photoreceptors and only pass on gradient information. The bipolar cells, in
turn, are connected to the ganglion cells. Those are leaving the eye through
the optic nerve head. This optic nerve head is solely made from ganglion cells
and retinal blood vessels, as it cannot perceive any light due to the lack of
photoreceptors; it is also referred to as the blind spot. Usually, humans do not
notice their blind spot, this is due to the same effect, that we only perceive
non-static images, and the brain fills in those areas occluded by vessels and
the blind spot. In some neuro-visual diseases like glaucoma, this effect leads
to late discoveries of the disease, as dead photoreceptors are not noticed easily
by affected patients.
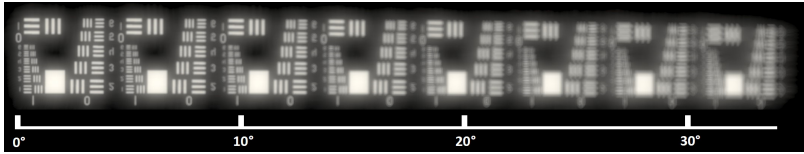
Figure 2.6: This simulation shows the accommodated Navarro eye. However, this time it has a dilated pupil. The acuity quickly drops for stimuli further than 10°. Especially further peripheral stimuli are resolved badly. However, more light is captured by the dilated pupil leading to overexposed images.

Source: *Previously published by the author in [11].*

Let us have a closer look at the layers of the retina. Several layers in the human retina can be distinguished. With modern technology, those layers can be inspected even in vivo using an optical coherence tomography (OCT) device. An example of the author's retina and a schematic of the layers is provided in Figure 2.7.

Especially when considering brain-computer interface experiments, we need to try to understand how the processing steps in the retina work. For color vision, three types of cones are distinguished in humans, the S-, M- and L-cones. S stands for short, M for medium, and L for long-wavelength receptors. According to Bowmaker et al. [BD80] the S-cones have their sensitivity peak at 420 nm (blue), M-cones at 534 nm (green) and the L-cones at 564 nm (red). In between those, the rods have their maximum sensitivity at a wavelength of 498 nm. When light hits these receptors, it interacts with rhodopsin, a photopigment protein molecule. The result is a neural stimulus that is the beginning of what we perceive as seeing.

The use of these photoreceptors is significantly influenced by ambient light. A distinction is made between **scotopic vision**, which is used in almost complete darkness, like starlight. The rods are mainly active, and the iris has its maximum aperture (of about 5-7 mm) to collect as much light as possible. In an intermediate stage, called **mesopic vision**, both rods and cones are responsible for vision, and this happens in poor lighting situations like, for example, moonlight. In **photopic vision**, the rods are supersaturated, the vision is mainly shifted to the cones, and the iris closes its aperture down to under 2 mm.
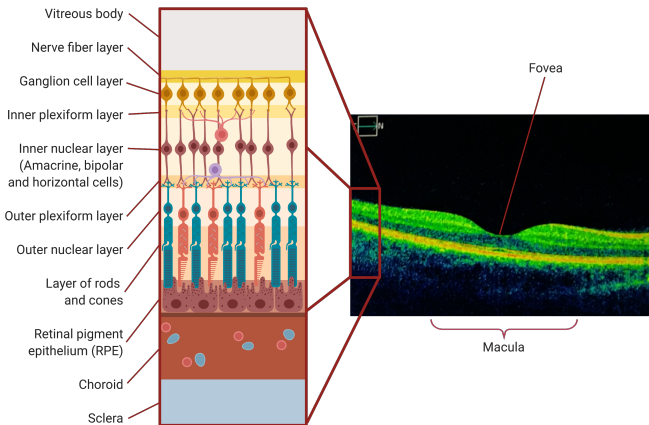
Figure 2.7: On the right side, a cross-section of the in vivo eye of the author is presented. The scan was retrieved from a ZEISS Cirrus 5000 OCT device. The area of the macula and the fovea in its center is shown. On the left side, the individual layers of the retina are presented.

Source: *Created by the author. Created with BioRender.com; OCT-scan recorded with a ZEISS Cirrus 5000, retina depicted is the authors'. Identifiers are based on [UCW⁺19, Oys99] and BioRender.com.*

The place with the highest density of cones in the retina is called the macula. It is positioned about 5 ° temporal of the anatomical axis. This ca. 5.5 mm (18°) large field is responsible for reading, recognition of structures and colors. The center-most ca. 0.5 mm (ca. 1°) is called the fovea. It is the area of highest acuity vision on the whole retina with many cones but no rods. It can be defined that this point is the oculomotoric zero point in vision which also defines the main viewing direction.

Additionally, to rods, cones, and bipolar cells, which we already covered, there are amacrine and horizontal cells. The horizontal cells are getting input from various photoreceptors, integrating signals, and adjusting information transported to the bipolar cells by regulating photoreceptor cell activity. The amacrine cells are similar to this, as they also regulate the activity of other cells. But in their case, bipolar and ganglion cells. They do this by integrating multiple bipolar cells. [UCW⁺19, Oys99]
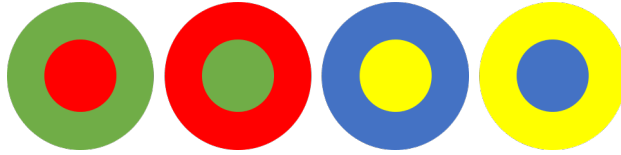
Figure 2.8: Center surround structure of retinal photocollectors. Red is only perceived in contrast to green, whilst yellow with blue in one receptive field.

Like already mentioned, the human eye is blind to static input, but this is not the complete story. The methodology of the sensor can be described as a 'center-surround'-detector. Multiple retinal cones together form a **receptive field**. Many bipolar cells are interconnected in an antagonistic center-surround fashion, like shown in Figure 2.8. The center can be on or off activated, and the surround is always reciprocal to that. If there is no light hitting the cells, the ganglion cells of the receptive field will not propagate any information. If the light is hitting the center and the surround, there is a weak response, and if the light is hitting the on-center while there is no light on the surround, the ganglion cell is firing rapidly. This is the very basic encoding of information performed by the retina.

2.1.4 Vision beyond the Retina

The visual pathway is functioning as the highway for visual information. Retinal ganglion cells are leaving the eye through the optic disc. The structure of the wiring is depicted in Figure 2.9. The right side of the figure presents the visual field information (black) for cross-sections of the optic nerve. Specific processing and controlling steps like pupillary reflexes are already performed before entering the primary visual cortex at the occiput [Csi05]. In the center of the head, the optic nerves are crossing in the optic chiasm (3). After that, they are separated to a left and a right hemisphere of vision, and are bundled in the lateral geniculate body, also called the lateral geniculate nucleus (LGN). The LGN is responsible for where we look and that both eyes are converging at the same position. Furthermore, it also performs the sensor fusion of auditory and visual perception. The LGN rearranges and separates the nerves to an upper (5) and lower (6) visual field. Finally, they are connected to the primary visual cortex, known as Brodmann area 17, or V1. It is located in the occipital gyri,
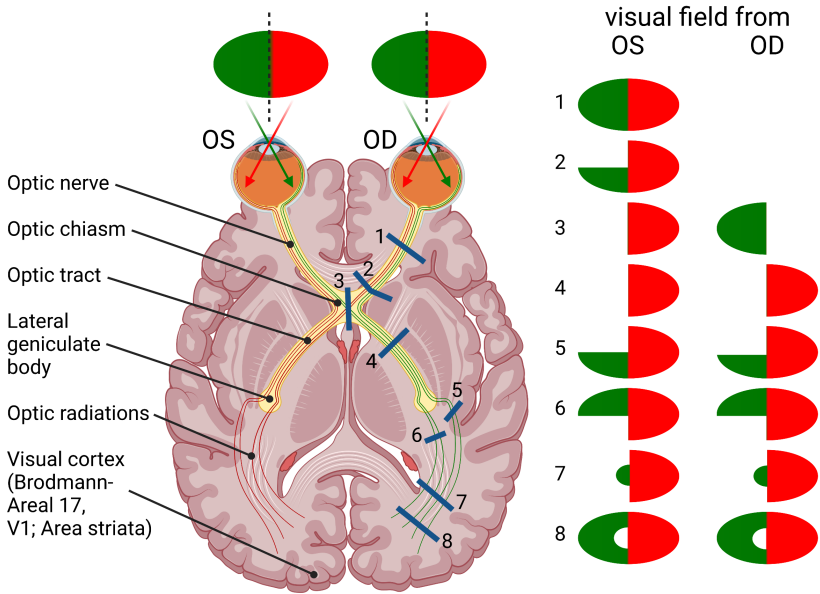
Figure 2.9: The figure shows a sketch of the visual pathway in the brain. Multiple sections (dark blue) are presented. The graphics on the right display active vision if a cut is performed. The input of the individual oculus dexter and oculus sinister (OD/OS) retinal signals are crossed at the optic chiasm, where the visual field is split into a left and a right hemisphere (3). Notice that the left side of the brain receives the information from the right side visual field and vice versa.

Source: *Created by the author. Created with BioRender.com; Inspired by [Lan00].*
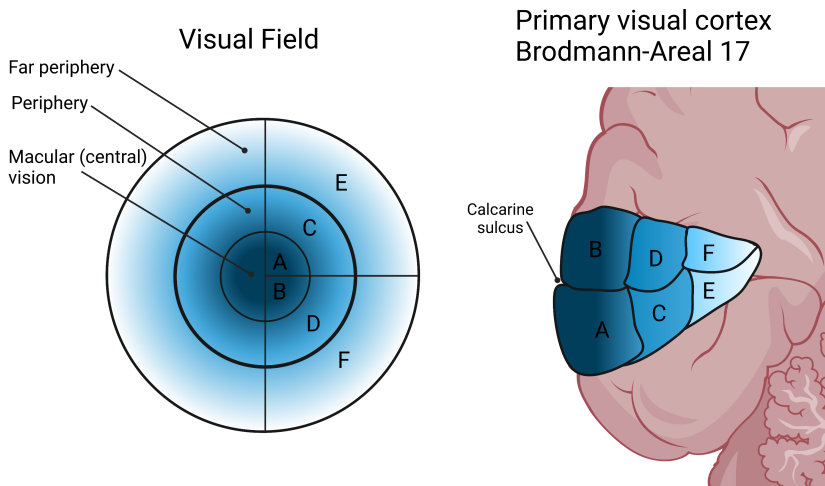
Figure 2.10: The figure shows the primal visual cortex retinotopic mapping. On the left side, the visual field of a human is shown. The right side shows the primary visual cortex of the left brain hemisphere. The letters represent which section of the visual field is being processed at which topological position. The macular (A,B) information takes up more brain area in relation to size in the visual field than the periphery.

Source: *Created by the author. Created with BioRender.com; Inspired by [PWK18].*

where all retinal information is retinotopically mapped topologically. Central vision resulting from this is mapped closer to the skull, whereas peripheral and far peripheral vision is processed deeper within the brain. This is illustrated in Figure 2.10. Furthermore, central, macular vision gets more dedicated brain area in comparison. This is called cortical magnification. The central 10° of the retina corresponds to only 2 % of the entire visual field and covers more than 60 % of the visual cortex. [HH91]

The average delta time from receiving light on the retina until it reaches the visual cortex is, according to Kaschke et al., approximately $95 - 115$ ms and the complete perception process takes about 300 ms [KDR14].

2.1.5 Processing of Visual Information in the Brain

> Vision begins with the eyes, but it
> truly takes place in the brain.
>
> *Fei-Fei Li*
> *(Director, Stanford AI Laboratory)*

Everything we experience is ultimately the activity of the approximately 86 billion neurons [ACG$^+$09] in the brain. It can be divided into sections, handling different operational tasks. For vision, the primary visual cortex located in the occipital lobe is where visual information is processed in an initial step.

The visual cortex is located within the neocortex, which is involved in higher-order information processing and sensory perception. It can be separated into six cortical layers, wherein layer IV can be subdivided into IVA, IVB, IVCα, and IVCβ. It is a layered mosaic structure that acts as sensors for line stimulations (slits, bars of light) of any given orientation. When within one column a cycle of 180° is completed, it is called a hypercolumn. Furthermore, there are specialized blobs of vision for different wavelengths.

There are three major known functions of Brodmann area 17, the primary visual cortex:

- Combine information from two eyes for three-dimensional stereo-vision.

- Divide information into short line segments with different orientations to further recognize motion and forms.

- Separate information related to color from motion and forms.

After this, the visual information is further processed in the parastriate cortex (Brodmann area 18) and later in the peristriate cortex (Brodmann area 19). The belt regions around the visual cortex are responsible for further higher-order processing. These are processes that happen in parallel and for different modalities of the visual image. [Oys99, KDR14, Csi05]

Although the visual cortex and the perception of the visual world are studied intensely by scientists all around the world, many functionalities of vision are a mystery to science still today [MM07].

From what we already know about visual processing in the brain, information technology has learned a lot about image processing. In the last couple of years, scientists and engineers could hugely improve vision-related tasks in computer algorithms by re-implementing the general functions of neural structures in the visual cortex of mammals. The main building block is the convolution of information [FS10], but other principles are also very alike, for example recurrent and skip-connections. Hence convolutional neural networks perform outstandingly at all kinds of image processing tasks [GBC16]. Details about these vision-inspired AI networks will be discussed in the machine learning section (Section 3.2).

2.2 The Human Visual Field

In this section, we will be introducing the concept of the visual field. In general, the visual field is defined as the environment that can be perceived visually by an observer while focusing on a fixed point in space (fixation point). Besides visual acuity (as described in Section 2.1.2), the visual field is the most significant other parameter considering the eye's functional status. The visual field examination is called **perimetry** and a device that is performing such a task is therefore called a **perimeter**.

2.2.1 Causes of partial vision loss

To better understand the visual field, let us first have a look on impaired visual fields. **Scotoma** (from ancient greek - darkness) is the condition where parts of the sensibility of the visual field are drastically reduced. Naturally, this is present for every human at the optical nerve head, where no visual information is collected. However, often a reduced field of vision is an indicator for a neurological or visual disease. Sometimes brain tumors or other issues within the brain have a direct impact on the visual field. It is even possible to derive the position of a problem from the pattern of the visual field deficit. Exemplary Figure 2.9 shows the individual functional sections of vision throughout the visual pathway. For example, a tumor pressing on position (4) will result in a loss in the right-hand side visual field in both eyes. Because of patterns like these, ophthalmologists often discover neurological problems within the

brain [MPV20]. Diseases like glaucoma, diabetic retinopathy, and age-related macular degeneration (AMD) are other common causes for visual field loss.

As an example, let us focus on **glaucoma** (from [Oys99, Qui06, WAM14]), which is a damage of the optic nerve. Due to an excessively increased intraocular pressure and disturbed blood circulation in the eye, this disease leads to a gradual loss of nerve cells in the retina due to high pressure. This intraocular pressure is regulated by the supply and drainage system of aqueous humour (the watery fluid filling the eye) within the vitreous body (compare Figure 2.2). New aqueous humour is formed in the ciliary villi and then released into the eye's posterior chamber. From there, it can also reach the anterior chamber of the eye via the pupil. At the pupil, the trabecular meshwork is located, which leads to Schlemm's canal, which is the drainage of 'old' aqueous humour. A disturbance of this inflow and drainage system can increase intraocular pressure, ultimately leading to dying photoreceptors due to malnutrition.

Glaucoma is the second major leading cause of blindness worldwide (after untreated cataract). It is estimated that there are almost 80 million affected by glaucoma in 2020 worldwide [Qui06]. As a side note, it disproportionately often affects women and Asians. Treatment methods include medications, laser therapy, and incisional surgery.

The disease symptoms usually originate in the periphery and slowly progress towards the center of the visual field. Affected patients are often unaware of the disease in the early stages. However, once lost, retinal photoreceptors cannot be restored with today's medicine and technology. This is why the early detection of glaucoma is crucial.

2.2.2 Perimetry of the Human Eye

Perimetry is a method to derive information about the status of the human visual field. It is also called 'visual field examination'. It measures where an observer can perceive visual information. Traditionally a fixation target is used, and different positions relative to it are assessed. There are different stimulation approaches to perimetry. A common method is examining differential light sensitivity (DLS), which is called static perimetry. Other perimetry approaches quantify color discrimination and peripheral motion vision (kinetic perimetry). The patient usually watches a uniformly illuminated bowl (L_{bg}), often at 10

cd/m^2 so that light intensity for photopic vision is present. As a result, only the contrast distinguishing ability of the cones is tested. The procedure is that several stimuli at various stimulation intensities (L_{th}) are presented at different observational positions until the whole retina is analyzed. A concept of this is presented in Figure 2.11. As many positions on the retina have to be tested, a perimetry test is relatively long and tiring. However, newer perimetry examination routines like *SITA-fast* [BH98] still have high reliability but only require about a third of the time needed compared to the *Full Threshold strategy*. The differential light sensitivity (DLS) of a patient at a stimulation positon can be calculated as:

$$\text{DLS} = 10 \log \left(\frac{L_{max}}{\Delta L_{th}} \right), \text{ with } \Delta L_{th} = L_{th} - L_{bg} \tag{2.3}$$

L_{max} is the maximum luminance achievable by the perimeter. With this scheme, a matrix of variables (notated in decibel) describing visual performance can be created. The current gold standard for such notations is the Humphrey 24-2 field [10]. An example for a 24-2 scan is displayed in Figure 2.12. The scan shows a typical pattern of advanced glaucoma. A value of '<0' in this case means that even the brightest stimulus was not perceived anymore. Whereas a 0 means that the brightest stimulus was perceived. The maximum achievable value is 40. The top left corner shows signs of advanced glaucoma, whereas the 0 on the right side is located at the optic nerve head, which is a blind spot and, therefore, cannot perceive light stimuli.

Perimetry testing aims to discover deviations from the normal hill of vision defined by a reference that is matched for age. Ultimately, the procedure is performed to diagnose and track glaucoma progression [BE02, NM14], other retinal diseases, and neurological disorders discoverable within the visual pathway. Furthermore, it is also used to certify fully functional visual performance for jobs where it is crucial, e.g. as a pilot or truck driver.

In the case of glaucoma and other slowly progressing visual diseases, standardized automated perimetry (SAP) is used to track the progression of such diseases with high accuracy.
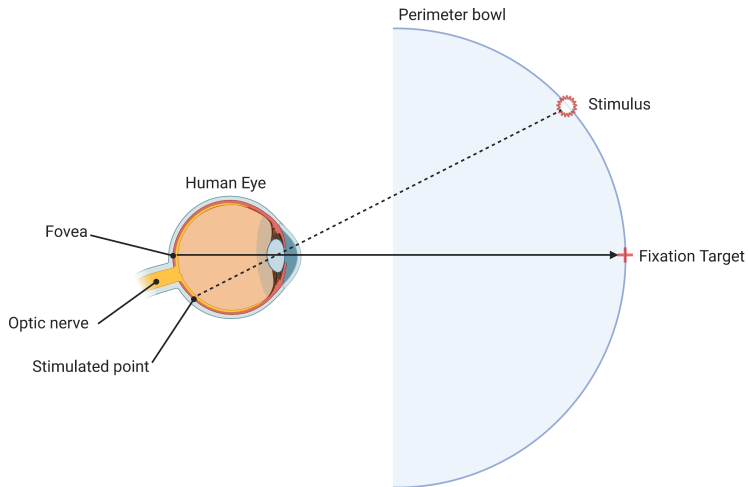
Figure 2.11: The figure displays the general concept of perimetry. An uniformly illuminated surface is presented to the observer. Stimuli are presented at different position in the visual field. If a stimulus is perceived, feedback is given by the observer.

Source: *Created by the author. Created with BioRender.com*

2.2.3 State-of-the-art Objective Perimetry

The SAP already offers a great possibility to measure visual fields. However, it only offers a subjective measurement value of the visual field. This is due to:

- Patient fatigue if the measurement times are too long and tiring.

- The learning effect: The more often a patient does the test, the better he becomes at it.

- Misalignment of the patient or wrong fixation on the target.

- Any higher mental processes which might intervene with the measurement.

```
                      <0   0 | 2    9
                  <0  <0   0 | 12   9   18
              <0  <0  <0  <0 | 7   12  14  22
  30°     <0  <0  <0  <0  18 | 30  14  19  25    30°
          <0   6  <0  16  29 | 31  29   0  28
              18  21   9   4 | 10  11  25  25
                  22  18  14 | 23  27  26
                      26  26 | 28  29
```
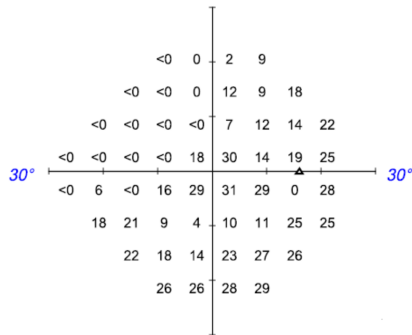
Figure 2.12: Visual field, represented as a 24-2 scan. The values represent the perceived intensity. The triangle on the right marks the position of the blind spot. The figure displays around 60° of the visual field. The sample was recorded with a ZEISS Humphrey Field Analyzer.

Source: *Already published by author in [10].*

Moreover, a patient who cannot understand the test (like young children, handicapped, or animals) cannot be screened. To overcome those drawbacks of SAP, techniques were developed that directly measure visual performance and the visual field. The most prominent ones are:

- Measuring electrical activity right on the retina using retinal evoked potentials (REP), which are usually measured by an electrode directly on the eye of a patient.

- Observing involuntary movements of the eyes of a patient by using subliminal stimulation.

- Measuring electrical potentials at the occipital lobe.

The advantage of REP are that they measure electrical activity on the retina with good performance. However, an electrode has to be placed *on* the eye, which is at least unpleasant for the patient. Furthermore, if there is a problem with the optical pathway or the visual processing in the brain, the REP are not suitable. The observation of bodily reactions like a pupillary contraction is another method to perform perimetry. Studies showed that it generally works; however it is not necessarily on a comparable level to SAP [TF07].

As vision begins in the eyes but truly takes place in the brain, we will focus on processing visual information within the brain utilizing brain-computer interfaces in combination with data-driven algorithms.

2.2.4 State-of-the-art Objective Perimetry using Brain Data

By directly using the visual processing data from the brain, the complete visual pathway can be examined. In 2010 Goetze [Göt10] published his dissertation about objective multifocal perimetry using EEG and pseudo-random patterns. He projected the created pseudo-random pattern on a large perimeter bowl. The stimulus was not a checkerboard or similar but white circles on a black background, which had an eccentricity of 5°and a diameter of 4°. The used projector had a refresh rate of 120 Hz at a resolution of 1280×1024 pixels. He performed two experiments. The first one with a total of 37 healthy individuals in the age range of 21 to 55 years. He used wet electrodes with an electrode paste, leaving residues in the hair of the subjects. In his second experiment, he used a fixation target with varying positions to decrease alpha-responses within the brain. He used five electrodes on the occiput in his experiments, which he cross-correlated with the pseudo-random input signal. The algorithms were developed in Matlab, which complicates cross-platform and real-time capabilities compared to more performance-oriented languages like C and C++. The work of Goetze is very focused on the mathematical part of a generation of pseudo-random patterns and the system aspect of projection perimetry using a large perimeter bowl. However, he could not produce reliable results from his setup using the perimeter bowl. He states that this might be due to the too little luminance of the projector system.

A more recent EEG-based perimeter system was introduced by Nakanishi et al. in 2017 [NWJ+17]. Their work is currently being further developed by the startup NGoggle Diagnostics. The central part of their system is a BCI fully integrated into a VR device. They use dry electrodes to obtain brain signals from six electrodes (Pz, PO4, PO3, O1, Oz, and O2) on occiput, and additionally, they recorded an electrooculogram from 4 channels. Their published study [NWJ+17] included 62 eyes from 33 glaucoma patients and 17 healthy control patients. They tested in total 20 sectors within the central 35° field of view while using flicker frequencies between 8.0 and 11.8 Hz. By calculating the canonical correlation analysis, they can understand the

correlation between glaucoma and the recorded strength of steady state visually evoked potentials (SSVEPs, see Section 2.3.3). They tested their setup against a Humphrey Field Analyzer from Carl Zeiss Meditec. A multi-focal canonical correlation analysis was performed on all data from a subject. The team around Nakanishi achieved a specificity of 80% with a sensitivity of 85% for distinguishing glaucoma in an eye globally from a healthy eye. However, they could not identify particular fields where vision loss occurred or perform any analytics on the intensity of vision loss. Furthermore, no data is provided on how strong the glaucoma cases in the patient group were and how well the system performs for early glaucoma in the periphery. Also, they do not provide any data about their used threshold and whether the threshold was achieved by overfitting the training data. Apparently, no train-test-validation split was performed. Nakanishi et al. state that their research is intended as a proof-of-concept of the feasibility of the NGoggle as a portable objective measurement device, which ultimately should enable telemedical home assessment of visual conditions.

2.3 Brain Computer Interfaces

The term *brain computer interface* (BCI) is the general expression for a system directly communicating with the human brain. Including input towards a computer but also to the brain from the computer. One could argue that every computer we interact with basically needs some form of a brain-computer interface. Ultimately, our brains interpret the information presented on a computer screen, and it is our brains that control the input to the computer by controlling our hands and hence the input through a keyboard. However, when talking about BCIs, the direct interaction with the brain, for example, through electrodes, is meant. A good definition for this is proposed in the paper of Wolpaw et al. [WBH+00]:

> 'A brain-computer interface is a communication system that does not depend on the brain's normal output pathways of peripheral nerves and muscles'

Therefore, a BCI is not dependent on a keyboard, mouse, or other peripheral hardware, controlled by the user's extremities, for input. We can go one step further than Wolpaw and define that a BCI also needs input modalities to be

considered an interface. In theory, all senses can be used to input information into the brain. As this book focuses on vision, the question arises, whether vision is a good way for communication with the brain? The retina is part of the central nervous system (CNS), which is the overall combination of the brain and spinal cord. Therefore, the retina is indeed the window to not only vision assessment but to the whole CNS [LBS13]. Therefore a direct, non-invasive, and high-bandwidth way to couple information directly into the brain is through retinal stimulation with light, for example, from displays. While it is relatively easy to excite the brain using patterned light on the retina, it is still a considerable challenge to directly induce images through electrostimulation at the optic nerve ([DOV03]) or direct electromagnetic stimulation at the visual cortex [JSL⁺19].

In general, there is a difference between invasive and non-invasive BCIs. In recent years the start-up Neuralink introduced implanting-robots [YS19] to directly implement thousands of electrodes into the brain. Their goal is to achieve a symbiosis between technology (like smartphones), complex AI algorithms, and humans. Enabling bidirectional communication between the brain and mobile devices. Today, animal trials were already successful, and results are promising to be adapted to humans in the future [Mus20]. As mighty as this sounds, there is a massive drawback to the technology developed by Neuralink, and it is invasiveness. A complicated implanting procedure, where risks are not yet completly known and understood, has to be performed. Therefore, it is very disputable whether such technology will reach human users in the next decade.

This book will focus on a more safe method of obtaining information from the brain in real-time: Electroencephalography, EEG in short.

For completeness, it is mentioned that there are also other non-invasive BCIs. These are based on effects such as pupillary reactions [MMvdLV16] which are not able to produce high information transfer rates. By applying light, we can use near-infrared spectroscopy (NIRS) [GAFT⁺14] which has a rather bad temporal resolution but can achieve impressive spatial resolutions. Another relatively new technology is functional Magnetic Resonance Imaging (fMRI) [ERW⁺94], an MRI technology that, analogously to NIRS, uses oxygen concentrations in the brain to determine which brain regions are active. However, while the spatial resolution of fMRI is comparably high, it suffers from poor temporal resolution. Another complex approach to overcome those

shortcomings is magnetoencephalography (MEG) [HRMB14], where an external device measures the brain's magnetic field. MEGs need to be perfectly shielded against noise from the outside, as the magnetic fields created by the brain are very small. Furthermore, superconductive sensors have to be used in the MEG machinery, which further complicates its use and ultimately makes the technology very stationary and expensive.

For the examination of vision, high temporal resolution is needed. Therefore only pupillary reactions, MEG, and EEG remain on the shortlist. As MEG is expensive, not portable, and not applicable in most real-world scenarios and pupillary reactions are somewhat limited in what kind of information can be gained from it, EEG is the kind of BCI that we will discuss in this book.

2.3.1 Electroencephalography

EEG is an up-and-coming technology for Brain-Computer Interfaces in all formats. In 1929 Prof. Hans Berger, a German Neurologist from Jena, invented the general principle [Ber29]. In Neurology, he is also famous for discovering the phenomena of frequency blockage of alpha-waves and the emerge of beta-waves after a patient opens their eyes after being in a relaxed waking state. This effect is named after him, the so-called Berger-Effect.

The working principle of EEG technology is the following [Coh17]: Millions of recordable neurons close to the skull are firing each second, leading to ionic currents within the brain. These currents transmit through the skull, which acts as a low-pass filter. The EEG electrodes are placed on the surface skin of the subject's skull. These collect the ionic current and transform it into an electric potential. High-quality, low noise, common-mode rejecting amplifiers are used to amplify the signal. In a final step, analog-digital converters (ADC) are used to obtain computer interpretable digital signals. Analog EEG systems do exist as well but are considered outdated and complex to interpret [VB98].

The main advantages of EEG systems are that they only require a relatively simple setup (electrode, amplification, ADC), the hardware costs are therefore comparatively low, and that they obtain a high temporal resolution. The disadvantages, however, are that the brain signal quality is considerably weakened through the skull and many skin layers. Furthermore, the spatial resolution is limited by the size of the electrodes and the scalp's surface area. Re-

garding spatial resolution, systems are ranging from only one measurement electrode [MPKSEP10] up to 256 channels in dense setups [YTT$^+$13]. However, the drawback of using many electrodes is that they tend to need more setup time and short-circuits are more likely to occur.

Signal quality and signal-to-noise ratio are the most important aspects of EEG systems today. Whilst potential discharges in the brain are in an order of 0.5 to 1.5 mV, the measurable potential on the scalp is only between 10 and 100 μV [Ern05]. There are several methods to improve the signal-to-noise ratio. Today, EEG systems are often equipped with a driven right leg (DRL), which acts as a ground. It is placed on a patient's body part with low biological activity, usually the earlobe. The DRL circuit decreases the potential difference between subject and amplifier and therefore causes a significant reduction in noise [MPG90]. Another method is to clean and prepare the skin of the patient with an abrasive substance. By removing the outer layer of the epidermis, the stratum corneum, which mostly consists of dead skin cells, the skin resistance can be reduced, resulting in a better signal quality [OPB$^+$17].

A sample EEG setup is displayed in figure 2.13. The images were taken in the KIT-antenna lab, which is isolated from electromagnetic fields. Initial experiments from this dissertation were performed in that antenna lab. However, it was concluded from the experiments that normal external fields only have minor influences on measured EEG signal quality and complete electromagnetic isolation is not necessary for EEG experiments.

In the early stages of this technology, users had to be trained for many hours to use it, like the experiments of Niels Birbaumer at the University of Tübingen in the mid 1990 [BGH$^+$99]. The first paradigm shift of EEG-based BCIs was using direct brain reactions to visual stimulation like the P300, which is a potential created in the visual cortex occuring approximately 300ms after stimulation [Joh86]. However, the information transfer rate of P300-based systems are usually beaten by steady-state visually evoked potential (SSVEP) based systems [ZLL15]. We will further discuss this kind of stimulation in Section 2.3.3. In the next section, we will focus on one of the most significant impacts on EEG signal quality: The electrodes transforming ionic current into electrical current.
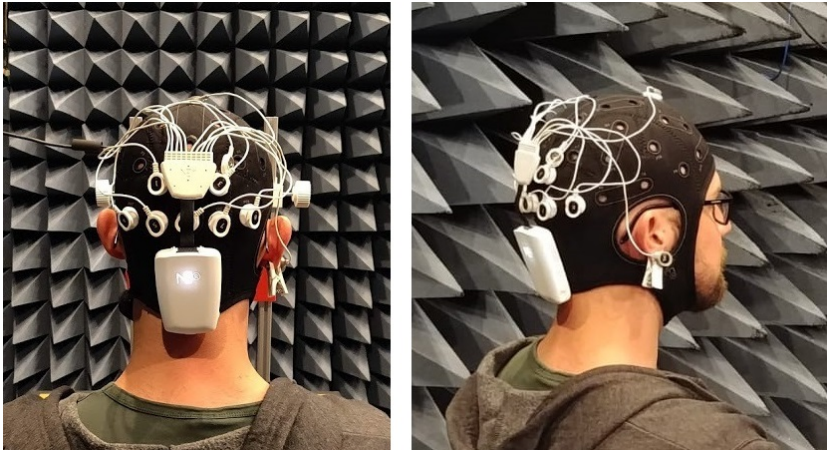
Figure 2.13: The figure shows the EEG set up like it was used during the experiments of this dissertation. Depicted is the author with an electrode cap, measurement-electrodes placed on the occipital lobe, reference-electrode on Fz (top of the head), and an ear-clip-electrode acting as driven right leg (DRL, ground). The combined amplifier, ADC, and wifi interface by Neuroelectrics® is located below the measurement-electrodes.

2.3.2 State-of-the-art Types of Electrodes

The electrodes used in an EEG setup are one of the most crucial elements for obtaining good signal quality. Over the years, different kinds of electrodes

Figure 2.14: Different types of commercially available electrodes. From left to right: Neuro-electrics® Sticktrode (adhesive), Neuroelectrics® Geltrode (wet or hard-gel), and Neuroelectrics® Drytode (dry)

31

emerged. This section will provide a brief overview of the most notable designs and their benefits and shortcomings. The term electrode in EEG, however, can have two meanings. One is the electrode as a whole, like the ones shown in Figure 2.14. The other is a block of material that transforms the ionic to an electric current; usually, this is a metal or alloy. In the following, the electrode is defined as the complete structure, whereas the material within the electrode responsible for the transformation will be referred to as the transformation electrode. Furthermore, this overview will only focus on non-invasive electrode technologies.

The primary property for the transformation electrode is that it has a low impedance in order to conduct the potential difference well to the amplifier. The physical principle behind the transformation electrode is that it acts as a capacitor that gets polarized by positively or negatively charged ions, causing a change in potential. Another important trait of a transformation electrode is that it should not or only slowly corrode with skin contact. Examples would be gold, stainless steel, or platinum. The downside of inert materials is that ions accumulate on the transformation electrode's surface, ultimately leading to potential drifts of the electrode [XMV+17]. Modern electronics can cope with these drifts. In older systems, however, this can lead to saturation and ultimately loss of signal. Therefore, in EEG systems, non-inert materials are used frequently. Silver-Silverchloride (Ag-AgCl) is a common used material. It buffers the accumulation of Cl^- ions at the electrode, avoiding drifts due to a chemical half cell reaction [XMV+17].

Adhesive electrodes like the ones on the right of Figure 2.14 have excellent signal-to-noise properties. They are one of the earliest types of electrodes and are still frequently used, and they also get used for conducting electrocardiograms (ECG). The transformation electrode is placed within an adhesive and conductive gel, which is glued onto the skin. Despite having outstanding EEG signal quality, they sport a considerable drawback, the surface skin area where they are placed has to be hair-free. This is usually no big issue for ECG recordings, as most of the body is not densely covered in hair. However, for EEG recordings, the electrodes have to be placed on the scalp. Therefore the patient would need to shave their head. Due to cosmetic reasons, this is often not desired and therefore only performed if the severity of the condition justifies the removal of hair.

Wet electrodes use a saline solution or a conductive gel to connect skin and transformation electrode. The gel itself can be a fluid or rigid gel. Usually, wet electrodes are fixated using an electrode cap like shown in Figure 2.13. Currently, **cup (wet) electrodes** are considered the gold standard for EEG measurements as the gel can penetrate through hair and therefore creates a good connection between the skin's surface and the transformation electrode. The used gel features a high electrolyte level like NaCl or KCl, enabling the fluid to conduct currents. The drawbacks of this form of an electrode are the rather lengthy preparation time as the electrodes often get filled while the head cap is worn. Furthermore, the gel-residues in the hair require a shower to get rid of [LGSMV14]. Instead of gel, saline solution can be used. However, due to the viscosity of the solution, it gets drawn onto the scalp, leading to potential short circuits. Another drawback of saline is that it evaporates quicker than gel, making long measurements impracticable. A solution to increase measurement times is to use more saline solution. This, in turn, however, also can lead to a decrease in signal quality [Ern05].

Dry electrodes try to overcome the disadvantages of wet electrodes. They are using neither a saline solution nor a conducting gel. As a result, they cannot dry out, are not leaving any residues, and, therefore, the general setup time is faster. The transformation electrode in a dry electrode is usually not located somewhere within the electrode, but the whole electrode consists either of the transformation electrode material or is coated with it. Usually, they have a finger-like structure to penetrate through hair. Those advantages are bought from not using wet elements, which, however, causes a significant downside. The contact between electrode and skin is worse than with gel or adhesive electrodes, leading to a low signal-to-noise ratio. That is why dry electrodes are not used in clinical settings despite their advantages [LGSMV14].

Other types of non-invasive EEG-electrodes which were not covered in this overview include electrodes with micro- and nanometer-scale-structures and electrodes with active components like amplifiers directly on or within the electrode [RDF+08, XMV+17].

2.3.3 Steady State Visually Evoked Potentials

Steady-State Visually Evoked Potentials (SSVEPs) are often the input-modality of choice for EEG-based BCIs since there is no need for users to train the evocation of SSVEPs. Furthermore, the signal is very robust, so that minor artifacts can be neglected. The general concept of SSVEPs is that a user is presented with a stimulation pattern, often a checkerboard. This pattern is then altered over time with a stimulation frequency f_{stim}.

There are several methods to evoke SSVEPs in the visual cortex. In Figure 2.15 the four main principles are displayed:

- Radial motion: Checkerboard rings move inwards or outwards, modulated with a sinusoidal frequency.

- Lateral/circular motion: The checkerboard is moving in a lateral motion. In the case of a circular checkerboard, it moves in a circular motion around the center of the structure.

- Pattern reversal: The color of each checkerboard field is inverted with f_{stim}.

- Pattern onset: The complete structure is presented at once and disappears again with f_{stim}.

A considerable benefit of motion to induce SSVEPs is that it is not perceived as flickering [HXX$^+$18], which can be perceived as exhausting. However, with flicker-based (pattern-onset and pattern reversal) SSVEPs, there are also methods that users do not perceive any flickering. One example for this is the work of Urpi et al. [AUS18], where they used high frequencies of flickering, which are near the flicker-fusion rate of the brain.

An example EEG recording with present SSVEP is provided in Figure 2.16. The subject under test was watching an 7.5 Hz stimulus for 10 seconds. The Figure is presenting the power spectrum (PSD) of the subject's EEG recording. The first response peak is visible at 7.5 Hz. At the same time, there is another peak at 15 Hz. The second peak is induced by the brain's non-linear properties, which excite responses at stimulated frequencies and harmonics. More about the analysis of EEG signals will be explained in Section 2.5.

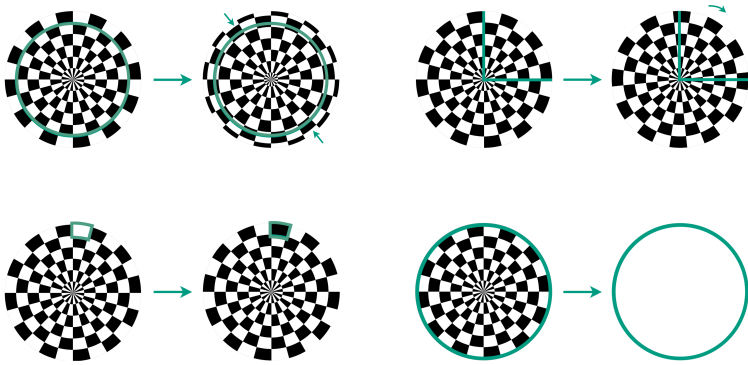Figure 2.15: SSVEP stimulation paradigms in the case of radial checkerboards. Top left: Radial Motion, Top right: Lateral Motion, Bottom Left: Pattern Reversal, Bottom Right: Pattern Onset.
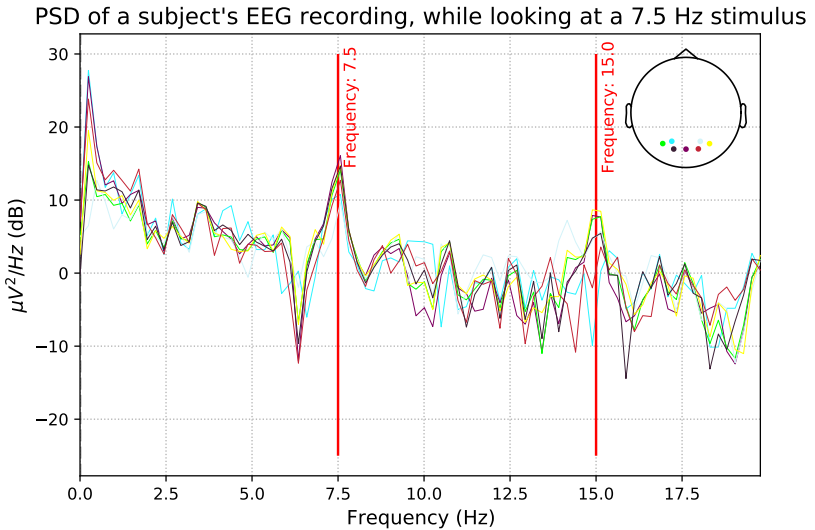

Figure 2.16: The figure shows the SSVEP response to a 7.5 Hz stimulus. The peaks at 7.5 Hz, as well as at 15 Hz are highlighted. The brain is a non-linear system that oscillates at the frequency of stimulation and their harmonics.
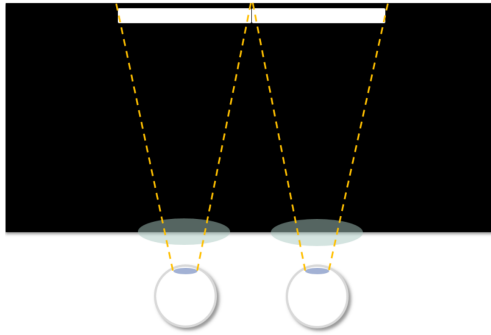
Figure 2.17: The general working principle of a VR HMD. A screen is viewed through lenses. The right eye only sees the right half of the screen and vice versa for the left eye. The display is really close to the eyes of the user. Normally, a user would not be able to accommodate such a close object. However, it is possible to view the VR display and realize a large field of view by introducing lenses. The current state-of-the-art VR HMDs cover a horizontal field of view of over 100 degrees.

2.4 Virtual Reality

It is often stated that the term *Virtual Reality* (VR) was first used by Damian Broderick in his novel *The Judas Mandala* in 1982 [Bro82, Kor17]. It can be used to describe a technology that enables the user to perceive a virtual world and interact with this environment realistically so that the user feels as if he is really in the simulated world. The feeling of being part of the virtual world is called immersion. Today, the state-of-the-art in VR technology are Head-Mounted Displays (HMD).

Virtual Reality gained traction in information technology in the last few years. Apart from the current interest in VR, the first consumer-grade VR HMDs hit the market in the 1990s, for example, the 'Forte VFX1' or Nintendo's 'Virtual Boy'. It can be assumed that the reasons for the disappearance of those systems from the market were technical deficits, high costs, and the high weight of the headsets. With the release of the Oculus Rift in 2013 and the technical progress in the last twenty years, the interest in virtual reality technologies has been reawakened. [Kor17]

The general working principle of such head-mounted displays is displayed in Figure 2.17. The user is viewing a screen through a set of lenses. The screen is physically separated in the center. Therefore, each eye is presented with an individual image, leading to a stereoscopic 3D impression within the brain. The lenses used are usually biconvex or fresnel lenses with positive dioptric power. With such lenses, virtual viewing distances farther than the actual distance of the display can be achieved. However, the image plane is in all commercially available HMDs non-variant. Therefore the eyes accommodate one viewing distance. This *vergence accommodation conflict* is often not a big issue, but it can cause the so-called VR-sickness (nausea and headaches) for some users [Kra16].

For vision-based BCI experiments, VR HMDs are particularly interesting due to the large field of view of about 90 to 110 degrees horizontally and vertically. Furthermore, no visual distractions are perceived while wearing the HMD as the user is entirely visually immersed. Therefore, VR delivers unique opportunities to research visual perception in humans. It can simulate visual experiences and stimulate the retina and brain at specific positions with complex patterns. Using a VR HMD, we can perceive a virtual world with a high level of immersion. Through this, we can experience visual perceptional states which we are not able to perceive naturally. An example is the simulation of visual diseases like glaucoma, age-related macular degeneration, and cataract. Potentially VR technology might even be used for the simulation of interocular lenses before transplantation. However, today, the resolution of VR-HMDs is still too low to realistically portray lowered contrast and artifacts from transplanted lenses. For example, for a VR display with a field of view (FOV) of $90°$ by $90°$, it would need

$$\frac{\text{FOV}}{1\text{arc min}} = \frac{90°}{\frac{1}{60°}} = 5400 \text{ pixels} \tag{2.4}$$

of horizontal and vertical resolution in order to meet a visus of 1 (healthy vision).

2.4.1 State-of-the-art Vision Loss Simulation

The simplest method to artificially simulate an altered state of perception is by using non-computer-based commercially available simulation goggles[23]. This technology enables healthy people to experience the influences of certain vision diseases first hand. However, those kinds of vision goggles are static for one stage of vision loss, and specific effects are hard to simulate by using optics and occlusion. Especially effects like increased light sensitivity are hard to implement.

Another particular interesting approach to vision simulation is the use of smartphone apps and screen-based applications. The Google Chrome browser plugin 'NoCoffee'[4] can simulate various visual impairments and apply them to a website. This tool is beneficial for developers when designing websites for visually handicapped people. 'GIP', glaucoma in perspective, is a smartphone application for the simulation of glaucoma [Cra16]. It displays the effect of glaucoma with the help of photographic representation of a domestic environment. Furthermore, the app informs about the disease and educates the user about the importance of treatment.

A novel method to simulate vision loss is through augmented, and virtual reality [AFF15, BZR05, LBCM11]. Thanks to the relatively large field of view, an immersive experience is provided. Digital technologies can simulate different causes of vision loss like macular degeneration, glaucoma, or diabetic retinopathy in one single tool. The work of Jin et al. [BZR05] is an early example for developing virtual reality vision simulations even before the emergence of consumer VR-HMD hardware. They used shutter glasses and table-mounted screens at a 45-degree angle to immerse the user into a stereoscopic experience. The drawback of such a VR system is its complexity and high cost with less immersion than the modern head-mounted displays. They used actual perimetry data as an inspiration for developing their software and, therefore, could display the progression of visual diseases within their software.

---

[2] https://www.lowvisionsimulators.com/ 'Our goal is to educate people about low vision' - last accessed in Jan 2021

[3] http://www.lowvisionsimulationkit.com/ 'Since 1979, the Zimmerman Low Vision Simulation Kit has been used by thousands of sighted individuals worldwide to gain a better understanding of the functional impact of visual impairment or low vision.' last accessed in Jan 2021.

[4] https://accessgarage.wordpress.com/ last accessed in Jan 2021

Lewis et al. [LBCM11] created a simulation of common vision diseases with the Unreal Engine 3. They used several post-processing effects provided by the engine, like motion blur and depth-of-field, to achieve their goal. Opticians and students evaluated the system. They conclude that despite inaccuracies within the simulation, the awareness and understanding of underlying effects in visual diseases were better understood by using the application. The drawback is that they did not implement any progressional vision loss, meaning simulation effects were static. There was no possibility to simulate the experience of a patient's vision loss over time.

Finally, we will have a brief look at the system by Ates et al. [AFF15]. This is particularly interesting as it uses a modern Oculus Rift HMD, which they further equipped with a camera to record, post-process, and represent the outside world. Users can select different levels of simulation intensity, changing the effects of the implemented filters.

2.5 Signal Processing

The signals obtained from the EEG brain-computer interface (compare Section 2.3) are received from multiple electrode channels. We obtain the differential potential between the measurement electrode and the reference electrode. Modern EEG devices receive many values per second and need to fulfill *Nyquist's criterium* for the measurement. In experiments regarding visually evoked potentials and steady-state visually evoked potentials, this maxes out at around 100 Hz, due to the decreasing response of the brain to higher frequencies. Therefore the sampling frequency should be at least 200 Hz. It is important to filter known sources of noise. One of those influencing frequencies is the frequency of the electrical power network. The European power network operates at 50 Hz, whereas the American at 60 Hz [Nei11]. Accordingly, a band-stop filter must be implemented. Modern EEG devices often integrate this filtering into their hardware.

One way to analyze data is to transform it into another representation space. Often signals are observed in the time domain. Using the Fourier analysis, we can transform them to the frequency-domain. From studying the transformed signal (compare Figure 2.16), we can develop an algorithm which is using thresholds to classify the existence of, for example, an 7.5 Hz stimulus-

frequency within the brain. However, the analysis of Fourier transformed signals has much uncertainty, especially when trying to define the threshold. That is why in neuro-informatics, the **Canonical Correlation Analysis** (CCA) is widely adapted to discover present steady evoked potentials in the brain (compare Section 2.5.2).

2.5.1 Fourier Analysis

Jean Baptiste Joseph Fourier introduced in 1822 the Fourier Series. The Fourier transformation translates a signal from time-domain into frequency-domain. The Fourier transformation of an integrable function $f : \mathbb{R} \to \mathbb{C}$ is defined by

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-2\pi j \omega t} dt, \qquad (2.5)$$

transforming the time domain function $f(t)$ to its representant in frequency-domain $F(\omega)$.

It is one of the most sophisticated tools of signal analysis. The computational pendant to the continuous Fourier transform used today is the discrete fast Fourier transformation introduced by Cooley and Tukey [CT65]. It is the algorithm for solving the following discrete Fourier series

$$F_d(\omega) = \sum_{k=0}^{N-1} A(k) \cdot W^{\omega k}, \quad \omega = 0, 1, ..., N-1 \qquad (2.6)$$

with the complex Fourier coefficients $A(k)$ and W as

$$W = e^{\frac{2\pi j}{N}}. \qquad (2.7)$$

Their algorithm made it possible to compute a Fourier spectrum effectively using machinery. The algorithm brought down the computational complexity of calculating discrete Fourier series from N^2 to $N \log_2 N$.

In signal processing for EEG signals we can assume a stochastical signal. Therefore, we do indeed not calculate the normal Fourier transformation but the Power Spectrum. The Power Spectrum (PSD) is defined as the Fourier transformed autocorrelation of the underlying function.

$$PSD = F(f_{xx}) \tag{2.8}$$

An example for such a calculated Fourier power spectrum for an EEG measurement is presented in the centered graphs of Figure 2.18.

2.5.2 Canonical Correlation Analysis

A more sophisticated approach towards data analysis of SSVEPs is called Canonical Correlation Analysis (CCA). It is a method that unfolds its full potential when a linear combination of frequencies needs to be detected in a multitude of channels. Especially in SSVEP experiments, we want to detect the stimulation frequencies Y_{ref} in the EEG channels X_{EEG}. Notice that X and Y are matrices containing the EEG channels and the stimulation frequency and its harmonics. We already learned in Section 2.3.3, that not only the stimulation frequency itself but also the harmonics are visible in the derived EEG signal. Hence we are searching for the stimulation frequency and the harmonics in X.

The original concept of CCA analysis as a mathematical construct dates back to Hotelling in 1936 [Hot92]. However, it took 70 years until it was first used by Lin et al. [LZWG06a] for the analysis of EEG signals for the detection of SSVEPs.

Now we will try to understand the general concept of the CCA in easy maths. After that, we will go a little more in-depth of its calculation.

First, we need to define the X_{EEG} EEG channels. We can write the channels as a vector of random variables, depending on the time t. Also, we can write it as a matrix with the dimensions of channels (N) and time samples (t), we can write:

$$X_{EEG} \in \mathbb{R}^{N \times t} \tag{2.9}$$

or formulate it to:

$$X_{EEG}(t) = \begin{pmatrix} x_1(t) \\ x_2(t) \\ \dots \\ x_N(t) \end{pmatrix} \tag{2.10}$$
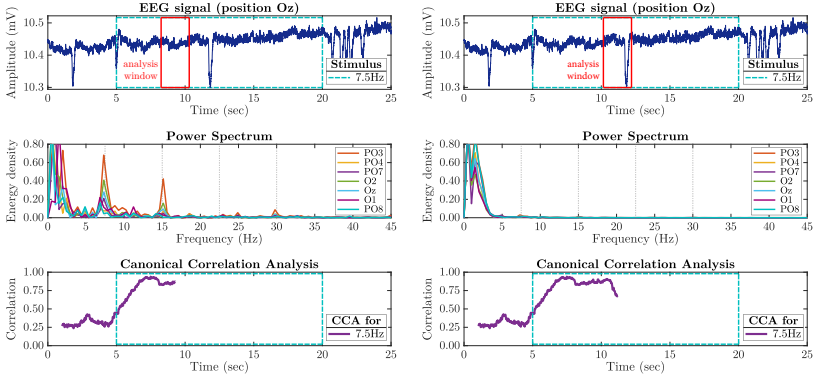
(a) SSVEP peaks are visible at 7.5 Hz and its harmonics in the Fourier power spectrum. For a good signal in the analysis window, the CCA value is respectively high.

(b) The power of SSVEP peaks is negligible compared to other signal components in the Fourier power spectrum. The CCA value is considerably decreased due to the blinking artifact.

Figure 2.18: The figures show the Fourier power spectrum and canonical correlation value of the EEG signal at the time of the red analysis window with a length of 2 s. During the blue window a 7.5 Hz stimulus was presented.

Source: *Previously published by the author in [5].*

with a total of N channels.

Now that we formulated the EEG signal, let us formulate the frequencies we are looking for in the EEG signal. For this we will use sine and cosine functions for the mapping:

$$Y_{ref} \in \mathbb{R}^{2N_{ref} \times t} \tag{2.11}$$

in this equation N_{ref} is the single frequency we are looking for. Notice that in reality, we can look for multiple disjunct frequencies. However, for the sake of simplicity, this example will only cover one frequency and its harmonics.

As we need both, sine and cosine representations of the function, we need the dimensionality of $2N_{ref} \times t$ for Y_{ref}. If we write the matrix as functions in a vector, this vector will look like this:

$$Y_{ref}(t) = \begin{pmatrix} sin\left(2\pi f_{ref}t\right) \\ cos\left(2\pi f_{ref}t\right) \\ sin\left(2 \cdot 2\pi f_{ref}t\right) \\ cos\left(2 \cdot 2\pi f_{ref}t\right) \\ \ldots \\ sin\left(N_{ref} \cdot 2\pi f_{ref}t\right) \\ cos\left(N_{ref} \cdot 2\pi f_{ref}t\right) \end{pmatrix}. \tag{2.12}$$

Basically, the last step to calculate the CCA is to correlate the two matrices. However, it is not a simple correlation, but we will introduce the weighting factors $w_{x,n}$ for each EEG channel and additional weighting factors for all reference signal channels $w_{y,n}$. Therefore we can write the equations X_{EEG} and Y_{ref} with the introduced w as X_w and Y_w:

$$X_w = w_{x,1} \cdot x_1(t) + w_{x,2} \cdot x_2(t) + \ldots + w_{x,N} \cdot x_N(t) \tag{2.13}$$

and

$$Y_w = w_{y,1} \cdot sin\left(2\pi f_{ref}t\right) + \ldots + w_{y,N} \cdot cos\left(N_{ref} \cdot 2\pi f_{ref}t\right). \tag{2.14}$$

The weighting factor is introduced so that there is the possibility to adjust the impact of individual channels and reference frequencies on the result. We are specifically interested in only the maximum obtainable correlation by adjusting the weights w:

$$CCA = \underset{w}{\arg\max}\left(corr(X_w, Y_w)\right) \tag{2.15}$$

and we can call the Euclidean norm of this CCA vector the CCA value.

In the following, we will perform a more in-depth dive into the CCA and its solution. Notice that it is more important to understand the mathematical concept of the CCA presented above than to understand its fine details. The

here presented method of solving the CCA by QR decomposition is heavily based on the publication of William H. Press [Pre11]. Most CCA implementations in today's statistical toolboxes like Matlab or Python are based on QR decomposition based on the algorithm by Golub [Gol69]. For the setup we will assume centered (mean substracted) data so that:

$$\sum_{i=1}^{n} x_{ij} = 0, \quad j = 1, \ldots, p_1 \tag{2.16}$$

and

$$\sum_{i=1}^{n} y_{ij} = 0, \quad j = 1, \ldots, p_2 \tag{2.17}$$

where x and y are the EEG signal and the reference signal and p_1 and p_2 are their dimensionality ($n \times p$). We are able to find multiple different linear combinations using the CCA algorithm. The total number of canonical coordinates is

$$d = min\,[rank(x), rank(y)] \tag{2.18}$$

with rank regarding column rank. However, for EEG experiments, only the linear combination yielding the largest correlation coefficient is usually of interest and is called the CCA value in the following chapters.

Following we will introduce $a \in p_1 \times d$ and $b \in p_2 \times d$ for the matrices u and v which denote the linear combination evaluated for each datapoint as:

$$u = xa, \quad v = yb \tag{2.19}$$

We can call the n rows of u and v *canonical coordinates*, they have d columns. By calculating the first column of a and b we get the single best linear correlation as:

$$\left(a_{(1)}, b_{(1)}\right) = \arg\max_{a_{(1)}, b_{(1)}} corr\,\left[xa_{(1)}, yb_{(1)}\right] = \arg\max_{a_{(1)}, b_{(1)}} \left[\frac{u_1^T \cdot v_1}{\sqrt{u_1^T u_1}\sqrt{v_1^T v_1}}\right] \tag{2.20}$$

with u and v as functions of a and b. We can see that all columns of u and v are cross-orthogonal, thus

$$u^T v = D \qquad (2.21)$$

with D as the d-dimensional diagonal matrix. By scaling the colums of a and b we get

$$u^T u = 1_d, \quad v_T v = 1_d \qquad (2.22)$$

with 1_d as the d-dimensional identity matrix.

For the solution using QR decomposition, we now need to decompose

$$x = q_1 r_1, \quad y = q_2 r_2 \qquad (2.23)$$

with $q_1^T q_1 = 1_{p1}$ and $q_2^T q_2 = 1_{p2}$ and r as the upper triangular. Now we calculate the SVD (singular value decomposition) as:

$$q_1^T q_2 = U S V^T \qquad (2.24)$$

with the column orthogonal U and V and the diagonal S. Using the ansatz: $a \equiv r_1^{-1} U$ and $b \equiv r_2^{-1} V$. By performing triangular backsubstituation we do not need to calculate the indicated inverse matrix. Now we can combine equation 2.24 with the equations 2.21 and 2.22 to the QR decomposition:

$$
\begin{aligned}
D &= u^T v = a^T x^T y b \\
&= (U^T r_1^{-1T})(r_1^T q_1^T)(q_2 r_2)(r_2^{-1} V) \\
&= U^T (U S V^T) V \\
&= S
\end{aligned}
\qquad (2.25)
$$

In equation 2.25 we used the associativity and orthogonality relations. Furthermore, it is noted that S and D are both diagonal. Similarly we can calculate

$$
\begin{aligned}
u^T u &= a^T x^T x a \\
&= (U^T r_1^{-1T})(r_1^T q_1^T)(q_1 r_1)(r_1^{-1} U) \\
&= 1
\end{aligned}
\tag{2.26}
$$

and similarly for $v^T v$. According to [Pre11] this is one of the fastest way to calculate the CCA. It is however noted, that whilst the SVD is generally well understood, the QR-decomposition which is also used in the ansatz presented is not trivial.

Finally, we can use the Euclidean norm of the CCA correlation vector, which is suitable to use as a feature for the classification of SSVEP signals [MMTC18].

A comparison between the euclidean norm of the CCA (in the following of this book called *CCA value* or *CCA correlation*) and the Fourier spectrum is illustrated in Figure 2.18 for an EEG signal with 7.5 Hz stimulus. From the results in Figure 2.18a, we can conclude that the CCA is a powerful and appropriate technique to analyze and classify the presence of known SSVEP features, which is, however, influenced by signal artifacts like blinking.

3 Concepts

> Our goals can only be reached
> through the vehicle of a plan, in
> which we must fervently believe, and
> upon which we must vigorously act.
> There is no other route to success.
>
> *Pablo Picasso*

This book will provide novel insights into human visual perception by using modern hardware and software tools like VR and AI. To fully comprehend vision, we need to know what it means 'to see'. In an initial step, different visual conditions are simulated in a VR platform. It is crucial to validate such developments in a clinical setting to understand whether the developed application is close to the actual perceptual experience.

A drawback of many applications in ophthalmology is that measurements often rely on user feedback and are therefore subjective. EEG-based BCIs can directly measure vision right at the occipital lobe and hence perform objective measurements.

In addition to classical signal processing approaches, machine learning artificial intelligence can be used to evaluate the visual cortex signals, which can be recorded using an EEG device.

3.1 Virtual Reality for Vision Simulation

Even with current technology, it is not possible to see through the eyes of somebody else. Unfortunately, there is currently no way to stream visual perception directly into the brain. However, current basic research in ophthalmology tries

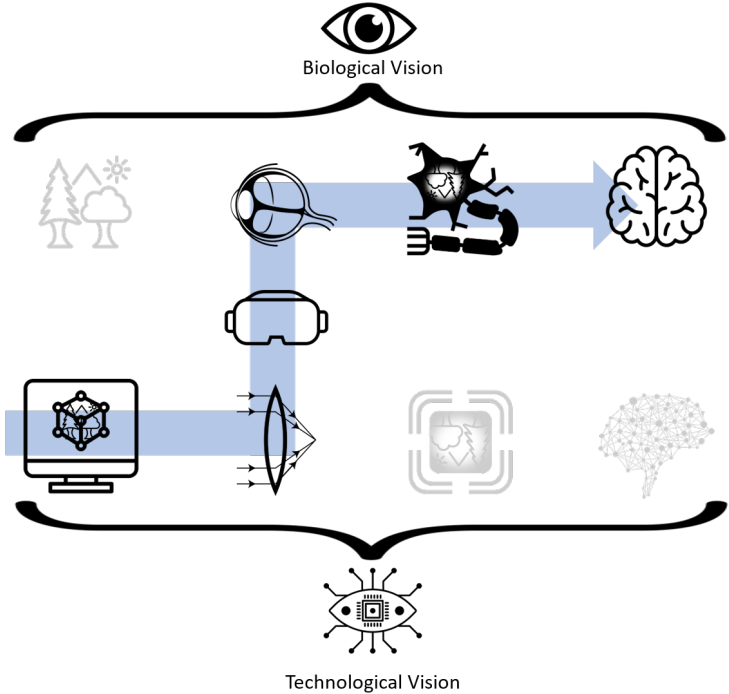Figure 3.1: When simulating an artificial environment and displaying it on the screen, the virtual world can be seen and perceived by our human eyes and brains. By changing properties in the simulated environment, like adding artificial glaucoma filters, a healthy person can experience the disease without having it. The VR device acts as an interface between the simulated world and the real world.

to implement systems enabling the perception of an altered state of vision. This can be achieved, for example, using virtual reality devices, which can immersively simulate vision experiences. For a better understanding lets extend Figure 1.1 to Figure 3.1. It portrays an artificial environment created by a computer. This artificial environment is rendered through artificial optics on an artificial sensory device (screen). The output is displayed on a consumer VR device. A user wearing that device will be able to immersively perceive the artificial environment using their eyes, retina, and brain. Figure 3.1 is a general concept for a vision simulation platform incorporating a user. Such a simulation should be non-invasive and reversible to the user. Therefore we are not actually altering the optics, sensory system, or perception of a user. However, we can alter what is happening in their field of vision using a VR device.

Several approaches have been implemented to simulate vision disorders. Those ranged from vision-altering goggles to VR and augmented reality (AR) applications [AFF15, BZR05, LBCM11] (compare Section 2.4.1). However, in most applications, visual impairments are statically implemented, making it hard to understand the development of visual disorders. Furthermore, the data displayed is generalized, and it is impossible to view individual cases.

Therefore, this book will focus on creating a VR application that can import visual fields from a medical perimeter and transform them into a perceptional experience. Section 2.2.2 describes how a perimetry scan of a human eye is performed. Aside from that, it introduced what a 24-2 scan is. Using such scans, we developed a post-processing filter simulating vision diseases.

3.2 Concepts of Machine Learning

This section will provide a brief introduction to machine learning and recent developments in artificial intelligence (AI). Goodfellow et al. provide an indepth analysis of AI in their book *deep learning* [GBC16].

Initially, we will define what machine learning is and what the differences to AI are. Because every machine learning algorithm is AI, however, not every approach in AI is machine learning. Generally speaking, every artificial system to which some form of intelligent behavior can be attributed is artificial

intelligence. Even complex mechanical structures, getting jobs done might be considered AI.

Regarding AI, people often think of the concept of *general AI*, which can solve various problems, just like humans do. This is somewhat unlikely to be implemented in the following decades. Therefore, this work will focus on more specialized AI. An example of a specialized AI would be a chess computer. However, chess is a relatively simple problem for advanced AI. Today, highly specialized AIs can beat even the best human players in the ancient game of Go [SSS+17], in which there is a vast space of possible moves (about $2.08 \cdot 10^{170}$ legal positions for a 19×19 board [TF07]). Another recent example is the mastery of Starcraft II, which is a complex digital strategy game [VBC+19]. The ultimate question is whether those computers already have something which we would call a 'mind'. Alan Turing already stated in 1950 experiments like the 'imitation game' [Tur50] for testing whether computers can think. Questions like these are rather philosophical and will not be answered in this book.

We will have a closer look at AI's advancements in recent years and how research and engineering are benefitting from it. But before we take a more in-depth look into this, we will face the big picture once again. Now that we know that AI is generally every intelligent behavior in machinery, the question arises of what machine learning is and how it is different from AI. As mentioned above, machine learning is a subset of AI. Complex if-else structures and algorithms crafted by humans to solve a problem are AI. The big difference in machine learning is that a machine learning model is not being programmed but is being trained. Therefore, the data scientist implements no *specialized* control structures and mathematical expressions to solve the specific problem. Only general mathematical expressions, the goal, and the dataset are provided to the machine learning model. The machine learning model learns from training *on its own*. There are several different methods for doing machine learning, ranging from Support Vector Machines to clustering approaches. More recently, Neural Networks gained much interest. The biological functioning of nerve cells inspired those. By layering multiple artificial nerve cells after one another, we achieve the 'deep learning' methodology. The breakthrough of that technology was novel ways to gradient-descent algorithms like the adaptive momentum (ADAM) [KB17] for the training of Deep Neural Networks. Mathematically, it is a complex task to get deep learning models
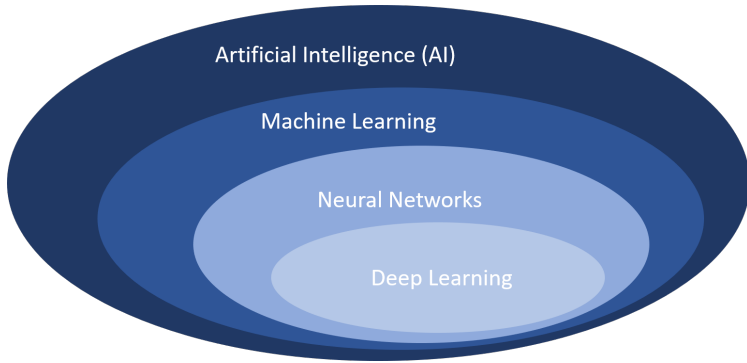
Figure 3.2: This venn-diagram shows how AI and deep learning are connected to each other. Machine Learning is a subcategory in AI, neural networks are a part of machine learning and deep learning is generally a subcategory of neural networks.

Source: *Own visualization, inspired by [GBC16].*

to converge, which would be impossible without modern optimization procedures. Figure 3.2 shows a venn-diagram for better visualization of the relations mentioned above of AI, machine learning, neural networks, and deep learning.

Machine learning can be divided into three different categories: Supervised, unsupervised, and reinforcement learning. In reinforcement learning, there is a *curious* agent, which is constantly exploring new inputs to improve its rewards which are defined by the developers' implemented reward function. Rewards might be points in a video game, whereas the inputs to the algorithm are the screen's output pixels, and the agent will provide outputs (inputs for the game) accordingly. Its ultimate goal is to reach rewards as efficiently as possible.

In unsupervised learning, generally, a large quantity of data is provided without labels. The machine learning algorithm tries to find structure in this provided Big Data. Different kinds of knowledge can be discovered in vast quantities of unlabeled data using unsupervised learning approaches like clustering. The aim is to group similar objects to one cluster and gain knowledge from the distributions and the clusters themselves.

In supervised learning, training data is provided in a labeled manner, which means that for the stated mathematical problem (also called loss-function), there is a clearly defined set of input parameters (features) and output parame-

ters (labels). Using methods like the ADAM-optimizer, as mentioned earlier, models are trained and converged to meet the training set's required outputs. Supervised learning is the method that is commonly used in data science for pattern recognition, especially in medicine. Therefore, this book will explore supervised learning in-depth. Section 3.2.2 will further introduce the principles of Pattern Recognition with neural networks and will introduce the general methodology of an artificial neuron and convolutions.

3.2.1 The Neuron

Before we look into more complex machine learning concepts, we first need to understand one of its most basic building blocks, the artificial neuron. The artificial neuron is inspired by real biological neurons like those in the human brain. However, the aim of the artificial neuron is *not* to exactly resemble the real world pendant. It is instead a simplification to a most basic mathematical principle of how a biological neuron functions.

The goal of a set of neurons is to map and approximate an underlying often unknown function with only the help of simpler functions, weights and biases. For this purpose, in the center of the neuron, there is the so-called *activation function* $y = f(x)$. Often non-linear functions are used for this like the Sigmoid

$$\text{sig}(x) = \frac{1}{1 + e^{-x}} \tag{3.1}$$

or the ReLU

$$\text{ReLU}(x) = \max(0, x) \tag{3.2}$$

or adaptions of the ReLU, like a leaky ReLU:

$$\text{leaky ReLU}(x) = \begin{cases} x & \text{if } x > 0 \\ 0,01 \cdot x & \text{if } x \leq 0 \end{cases} \tag{3.3}$$

The reason for using non-linear activation functions is so that complex non-linear functions can be approximated using those functions. The sigmoid was rather popular in artificial neural networks in the past. However, today the ReLU function is often a more favored activation function. However, the drawback of the simple ReLU is that for values below zero, the output is always
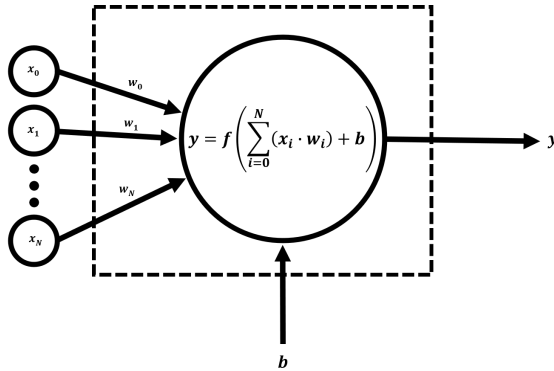
Figure 3.3: The figure shows the graphical representation of a artificial neuron. Information from previous nodes (x_n) are multiplied with wights (w_n) and a bias (b) is added. The result (y) is passed to the next layer of the neural network and in turn is handled as an (x_n) input for the next neuron.

zero. Therefore, information might be lost. The leaky ReLU tries to overcome this issue this by introducing a *leaky* part to the function so that values below zero are considered, however not as strong as the values above zero. The input x into the activation function is calculated as the sum over all previous outputs multiplied by a weight. Additionally, a bias is added. The result of the activation function is then passed to the next neuron layer, this procedure is called *feed forward*. A graphical representation of the neuron is provided in Figure 3.3.

If multiple neurons are connected together through multiple layers, it is called an *Artificial Neural Network*. The most straightforward implementation of such Artificial Neural Network is the Multi-Layer Perceptron, a Neural Network composed of an input layer, a hidden layer, and an output layer. Cybernko [Cyb89] proved in the *Universal Approximation Theorem* that with just this one hidden layer and a finite number of neurons, any continuous function could be uniformly approximated to arbitrary precision.

Modern neural architectures are commonly built from multiple hidden layers. If a neural network has more than one hidden layer, we call that kind of machine learning model a *deep learning* model. All neural network architectures are trained using the method of *backpropagation*. The goal of it is to minimize

the underlying cost function C with respect to the weights (w) and biases (b) of the neural network. Note, that every neuron n, each connection n, m in each layer l has their own variables. Therefore the gradients used for the gradient descend are:

$$\frac{\partial C_x}{\partial w_{n,m}^l} \tag{3.4}$$

and

$$\frac{\partial C_x}{\partial b_n^l}. \tag{3.5}$$

The weights and biases are updated in iterative steps in order to improve the cost function. The step size μ is defined in the set of hyperparameters of the model. However, optimizers like ADAM change the step size over time to increase the possibility of converging to a solution for the artificial network. To better understand the underlying optimization process in deep learning, the following Subsection will introduce the concept of pattern recognition and classification utilizing Deep Neural Networks.

3.2.2 Pattern recognition with Deep Neural Networks

As we learned in the previous section, deep neural networks are composed of multiple layers of neurons. An exemplary Deep Neural Network is presented in Figure 3.4. This neural network has eight input parameters. The input itself is then adapted to the activation function of the neural network. For example, when using sigmoids, the input should be scaled to match the range of the function, in that case, $[-1, 1]$. However, when using ReLU activation functions, any input above zero can be used without further tuning. Although scaling and standardization might improve the convergence rate. The presented example in Figure 3.4 has two output neurons and is used for classification. Any combination of inputs $\vec{x_{in}}$ can be mapped to an output $\vec{y_{out}}(\vec{x_{in}})$. Inputs in Machine Learning terminology are called *features*, whereas known outputs to features are called *labels*. The cost function C is also referred to as the *loss* function. A common function is the *Mean Squared Error (MSE)*:

$$C = \frac{1}{n} \sum_x \left\| \vec{y}(\vec{x}) - f^l(\vec{x}) \right\|^2 \tag{3.6}$$

Input Layer ∈ ℝ⁸   Hidden Layer ∈ ℝ¹²   Hidden Layer ∈ ℝ¹²   Output Layer ∈ ℝ²
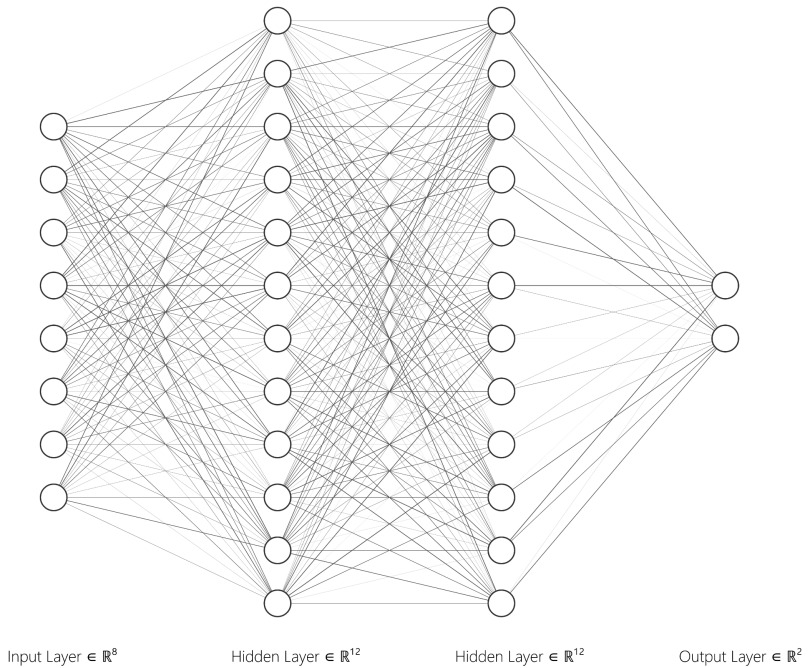
Figure 3.4: The example shows a fully connected Deep Neural Network. It has eight input parameters, two hidden layers with 12 neurons each, and two output neurons. A neural network like this could be used for binary decision-making processes. The intensity of the graphs' edges represents different weights. Biases are not displayed.

Source: *Created with http://alexlenail.me/NN-SVG/*

where n is the number of output neurons which is equivalent with number of classes and $f^l(\vec{x})$ is the activation of the output neuron given the input $\vec{x_{in}}$ to the Neural Network. By utilizing backpropagation, we try to converge to a well perfoming solution for the deep learning model.

In order for this to happen, many iterations need to be calculated. In reality, for each iteration, not only one input vector from the dataset is used, but *batches* and resulting gradients are averaged. Therefore the underlying cost function of the whole dataset is approximated in every iteration. Sebastian Ruder [Rud16] published an in-depth overview of gradient descent optimization algorithms in 2016. Methods like these enable Neural Networks to handle even complex cost functions.

One major drawback of fully connected Neural Networks is the massive amount of weights and biases that need to be tuned. Let us think of an image with three color channels and 128 times 128 pixels that we want to classify. Therefore, the image has the dimensions of $H \times W \times d = 128 \times 128 \times 3 = 49,152$ input nodes. Let us assume in an initial step we want to cut the dimension in half; the first hidden layer needs half that many nodes ($\frac{49,152}{2} = 24,576$). Because of the fully connected structure of the approach we end with $49,152 \cdot 24,576 = 1,207,959,552$ parameters for weights plus an additional 24576 biases. Further layers only add to this problem. The more parameters there are to train, the longer the training and the more likely the system is to overfit or not converge if not enough representative training data is fed.

Another disadvantage of fully connected Neural Networks regarding pattern detection is that they learn locally and not globally. Let us assume we built that network for the input of $128 \times 128 \times 3$ and trained it on an outstanding amount of cat photos. Additionally, let us assume the cats were always centered in the image. We will end with a cat-detector, which will only work for cats that are centered. If the cat is scaled or moved to one of the image's corners, the detector will most likely fail because the weights and biases of cat features in that scale, rotation, or position were not present in the training set.

As we can easily see from those two examples, fully connected neural networks are not ideal when trying to classify images. When we reduce images to one dimension, we get a one dimensional series. This is very similar to time series, where usually adjacent points correlate with each other (e.g. in a electrocardio-
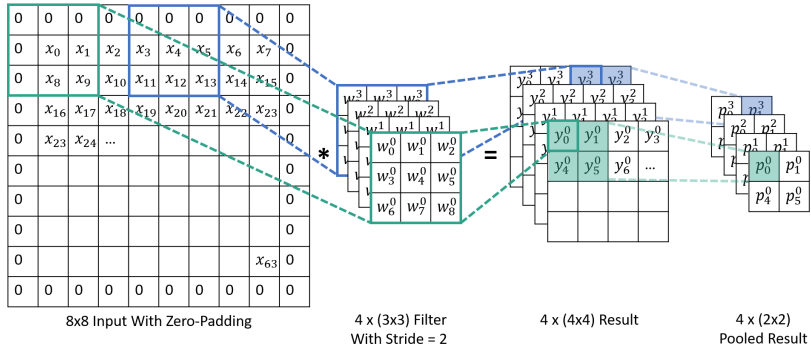
Figure 3.5: The figure shows the graphical representation of the inner workings of a CNN. An input image is padded with zeros and convolved with four different (3x3) filters. This is resulting in four (4x4) feature maps. To further decrease the dimensionality, pooling is applied. The combination of convolution and pooling layers is very common in CNNs.

gram or in electroencephalography). To overcome these issues, the following section will introduce the concept of convolutional neural networks (CNNs).

3.2.3 Convolutional Neural Networks

Like the name already implies, convolutional neural networks (CNNs) are heavily based on convolutions. This drastically reduces the number of needed parameters and provides spatial independence. In fully connected architectures, we learned about activation functions and single neurons; however, CNN's basic building blocks are convolution-filters, called *kernels*, and *pooling* operations. In the following, the basic principles of CNNs will be discussed. As a reference, please refer to Figure 3.5 in order to better understand the following explanations.

The input image is usually provided as a two- or three-dimensional matrix. In Section 2.1.3 we learned about the three kind of cones in the retina and their importance to color perception. Therefore, in information technology, images are often encoded using one to three color channels (d), leading to a $x \times y \times d$

57

dimensional input. However, the channels are often reduced to one ($d = 1$), especially when no additional information is expected from color.

Before the first convolutional filter processes the image, data preprocessing is often performed to normalize and standardize the dataset's data. Another important preprocessing step might be to scale the image to have the correct dimensions for the Neural Network to function. However, for now, let us assume that all preprocessing is handled. So the first convolution layer and its kernel masks are executed. In general, kernels, in analogy to images, are defined as

$$K_x \times K_y \times K_d \tag{3.7}$$

with K_x and K_y being the size of the filter and K_d being the number of two-dimensional filters in the kernel. In Figure 3.5 a $3 \times 3 \times 3$ filter is presented. Similar to a moving filter in any other signal processing method, the kernel is processed over the whole image step by step. The parameter *step size* is an important parameter in this case, as we can decide to jump and neglect some data points. In machine learning terminology, the step size is called *stride*. Finally the filter is executed over the whole image and moved to the right in accordance to the stride. After reaching the right end of the image, the filter proceeds to the next lines, resulting in a new matrix of values. In the case of multidimensional filters, the result will be multidimensional, too. An exemplary calculation for y_0^0 is:

$$y_0^0 = x_0 \cdot w_4^0 + x_1 \cdot w_5^0 + x_2 \cdot w_7^0 + x_3 \cdot w_8^0. \tag{3.8}$$

The result of the convolution is a new result matrix which is called *Activation Map* or *Feature Map*. We notice that in Figure 3.5 many zeros are added around the input image's pixels. This technique is called padding, or in this case, more specifically, zero-padding. Because we used zero-padding in the example provided, Equation 3.8 contains only four multiplications and not nine as the other multiplications are with zero. Padding is an often-used method to increase the size of the input matrix and, therefore, increase the size of the feature map. Furthermore, padding allows the features at the image's border to gain just as much attention (passing of filters) as center features.

We can calculate the resulting size of the resulting feature map $(Y_x \times Y_y)$ quite easily if we know the input size $(X_x \times X_y)$, the amount of added pixels from padding (P) the used stride (S) and the kernel size $K_x \times K_y$:

$$Y_x = \frac{((X_x + P) - K_x)}{S} + 1 \qquad\qquad (3.9)$$

$$Y_y = \frac{((X_y + P) - K_y)}{S} + 1. \qquad\qquad (3.10)$$

The dimensions d of the resulting $Y_x \times Y_y \times d$ feature map is defined by the number of filters used, which is four in our example. The reduction of dimensions from the convolution operation is often not enough, and we want to reduce the dimensions more drastically. We already learned that we could use a higher stride value to achieve this; however, some valuable information is lost in this process. The larger the stride, the more information is lost without regard to the importance of said information. Therefore, the standard CNN structure implements pooling operations. Pooling operations are a reduction in dimensions adhering to certain rules. Often the *max-pooling* or *average-pooling* methods are used. They are pretty self-explanatory, with max-pooling directly transferring the highest value to the next feature map and average pooling the average of all values. In our example, a pooling layer with a pooling size of 2×2 and a stride of 2 are used to obtain the pooled result.

This presented combination of convolution-filter and sub-sequential pooling is the foundation of most modern CNN architectures.

Please note that we assumed a mono-color image as the input, but the same principle will work for images with multiple dimensions. Furthermore, despite describing the matrices as 'images' in this Section, the matrix can also be a one-dimensional time-series or any form of a multi-dimensional matrix with local correlations. CNNs excel at problems with local dependencies and correlations.

As the above explanation is rather theoretical in nature, we will have a exemplary look at a very prominent CNN which has been around for some time. The network is called LeNet-5, was published in 1989 by Le Cun et al. [LBD+90] and is presented in Figure 3.6.
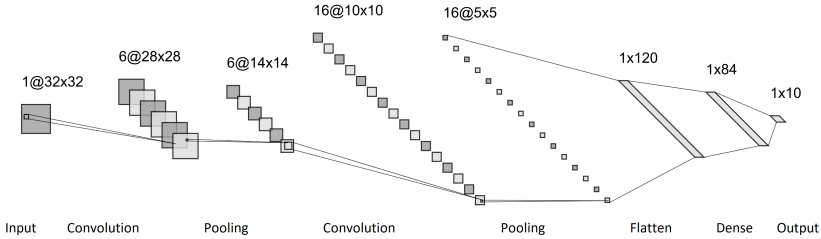
Figure 3.6: LeNet-5 is a prominent example for CNNs. It was first applied in the late 20th century for the detection of postal codes on letters.

Source: *Created with http://alexlenail.me/NN-SVG/ based on [LBD+90].*

It is a CNN for the recognition and classification of written numbers. Its most prominent application was the recognition of ZIP codes on mail. As of today, a lot of beginning Data Scientists do their first steps alongside number-recognition datasets like the MNIST database (Modified National Institute of Standards and Technology database) [Li 12]. MNIST is a dataset composed of hand-written digits between 0 and 9 with 60,000 training and 10,000 test samples. When looking at the structure of LeNet-5, we can see the previously introduced structure of convolution followed by pooling layers. However, later in the network a *flatten* layer is introduced. This concatenates all remaining values of the previous activation maps in one vector. This vector is then used as input for a fully connected Neural Network to produce an output vector. The output vector is a categorical vector with one category for every recognizable digit, therefore ten classes. Using the MNIST training data, the weights and biases of the LeNet-5 network can be trained to reach a human level of digit recognition.

In the next section, a systematic process for the creation of CNN models is provided. After that section, a computer vision concept study is provided to better understand the functionality of CNNs, which will be used to classify brain data later in this book.

3.3 Machine Learning Development Process (QUA^{3}CK)

When trying to build a machine learning-based solution, a huge obstacle is that there are few standardized processes for developing an artificial intelligence model in an academic setting. There are standardized methods like KDD, which stands for Knowledge Discovery in Databases, and the CRISP-DM (Cross-Industry Standard Process for Data Mining) process. However, those processes are catering towards data mining and development in a business setting [AS08, She00]. In industry and science, a defined, systematic approach towards a complex problem is helpful. However, there is a difference between industrial development and scientific research. First and foremost, a scientific approach needs to cater to more uncertainty.

The **QUA^{3}CK** machine learning process, which will be introduced in the following, was initially developed for the *Laboratory for Applied Machine Learning Approaches (LAMA)*. LAMA was the first hands-on laboratory course teaching modern machine learning methodologies in electrical engineering at the Karlsruhe Institute of Technology. It has also been awarded the prestigious Faculty Teaching Award 2020 [6].

The QUA^{3}CK process was publically introduced during the LAMA in Summer 2019. Following, in October of 2019, the laboratory and the development process were introduced to the scientific community at the Symposium Artificial Intelligence for Science, Industry and Society [7]. The resulting scientific paper was published at Proceedings of Science (PoS) later [8].

The basic structure of the development process is presented in figure 3.7. In the following, the individual steps of the QUA^{3}CK machine learning development process are described. The following description of the steps of QUA^{3}CK is partially recited, adapted, and shortened from [8]:

Question:

The first step of the process is to specify a problem, which needs to be solved. The goals, requirements, and expected outcomes have to be stated. This step helps to avoid the common mistake of realizing too late that the conditions and specifications are not entirely defined. Once the requirements are precise, it is possible to continue analyzing the data.

Understanding the data:

If a dataset is not available from the beginning, the first step is to formulate it. An established database can be based on open-source data, which already fulfills a part of the predefined requirements. On the other hand, it can be collected from sensor measurements or by information retrieval. Once a dataset is available, the raw values must be analyzed to discover and understand possible hidden information, such as irrelevant data or unexpected trends. Data augmentation might be executed as well to increase the diversity of the samples. In general, consistent formatting of the input dataset and labels are crucial for the following steps.

Algorithm:

These three steps can be seen as a loop that can be executed as many times as necessary until the model fulfills the expected outcomes.

In the first step, an algorithm is selected based on the application. There are specific rules of thumb for this. For example, for images and videos, Convolutional Neural Networks are usually set in place; for time series forecast, Recurrent Neural Networks, for classification of numerical features, Feed-forward Neural Networks. However, some applications do not match these rules and it is always important to think outside the box.

Once an algorithm has been selected, the dataset should be adapted correspondingly (Data Adaptation step). Each algorithm expects a particular input format. For example, the given dataset represents a time series with a single feature, and the task is to predict the following ten steps. It is crucial to rearrange the data so that input (the single feature) and label (the following ten steps) are available to the machine learning model.

Hyperparameters, such as the number of training epochs or neural network architecture, are present in most deep-learning algorithms. They control different aspects of the model and hence have a significant impact on its performance. However, they are not adapted by the learning process and have to be tuned either manually or automatically by another program (Hyperparameter Adjustment step).

As soon as a set of hyperparameters has been chosen, the model can be trained on the training dataset. For validation, a separate, equally distributed dataset

is extracted, called the validation set. The validation error is calculated for various hyperparameter configurations. The best performing selection is used to move to the next step, conclude, and compare.

Sometimes it might be necessary to return to the *Understanding the data* step in case none of the algorithms selected is performing as expected. Changing the set of features or increasing the size of the database might be a solution. Problems, such as missing important information or interdependent variables, increase the complexity of the model but not its performance.

Conclusion:

In this step, it is checked whether it is necessary to execute another round from the A^3 loop. Therefore, the performance of the trained model is tested on another, separated, equally distributed dataset, called the test set. The results are compared to reference models, which might be taken from literature or other models created on the same QUA^{3}CK process.

If necessary, it is possible to go back to the Algorithm Selection step. Otherwise, if the model already satisfies the initial problem's requirements, the next step is Knowledge Transfer.

Knowledge transfer:

Once the results meet the requirements, the model can be considered as finished. In this step, it is of importance to prepare detailed documentation. The algorithm, starting values, the final hyperparameter set, the training, validation, and test sets must be described in detail. Additionally, they should also be secured and stored as well as all other relevant information and data. This will ensure the replicability of the study in the future. The last step can involve the deployment of the algorithm on a target platform. This can be on embedded hardware, as an online solution in a data center, on a mobile device, or another system.
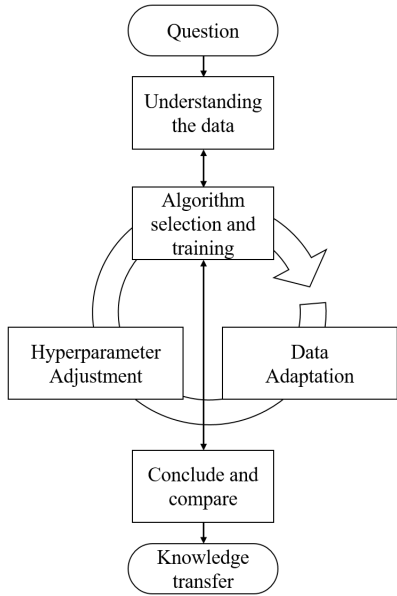
Figure 3.7: The QUA³CK machine learning development process.
Source: *Previously published by the Author in [8]*

3.4 A Visual Perception Study with Deep Learning: DeepBees

Hopefully, AI algorithms' importance and power, especially in pattern recognition and computer/technological vision, is evident after the last sections. However, the explanations which introduced convolutional neural networks were on a somewhat superficial level so far. In order to better understand to what extent we can utilize CNNs today, we will take a look at a project which involves the technological side of vision.

"If the bee disappeared off the face of the Earth, man would only
have four years left to live"
Albert Einstein

This dissertation is not about solving the pollinator situation. However, there is much vision research can do to understand the general situation better. In the following, we will explore how we can harvest the power of technological vision to understand Apis Mellifera better, the western honey bee.

The pollinator population is on a steady decline for decades [KP07]. We still do not understand all underlying mechanisms and influences on the decline of honey bees. However, we feel confident that pesticides and biodiversity play a considerable role in the equation [BVB+10]. To better understand this role and all of its details, we need to first understand the behavior of the bees on a day-to-day basis. Starting at simple steps like: How many bees are leaving the hive and how many are entering? Are they carrying pollen? Are there natural predators attacking the hive? Are the bees ill? For this purpose, the hive entry seems to be the ideal position for positioning a camera system. In this case, we worked together with apic.ai, a Karlsruhe start-up specializing in monitoring bees and using them as 'flying sensors'. In collaboration, we harness the power of technological vision systems to further automate their systems utilizing the strengths of CNNs and AI.

The following example will serve as an introduction to AI by utilizing the QUA^{3}CK machine learning process. However, before we start with the AI, first we need to define the overall problem in an overview manner. We start in the real world, where bees and hives exist. That is our starting point in Figure 3.8 on the left on the biological side. Then we convert this real-world perception using a camera system made from lenses (technologically shaping light) and a CCD (image sensor) to a virtual representation of made from pixels. Furthermore, on the last and final step, this pixel-image-representation is processed using artificial intelligence. This artificial intelligence will be called 'DeepBees' and we will use this example to introduce computer vision in a real-world use case. However, the explanations provided here will be on a superficial level, while all the details can be read in the published papers [12] and [13].

As already mentioned, we will use the QUA^{3}CK ML process. In which we will start with the

Q - Question

We want to understand honey bees and their behavior better. This includes them entering and leaving the hive, collecting pollen, and interacting with other
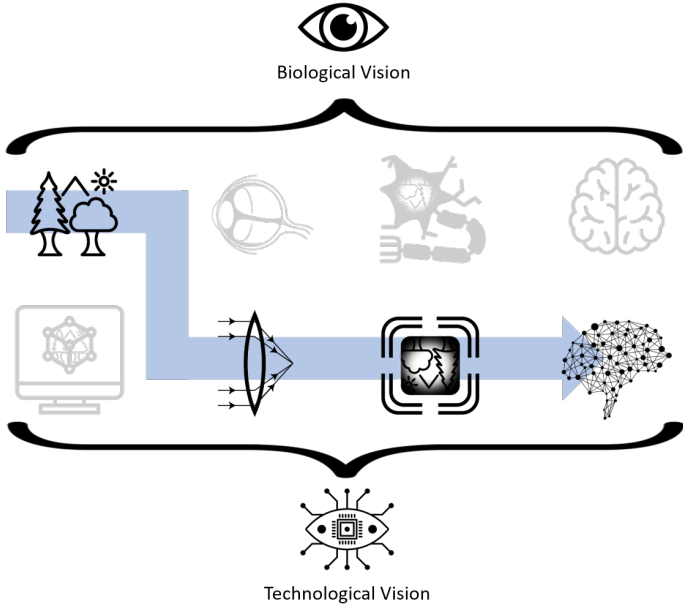
Figure 3.8: Another example for visual perception is presented in this model. We have a real-world environment that is recorded through technological optics and imaging sensors, therefore a camera. The resulting image is then post-processed with AI computer vision algorithms. In the example of DeepBees, bees (real-world) are recorded (using a camera) when entering and exiting the hive. Computer vision algorithms are used to extract further information, like whether they carry pollen or not.

insects like wasps and hornets. We will be using a system provided by apic.ai for the monitoring of beehives. To automatically detect bees, their pollen, and their health status, we need first to detect the bees within the recorded image. Next, we want to extract those images of bees and then further classify the extracted image. Specifically, we want to classify the genus of the insect (whether it is a bee, wasp, bumblebee, or hornet), whether it is carrying pollen, and where the pollen are located within the image. We want to further classify if the bee is pollen-bearing, a drone, or even dead. Additionally, we will perform a pose-estimation, with that possible malformed bees could be detected and certain behavioral features could be classified in the future.

Figure 3.9: The used monitoring systems of the DeepBees study. When bees enter the hive through the entrance below the yellow boxes, they get recorded with the attached hardware. The hardware includes light sources and a camera.

Source: *Previously published by the Author in [13] in collaboration with apic.ai*

U - Understanding the data

As before mentioned, we will use data acquired by the Karlsruhe start-up apic.ai. For this, they are using their in-house know-how and technology to record bees directly at the hive entrance. An exemplary setup is provided in Figure 3.9. The yellow boxes are the entry-areas for the bees, and within the black boxes, all recording hardware is hosted. Those entries are the only entrance to the beehive. Using this setup, we were able to collect a large quantity of bee data. Furthermore, publicly available datasets were added to increase the size of available training data further. Labeling is of utmost importance. The data was collected in the form of video streams, train and test must be split to prevent unwanted overfitting. Using streams of multiple images instead of single images is a form of augmentation in itself as it greatly improves available training data. Moreover, to further regularize the training, many common augmentation techniques like rotation, shifting, shearing, and flipping need to be implemented. Furthermore, color transformations were performed to tackle the problem of potential tint bias. Currently, the dataset is still under construction and not publically available. However, apic.ai plans on releasing it.

A - The Algorithm

The selection of the deep learning algorithm is one of the most crucial steps of the QUA^{3}CK development process. Furthermore, it is an iterative process including data adaptation towards the selected algorithm and its optimization, also known as hyperparameter adjustment. The first decision was to evaluate CNN-based architectures, these usually excel in image recognition tasks. Therefore VGG, Inception, ResNets, MobileNets, and PNASNets were evaluated at different sizes, with MobileNet-v2 outperforming the others in terms of a trade-off between accuracy and required resources.

Hence we will use MobileNet-v2 as a feature encoder. A feature encoder extracts all relevant information from an image and reduces them to a latent space representation. Using that representation, we can classify specific properties of the recorded bees. The structure of the overall architecture is provided in Figure 3.10. As we can notice, there are multiple modules attached to the encoder. With these, a machine learning regularization technique called multi-task learning can be utilized. Using this principle, multiple related tasks help to improve the results of similar tasks further. As we can see, there are special modules, for example, for the pollen, genus, and classification. The pose module is a reverse CNN which outputs are confidence maps for key points on the bee. We also notice the decoder, which is used for so-called structural learning. The decoder tries to rebuild the input image from information provided in the latent space representation alone. By observing the decoder, we can observe what kind of information is present within the latent space. Furthermore, the hyperparameters were optimized by using a random search optimizer. A total number of 40 sets of hyperparameters were evaluated. The most interesting finding was that the ADAM learning-optimizer significantly outperformed Momentum, which relies on joint learning, probably due to the network setup. Therefore, keeping individual learning rates is crucial.

C - Conclusion

The final output of the presented convolutional multi-net is presented in 3.11. We can observe the fitted skeletons on the bees. Also, we can observe whether the bees are carrying pollen, including bounding boxes for pollen. Furthermore, the network is classifying whether the bees are drones or even dead. The network provides overall good performance. However, some errors can be observed in the fitted skeleton on the lowermost bee. The results for comparison
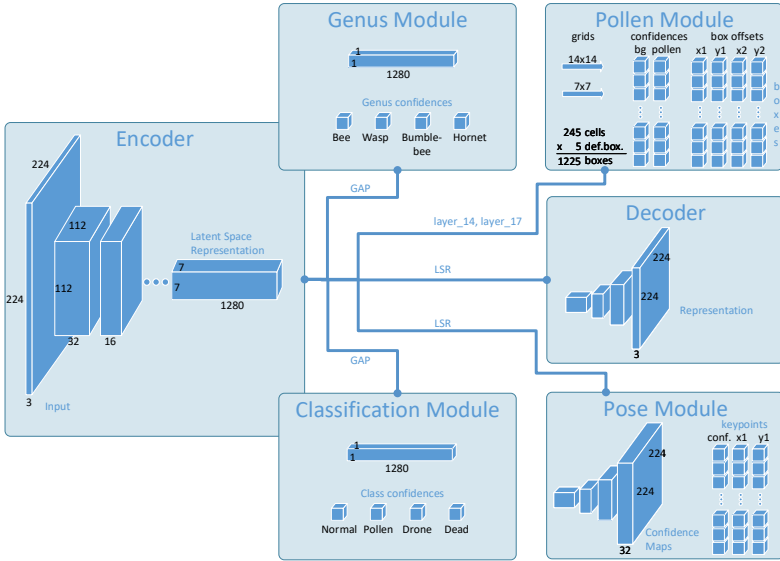
Figure 3.10: DeepBees, a MultiNet consisting of a shared feature extractor and modules for genus identification, pollen detection, unsupervised learning, pose estimation and classification.

Source: *Previously published by the Author in [12]*

of a single-trained version of the network (individual modules) and multi-task learning (DeepBees) are presented in Table 3.1. We can observe that the multi-task learning overall outperformed the single-task learned network model due to better regularization. Only the average precision for the pollen detection (AP_{pol}) was better using single-task learning. However the area under curve for classification (AUC_{cla}), genus (AUC_{gen}) and the relative distances ($1-AD_{Pose}$) as well as the overall accuracy result ($\mathcal{P}_{Avg}$) were significantly better utilizing multi-task learning.

K - Knowledge Transfer

In terms of the K-step of the QUA^{3}K process, we published the findings and a dataset paper about bumblebees at international conferences [12, 13]. What can we learn from these experiments involving computer vision in a real-world
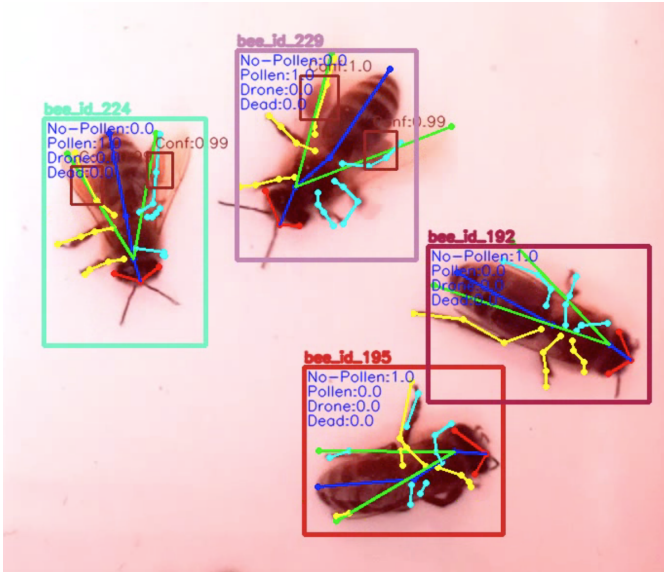
Figure 3.11: This figure shows an exemplary output from the DeepBees CNN multi-task architecture. The output includes pose estimation, pollen detection and genus classification.
Source: *Previously published by the Author in [12]*

Table 3.1: Task performance comparison for individual training (Single-Task) and DeepBees.

Model	$\mathbf{AUC_{cla}}$	$\mathbf{AUC_{gen}}$	$\mathbf{AP_{pol}}$	$\mathbf{1-AD_{Pose}}$	$\mathbf{\mathcal{P}_{Avg}}$
Single-Task	70.34%	76.05%	**46.64%**	90.12%	70.78%
DeepBees	**82.41%**	**76.19%**	40.14%	**93.61%**	**73.08%**

Source: *Previously published by the Author in [12]*

setting with a real problem? First, we can stand in awe of how far machine learning brought computer vision in just the last decade. Only ten years ago, work like this would have been almost impossible with the same resources. Additionally, this chapter and its research provided a good foundation on CNNs and their inner workings. Furthermore, we learned about multi-task learning and how it can be used as a potent methodology for regularization.

This work was mainly presented here as an example of how machine learning is leveraged in a scientific, methodical approach. Furthermore, the learnings from this computer vision project significantly influenced machine learning within the Metis Videre platform described in the following sections.

3.5 The Metis Videre Brain-Computer Interface Platform

In the previous sections, we first conceptualized vision in a simulation environment for different visual diseases and other influences that can alter human vision. Afterward, we introduced machine learning to visually perceive using CNNs and large quantities of data. Now, we will combine the human visual experience and the perceptual capabilities of computers to establish a platform for objectively measuring human vision parameters. The platform is called *Metis Videre*. Metis is inspired by the old greek goddess of practical, complex, tacit knowledge and Videre from Latin *seeing*; therefore, the platform's name is an analogism that we want to generate knowledge and insights into vision in humans.

The general concept of visual experiments involving EEG brain-computer interfaces is quite simple. We only need a displaying device like a monitor, smartphone, or VR-glasses to present a stimulus. A person is looking at those stimuli, and those in return excite potentials in the brain, primarily in the visual cortex at the back of the head. If the stimulus is periodic, we obtain an SSVEP in the brain (compare Section 2.3.3). This SSVEP is not visible to the outside world. However, we can measure it using electroencephalography. The EEG is connected to a subject, with electrodes over the visual cortex at the occiput. Now, we can measure brain data in real-time using wireless transmission and a connected computer. It is the same computer that generates the stimuli observed by the user; therefore, we achieve a closed-loop capable system with a human in the loop. By recording, labeling, and evaluating a lot of data, we can train machine learning models to predict the functionality of the person's visual system. Those algorithms are trained offline on many computers in parallel, and the final trained model is shared back to the BCI-computer. This principle of the human in the loop as well as the controlling computer and the learning process is displayed in Figure 3.12. However, this schematic view of the system is simplified as many components need to interface.

We will reflect on the biological and technological vision chart once again. Though, this time we will need to add the components of a presentation device and an EEG. Therefore, we need one connection from the technological vision world to the real world and a way back to handle the vision information technologically. The overview is presented in Figure 3.13. We start with a
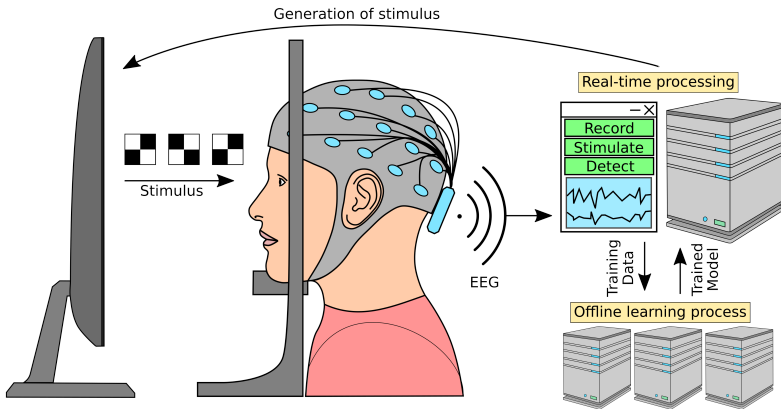
Figure 3.12: The figure shows the basic working principle of the Metis Videre platform. A user is presented with stimuli which in turn evoke visual potentials. Those are recorded by an EEG and transmitted to a computer. The same computer is responsible for the classification of recorded signals and the production of subsequent stimuli. The algorithm information and models are trained on other computers, and only the trained model is copied back to the main BCI computer.

Source: *Previously published by the Author in [5]*

computer-simulated environment. In this case, it is used to produce stimulation patterns like checkerboards, flickering at a certain frequency. The stimulation patterns are produced from a shader in a 3D environment (Unity) and then rendered to a screen; let us assume a VR-HMD. This technological interface device connects with the help of electronics and optics the technological vision world with the biological world. Real humans can now watch the stimuli with their very own eyes in the real world. Just like the biological function of vision was described in the Fundamentals Chapter.

There is, even with modern technologies, no way to record the actual perception of the person. However, we can take a glance at it with EEG. Using electrodes connected to the occiput, we can observe at the surface level what is happening within the brain or at least approximate it. Especially if the stimulation is periodic, we can aim to discover the same periodicity within the derived EEG signals. From the wave-form of the recorded curves alone, it is hard or merely impossible even for a skilled professional to discover the portion of the correct

frequency within the signal. That is where technology needs to support us once again. A simple way to analyze the brain signal is by performing a Fourier transformation. However, this has some drawbacks, which were explained in Section 2.5. A more sophisticated statistical approach uses the Canonical Correlation Analysis, which already proves to be efficient in many SSVEP BCI-related experiments. As we learned in the section about Convolutional Neural Networks, they excel at pattern recognition tasks with local dependencies. Therefore, we can experiment with CNNs and brain data by assuming the time series to be just like a 1D image. Furthermore, there is not only one channel we are looking at but multiple, meaning there is a 2D *image* of EEG data over time. We will explore the possibilities and concepts of CNNs for SSVEP based ophthalmologic application in the following subsection.

Before we can jump to the data processing of recorded data, we first need to collect said data. We used the Neuroelectrics Enobio® 8 EEG and an OpenBCI Cyton Biosensing Board. Both devices support 8 Channels, a driven right leg, and a reference electrode. However, we are using a Neuroelectrics neoprene cap to fix electrodes, which is restricting the possible electrode positions. The electrode placement was performed in compliance with the most recent version of the ISCEV standard for clinical visual evoked potentials [OBB$^+$16]. The standard suggests putting the reference electrode at the central position on top of the head at F_z. A graphical representation of the electrode placement is provided in Figure 3.14. The figure portrays the positions of the selected electrodes on the international 10-10 notation system. All iterator integers on the left hemisphere seen from the subject are odd and even on the right side. The most interesting areas for positioning our electrodes for the measurement of visual stimuli are on the occiput. We choose Oz as the central electrode position. Naturally, POz, which is a little closer to the brain's parietal region, would be a pretty good position as well. However, the neoprene cap we are using has no slot for POz, which is why we chose the neighboring electrode positions PO_3 and PO_4. To further increase the spatial resolution, the electrode positions $PO7, O1, O2,$ and $PO8$ were used.

Now we can present stimuli to a subject. However, we need to consider that 0.3 to 3 % of the human population are prone to photosensitive epilepsy [Erb01]. Guidelines and precautions are therefore important whenever working with flashing lights [FHE$^+$05]. To further reduce the possibility of an epileptic seizure, the frequencies between 15 Hz to 25 Hz should not be used for the
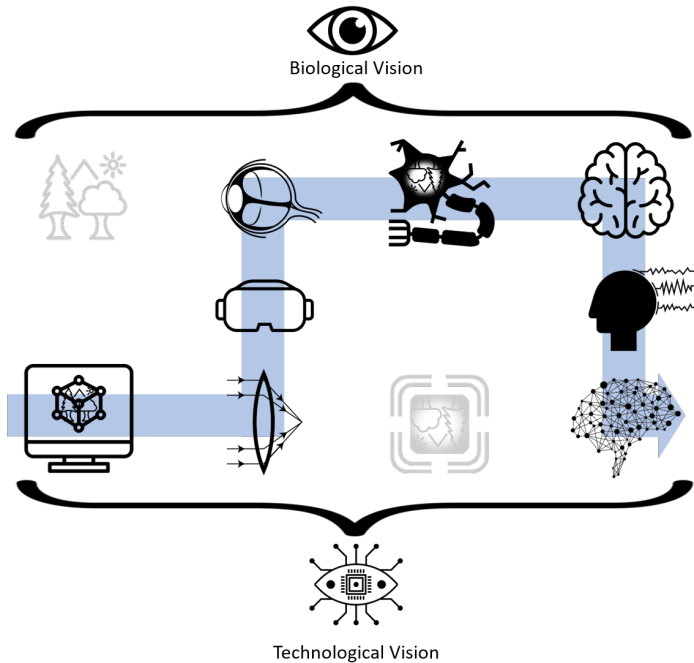
Figure 3.13: The figure shows the meta-level structure of the Metis Videre Project. We start with a computer-simulated environment, which an observer perceives through VR goggles. An EEG-BCI is used to collect information about the perceived stimuli at the back of the head and is finally processed and computed by artificial intelligence.

induction of SSVEP, despite the brain being sensible to stimulation in that frequency range [FHE⁺05].

In Figure 3.12 the setup of Metis Videre is presented. Now that we covered the used EEG device, there are additional hardware components that we need to consider. The following list will provide an overview of all hardware components needed for the platform:

- A Computer to connect to the EEG and run software on
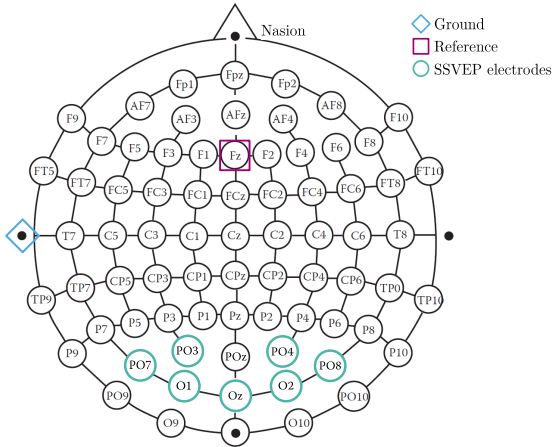
- Visualization devices

Figure 3.14: The electrode placement for the Metis Videre research project is presented in this graphic. The electrodes used for SSVEP measurements are placed at the back of the head above the visual cortex. All measurements are performed with F_z as the reference and a ground electrode positioned at the left earlobe.

Source: *Based on international 10-10 system [JTD07].*

- Computers or server to train AI models

- A headrest with a phoropter

- EEG electrodes

Like most technological systems of today, we need a computer to run the software and the algorithms. This project uses a desktop computer with an integrated NVIDIA GTX1080TI graphics processing unit (GPU) to run machine learning models while rendering stimuli. The rendered stimuli also need to be presented to the user. As the software shall be modular, we will focus on virtual reality HMDs and a high-frequency monitor. Today, most VR devices have a refresh rate of 90 Hz or more; most standard monitors only got 60 Hz. For computer game enthusiasts, there are special monitors with higher refresh rates. Why is that the case? Because if we want to present complex stimuli like in a fast-paced video game, we might experience image tearing at refresh rates of 60 Hz. However, those effects are drastically reduced for refresh rates at 90 Hz and above. In BCI experiments, we rely on displaying clear

stimuli, especially as we stimulate with different frequencies. Furthermore, the refresh rate is restricting the useable frequencies for pattern-onset stimulation. Pattern-onset means the same amount of frames, the stimulus is turned off, and then it is set on for the same amount of frames again. For a device with a refresh rate of 90 Hz that would be all frequencies, where

$$f_{\text{pattern onset}} = \frac{f_{\text{display}}}{2k}, k \in \mathbb{Z} > 0. \tag{3.11}$$

Therefore 22.5 Hz is a valid frequency for a 90 Hertz device, however not for a 60 Hertz. The general rule of thumb is that the higher the displaying device's frequency, the more possible pattern-onset frequencies are at our disposal. Another method of presenting stimulation is sinusoidal stimulation patterns; here, the luminance of the stimulation pattern is adjusted every frame to get a sinus-like shape. The higher the frequency of the monitor, the better the approximation of the sinus. We do not want to use the computer running the lab software for the training of AI models, since algorithm training would prevent lab-recordings with the computer. Because of that, multiple computers with RTX2080 GPUs at the institute were used for AI training. Furthermore, the utilization of computational clusters like the forHLR II (German: Forschungshochleistungsrechner II) was explored.

For brain experiments, we want to reduce the effects of external influences. That is why a headrest is needed to minimize artifacts from bodily movements. An attached phoropter increases the possibility for doing experiments where a varied focal point is needed. This could, for example, be for researching visual acuity, refraction, and astigmatism.

Finally, there are the EEG electrodes which have a tremendous influence on the experiment. They influence the signal-to-noise ratio, and they are also influencing how long the set-up time of the experiment is and whether the user needs to take a shower right afterward.

After covering all hardware aspects of the system, let us have a deeper look at what is making the platform run internally. Again we will first look at the list of components, and then each component will be defined and explained.

- Software-component for interconnecting all software pieces

- Interaction with a user - graphical user interface (GUI)

- Visualization and rendering software for visual stimuli

- Data-structure to define and structure vision-based EEG-experiments

- Utilization of AI

As the Metis Videre platform appears to be a big project from the get-go, we must first and foremost consider how the individual software modules will communicate and interface. For this, a central software control panel in the form of a manager class is needed. For this component, the programming language C++ appears to be the perfect choice as it is platform independent and fast in execution.

The conductor of the experiment also needs a way to communicate with the program. In theory, a command-line prompt could achieve this effect. However, due to the complexity of the platform, a GUI is more advisable. To easily integrate a GUI to the manager class, the open-source software framework Qt is used.

A more complex challenge revolves about displaying the stimuli frame-by-frame in real-time. Libraries intended for C++ often lack the possibilities to display complex stimuli easily (e.g., OpenCV) or are hard to handle and work with due to their complexity (e.g., OpenGL and DirectX). We can obtain a well-balanced software tool when looking at the gaming industry, which has reasonably similar requirements as we do in BCI experiments. The Unity IDE by Unity Software Inc. was built to implement complex graphical simulations in computer games quickly. A major downside of the IDE is that it runs in C. Therefore we will need to interface from the general manager to the C-Unity project.

Another requirement is the ability to define different experiments without re-writing the code. That is why we will need a 'protocol'-file-format. In this protocol, all relevant information about an experiment is saved. This way, an experiment can be redone at a later point in time with the same parameters. Example information within the protocol file is what kind of stimuli is presented, at which frequency, and how long.

Additionally to classical algorithms, which run in C++, we want to focus on AI algorithms. For that, we will need to interface to state-of-the-art AI libraries

in order to run those models. In the next section, we will discuss the concepts of utilizing AI for vision-based BCI experiments.

3.5.1 Machine Learning for EEG Signal Analysis

In Section 3.2 we learned that machine learning is a powerful concept for the realization of pattern recognition in data. Ultimately, the aim of the Metis Videre platform is exactly that, to discover patterns in brain data. However, EEG data is different from the data AI research is usually focussing on. Exciting and promising leaps of technology were made in CNN technology in computer vision. So there is naturally the question, how can we harness this knowledge and apply it to brain data? Remembering the QUA^{3}CK ML process, we already defined the **Q**uestion in place: We want to discover patterns in brain data. In this case, those patterns will prominently be SSVEPs. Therefore, we can already step to the second step of **U**nderstanding the data. In the conception phase, this might even be the most sensitive step, as it is the foundation stone towards the result of the entire system. We know that brain data is recorded using an EEG device. In the case of this book, data was predominantly collected with a Neuroelectrics Enobio® 8 EEG device. The device is measuring the fluctuations in potentials on the scalp where passive electrodes are placed. Generally, we are talking about 5 to 100 μV of potential differences. Therefore the overall setup is prone to artifacts and influences. The electrode placement was already discussed in the previous section and is displayed in Figure 3.14. Furthermore, we introduced CNNs in Section 3.2.3. Therefore, we have to put the EEG data collected from the occiput in an image-like pattern to utilize temporal *and* spatial dependencies. If we look closely, we can see that the EEG data is quite similar to video data. We can assemble all blocks in a 2D grid and add the time/the samples as a third dimension. This procedure results in a data format like presented in Figure 3.15.

Generally, obtaining EEG data is a tedious process. We need to set up the experiment environment, prepare the subject and conduct the experiment. Furthermore, EEG public datasets are a rarity in comparison to labeled image databases. The amount of external data is very limited. Therefore, data augmentation methods like slicing are crucial in order to increase the pool of training data.
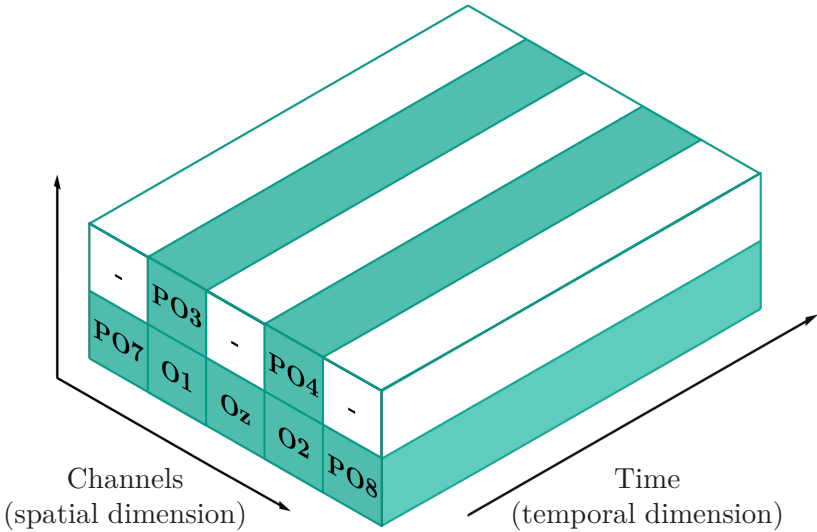
Figure 3.15: The figure shows how the input EEG-data can be organized in a 3D array. The individual channels represent the spatial dimensions, whereas the temporal dimension is produced from samples.

Source: *Previously published by the author in [5]*

Due to the inherently low SNR in EEG recordings, several methods in the **A**lgorithm step of QUA^{3}CK are of utmost importance. We already decided on an image-like approach utilizing CNNs. Many hyperparameters need to be considered, and regularization is crucial to obtain performing models. Therefore, methods like dropout, L1 and L2 regularization, and multi-task learning need to be explored.

In the implementation chapter, we will do a deeper dive into AI and brain signals and explain how CNNs were used to process EEG data.

3.5.2 Handling of EEG Artifacts

As already mentioned, EEG data often has a bad signal-to-noise ratio. It is hard to classify what an artifact or noise is and what is part of the brain's signal. Many things might influence the signal obtained in EEGs. The most significant influence probably comes from movement. Generally speaking, everything that will produce an EMG (electromyography) signal disturbs the sensible EEG signals.

It might cause electrodes and connection cables to wiggle, changing electrical properties. In most modern environments, there is much electro-smog, radiation produced by nearby electrical devices, cables, outlets, smartphones, computers, et cetera. All these devices have inductive effects on the acquisition of the EEG reading. An example of an EMG artifact is the clenching of the jaw, which produces muscle-activation artifacts in the EEG signal. The eyes themselves are electronically speaking dipoles, and looking around is already influencing the EEG signal. Furthermore, our cornea wants to stay healthy and moist. We have a natural blinking reflex, which produces blink artifacts.

So how can we handle those artifacts which might occur while recording EEG signals? The best way to manage artifacts is to not produce artifacts in the first place. As the experiments we conduct are in a laboratory setting, we can define some general precautions. Sure we cannot force subjects to stop blinking during measurements, but we can encourage them to minimize blinking while being stimulated. Also, we cannot eliminate all other electrical devices around. However, we did some measurements in the antenna measuring chamber of the institute. There were *no* significant differences in SNR, so we assume that the before mentioned electro-smog has little influence on EEG-acquisition quality if no conducting wire, phone, or similar is closeby to the EEG device. What we can manage are physical artifacts from the body and head movements. We can handle them by utilizing chin rest where the subject is asked to rest their head upon. Those chin rests are common in ophthalmology. Therefore, using a chin rest is the easiest way to eliminate the influences from movement-induced artifacts.

However, we can be confident that we will never be able to eliminate all artifacts. Therefore, for algorithm development, we got two possibilities. First, we could try to classify which signals are good for processing and which ones are heavily obstructed by artifacs. Alternatively, we could try to build

algorithms that can handle or even understand artifacts. AI excels in the task of handling sparse and noisy data. Therefore, the second approach will be discussed in this dissertation. This is especially the case for the approach of end-to-end learning where labels are known or provided.

3.5.3 EEG Electrodes

As previously discussed in the Fundamentals (Section 2.3.2), the electrodes used are crucial for the overall BCI setup. They are the essential piece of equipment connecting humans with the machine. Their influence on the signal-to-noise ratio is immense. As we already learned, adhesive electrodes attached directly to the skin are state-of-the-art in terms of signal quality. However, they bring one major downside with them: The subject needs to shave their head. From a strictly natural-scientific point of view, there is no issue with that.People tend to be emotionally attached to their hairstyle from a psychological aspect as part of their personality. Furthermore, their hair keeps their head warm in the cold. Therefore hair is often only shaved to match a style or when the severity of a condition defies the means for obtaining better signal quality.

On a side note: The author has a shaved head and is, therefore, an excellent candidate for evaluating and comparing all kinds of electrodes. Furthermore, even some students who wrote their theses within the Metis Videre project shaved their heads for science. However, most students and subjects, especially females, are not that willing to do that. Ultimately, when going to the ophthalmologist in a real-world setting, no one probably wants to get their head shaved for an eye exam.

The second best passive electrodes for the measurement of brain signals are gel-filled electrodes. They still provide a reliable signal, and we do not need a shaved head for their application. However, they leave residues behind in the hair. They are making a shower obligatory after a measurement. Again, this kind of inconvenience is already too much when considering the application in ophthalmology. The third electrodes we might want to consider are the dry electrodes. Usually, they are made from plastic and coated with a layer of Ag-AgCl or any other conductive material with free charge carriers. Despite the considerable advantages in recent years in dry electrode technology, the SNR is still not closely en-par with their wet or adhesive counterparts. This is due to

high resistance between electrode and skin and poor conversion from ionic to electrical current [LGSMV14]. The main question is whether dry electrodes still produce enough signal quality to produce the results we are looking for. Moreover, even machine learning methodologies cannot discover signals where almost no signal is left. Therefore, we will focus in this dissertation on innovating electrodes as well.

We want to combine the desirable features of wet electrodes:

- Good signal-to-noise ratio

- Reuseable by just exchanging gel

with the benefits of dry electrodes:

- No residues on the head

- Functionality independent of existing hair

- Short setup time

How do we achieve all those effects? What structure should an electrode fulfilling these requirements have? Let us work through the requirements one by one and do some thinking. First, we want to achieve a well SNR. There are a lot of new materials which could be considered for the electrode like carbon-nano-tubes [RDF+06]. However, due to the skin, dry materials will always be prone to artifacts and fluctuations in SNR. The benefits of dry electrodes are, in fact, just all the disadvantages of off-the-shelf wet-electrodes. Therefore, the only reasonable approach for us is to use a liquid or conductive gel.

To have no or little residues in the hair or on the head, we need a design that allows only a minimal amount of gel to get into contact with the skin. Furthermore, the functionality should not be compromised by hair. Therefore, state-of-the-art dry electrodes have long slim legs all around them, which stick through hair and reach the surface of the skull. However, due to the last sub-millimeters of hair follicles and hair, the contact is often imperfect. However, let us take the idea of legs and combine it with gel. The straightforward approach would be to fill the electrode with electrolyte and use the drilled holes in the legs to exploit capillary forces to transport electrolyte to the skin. Thanks to manufacturing techniques like 3D printing, we do not need to drill

at all. We can design models in a CAD software tool and print the results right away.

Furthermore, we can even make the electrodes compatible with the electrode heads of state-of-the-art wet-electrodes, thereby improving compatibility. With these steps, we probably will already reduce the residues significantly on the head while still maintaining robust signal-to-noise quality. However, the more gel we introduce, the more the head will get soaked. So the next step is to make sure that not too much gel is left on the head. To solve this issue, we might want to consider multiple tubes within the electrode and deflection volumes.

4 Implementation

This chapter will provide insight into the performed implementations. Initially, the VR vision simulation platform will be discussed. Afterward, we will review the implementation of the Metis Videre Platform from a software, hardware, and data science perspective.

4.1 Vision Simulation Environment in VR

Visual fields like the one presented in Figure 2.12 represent visual sensitivity of a patient. The values are typically between <0 and 40 decibels (dB). Where <0 occurs if a patient is not able to perceive even the brightest possible stimulus of the device. The grayscale representation of the scan is presented in Figure 4.2 (A). The 24-2 scan performed on a standard perimeter like the ZEISS Humphrey Field Analyzer records data of the visual field, usually in a range of 54° horizontally and 48° vertically. A modern VR-HMD has a field of view of at least 90°. Therefore, the grayscale map has to be extrapolated, compare Figure 4.2 (B), to match the screen size of the VR-device. In a last step, the final grayscale 'maskvalue' map is generated. The individual pixel intensities are calculated using the value of the tile where the pixel is located (KW_e) and all neighboring tiles (SW_x, KW_y, KW_d):

$$maskvalue = F_e \cdot KW_e + F_x \cdot KW_x + F_y \cdot KW_y + F_d \cdot KW_d, \qquad (4.1)$$

the factors (F_e, F_x, F_y, F_d) are describing how strongly a neighboring tile is influencing the maskvalue at a certain position and they are calculated as follows:

$$F_x = \begin{cases} 0,5D_x - 0,25D_y & \text{for } x > y \\ 0,25 & \text{for } x = y \\ \frac{D_x^2}{4D_y} & \text{for } x < y \end{cases} \qquad (4.2)$$

$$F_y = \begin{cases} \frac{D_y^2}{4D_x} & \text{for } x > y \\ 0,25 & \text{for } x = y \\ 0,5D_y - 0,25D_x & \text{for } x < y \end{cases} \qquad (4.3)$$

$$F_d = \frac{D_x + D_y}{8} \cdot \left(1 - |D_x - D_y|\right) \qquad (4.4)$$

$$F_e = -0,5 \cdot \left(\max\{D_x, D_y\} + F_d\right) + 1. \qquad (4.5)$$

For better understanding Figure 4.1 explains the dependencies of a pixel P in a tile KW_e. Furthermore, the presented software lets the user experience a virtual domestic environment and experience and interact with it with common VR-input modalities. During those experiences the glaucoma filter can be adapted to show earlier or later 24-2 scans of the simulated or viewed patient.

The final result of applying such post-processing render mask to a VR scene is displayed in Figure 4.3. The mask shown in Figure 4.2 is herein applied to the right eye (B). It shows an advanced case of glaucoma developed in the top left corner of the visual field and is also prominent within the central macular vision. The graphical user interface (GUI) of the developed simulation environment is presented in Figure 4.4. All implemented effects can be selected on the right-hand side using slide controllers. While different presets for individual progressive states of glaucoma, cataract and AMD can be adjusted on the left side. Furthermore, for glaucoma, there is the possibility to load real-world data into a folder of the software. For this case, the user can choose between rendering the between left eye mode, right eye mode or binocular mode. It is interesting to experience how perception is experienced in the brain in binocular mode if one eye is affected more severely.

In the following the effects implemented will be presented and briefly explained.
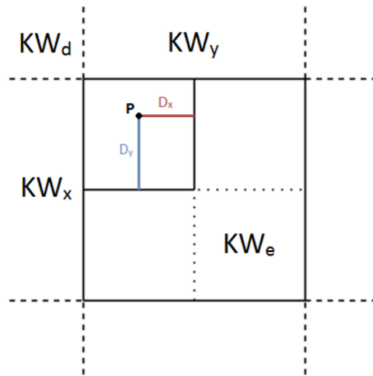
Figure 4.1: Graphical representation of the used variables for the calculation of the post-processing filter for the glaucoma effect.

Source: *Previously published by the author in [10].*

Figure 4.2: (A) shows a rasterized, transformed 24-2 scan (from Figure 2.12). It is extrapolated in (B). In (C) the post-processing mask is smoothed.

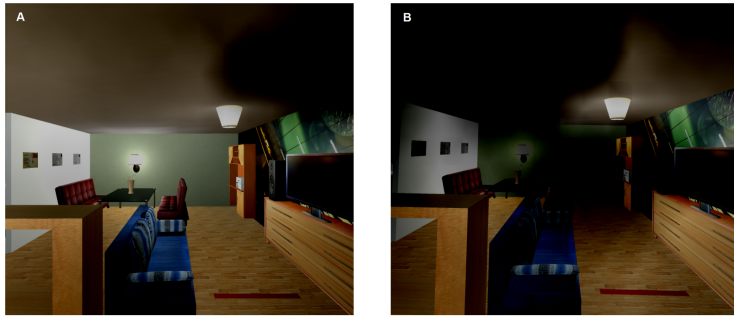**Source:** *Previously published by the author in [10].*

Figure 4.3: A screenshot of the VR environment like experienced by a user is presented. (A) displays the viewport of the left eye whereas B shows the right eye. A mild case (A) and a severe (B) case of glaucoma were rendered. The right (B) rendering was generated using the 24-2 scan from Figure 2.12 which was extrapolated like displayed in Figure 4.2.

Source: *Previously published by the author in [10].*

Figure 4.4: Display of the settings in the menu scene. On the right side, the strength of the different disease effects can be adjusted using sliders. On the left side, presets for progressive vision diseases can be selected. For glaucoma, it is possible to display imported perimetry data. A menu for basic functionality is provided in the center.

- **Gaussian Blur** is used to simulate reduced contrast sensitivity. It is implemented using a two dimensional gaussian filter mask where the dimensions are defined as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}. \tag{4.6}$$

Which is the continuous Gaussian function. The higher the selected variance σ, the stronger the spread of the Gaussian function and therefore the reduction of contrast. To apply the function to discrete pixels we need to perform a convolution with a discrete Gaussian mask. This kernel mask can be calculated using

$$f(x) = \int_{x-0.5}^{x+0,5} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, X \in \mathbb{Z} \tag{4.7}$$

where x is iterated through all elements of the kernel on the interval $[(\text{Kernelsize} - 1) \cdot (-0.5), (\text{Kernelsize} - 1) \cdot (0.5)]$. The kernel size has to be a natural non-even number. Finally, the sum of all weights has to be 1, thus the weights are normalized over the kernel. The next step is the convolution of the screen image with the filter mask and to sum up the results. The value of this sum is assigned to the center pixel. In the case of this work, we first convolve in the x-direction and then in the y-direction to obtain a two-dimensional convolution. This separate one-dimensional convolution in each direction allows the computation time to be reduced from $O(n^2)$ to $O(2n)$, in contrast to direct two-dimensional convolution [Jäh05].

- **Distortion** Distorted vision (metamorphopsia) is a characteristic disease effect for wet AMD. In this disease effect, lines are not perceived as straight but wavy. This effect is simulated by rotating the pixels within the range r around the center of vision. The effect was implemented analog to Banks et al. [BM08].

- **Yellowing and Greying** are modifications of the RGB color values of the final image. The easiest method to achieve greying is using a gamma

value Y for each pixel as a grey value. The easiest way to achieve this is by equally weighting each color

$$Y = \left(\frac{R + G + B}{3} \right).$$

(4.8)

However, the human eye's color perception is not equal in the three color channels. Therefore, we will use the following notation from [BB06]:

$$Y = (w_R \cdot R + w_G \cdot G + w_B \cdot B)$$

(4.9)

with $w_R = 0.2126, w_G = 0.7152, w_B = 0.0722$. To calculate the final pixel values, we introduce the effect intensity I. I is either chosen in the GUI settings or set by a preset selection (for example Cataract of level 3). The final red, blue and green color values C are calculated as

$$C = C \cdot (1 - I) + I \cdot Y, I \in [0, 1], C = [R, G, B].$$

(4.10)

. The yellowing effect is technically the same as the greying, with the only difference that the blue color channel is modified.

- **Loss of central vision** like exhibited regularly in age-related macular degeneration, the central part of the vision is without information. The size of the blacked region is called r_b. Up to $\frac{r_b}{2}$ pixels are painted black. To prevent blackening looking uniform, we do not use a linear decrease but a cosine function.

- The **Lens-flare-effect and halo-effect** were implemented using existing functionalities of the Unity software development kit. The light components implemented in the scene are *point lights*. Through the properties of those lights, a *lens-flare* with adjustable brightness and color is added. Analog to this, the halo-effect was achieved by using the standard effect *halo* for light sources in unity.

- The **Tunneling effect** was implemented reciprocal to the 'loss of central vision'.

- **Distortion of central vision** was implemented similarly to the general 'distortion' effect. The difference, however, is that instead of a rotational
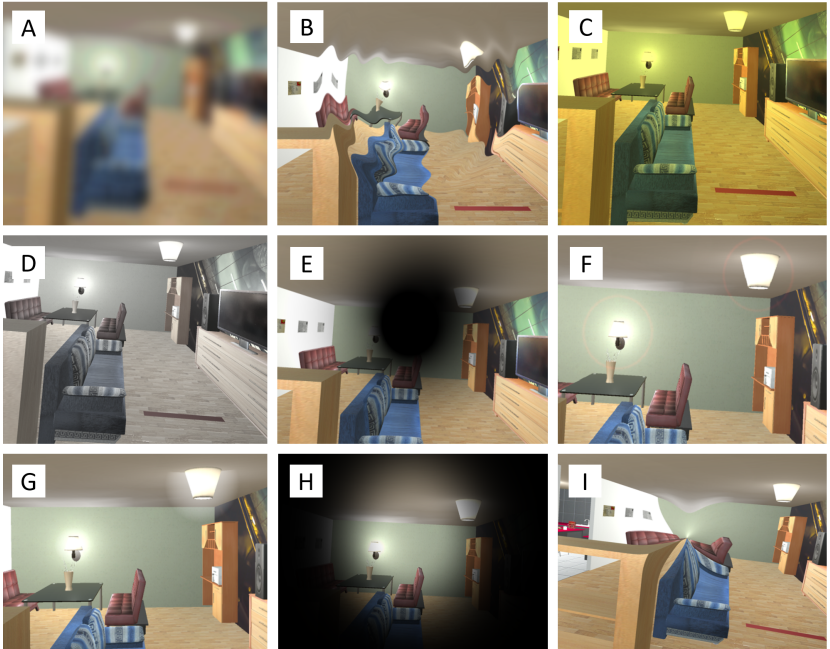
Figure 4.5: The figure shows all implemented simulation effects in the VR environment. In alphabetical order: Gaussian blur, distortion, yellowing, fraying, loss of central vision, halo-effect, lens-flare-effect, tunneling effect and distortion of central vision.

transformation of the pixels, a plain translation vector $\vec{P}$ was used to achieve the effect of a sinkhole-like distortion pattern.

4.2   Software of 'Metis Videre'

This section will focus on the software developed for the Metis Videre Platform we introduced in Section 3.5. The Metis Videre Platform in itself is the apex of vision research performed in this work, and the most crucial part of it is the inner workings of the system's software. Furthermore, we need to consider hardware-software co-design challenges as the overall system comprises hardware and software modules. The hardware aspects will be discussed in the next Section (Section 4.3). Due to the increasing complexity of artificial intelligence and algorithms, we will dedicate a whole section to data science aspects in Section 4.4.

We already introduced the system's overall concept in the Figures 3.13 and 3.12. Now we will do a deeper dive into the software's inner mechanics and what components are necessary for the platform to function properly. As already mentioned, the software revolves around computing machinery, displaying hardware, and a BCI. While the BCI and the displays are hardware like an HTC Vive Pro VR-HMD or a Neuroelectrics Enobio 8, the software is the thinking brain that interacts with those peripheric devices. Pictographic representations of the devices and their interaction with other modules are displayed in Figure 4.6. The Software Development Kit (SDK) from Unity Technologies called *Unity* is used for the rendering of complex stimulation patterns (*Graphics Engine*). The SDK is famous for its cross-platform capabilities, which allow the simultaneous development of solutions for desktop displays, immersive VR-HMDs, and even smartphones. Furthermore, it uses the modern programming language C# for all target platforms, meaning that one code backend can run on different setups.

The EEG BCI (e.g., Neuroelectrics Enobio 8 or openBCI Cyton) interfaces with the computer via wifi. Compared to a cable-bound EEG, the advantage of wifi is that fewer movement artifacts are produced. Most modern research-grade[1] EEG systems provide an LSL (Lab Streaming Layer) stream with the information gathered in the BCI (EEG readings and accelerometer data). Using the LSL, we can access the brain data in real-time and run algorithms on the data. Furthermore, the Metis Videre backend is able to write labels

---

[1] EEG devices that are safe to be used on humans have a good signal quality but do not necessarily have a medical device regulatory clearance.
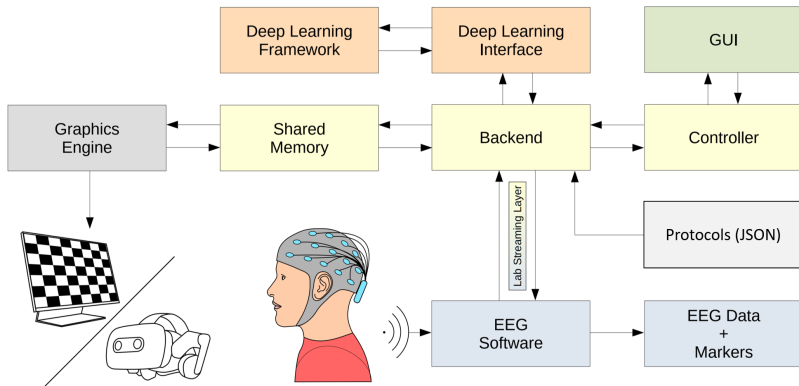
Figure 4.6: The figure shows the individual software components of the Metis Videre platform.
Source: *Adapted from previously published by the Author in [5].*

into the LSL. The final stream, including labels, can then be saved onto the
hard drive for later use. The already mentioned backend is the central piece
that holds all the pieces together. It manages the basic functionalities of the
software and also manages all data flow. To its left (compare Figure 4.6), it
communicates using a shard memory handle. This is because the backend
is written in C++. C++, just as Unity, is very versatile regarding platforms;
however, the projects do not share the same code-base, nor are C++ and C#
entirely compatible. Especially when handling real-time algorithmic processes
and machine learning tasks, a C++-based project has quite some advantages
for high-performance algorithms. Even Google's Tensorflow Deep Learning
library is running in C++. So for the main project and the Unity project to
communicate with each other, we need a fast and reliable method for the ex-
change of information between the two individual threads. One possibility
would be to use TCP (transmission control protocol). However, this will take
a route over the computer's network adapter. Another method would be to
write information to a file destination. On the one hand, this will take a long
time for the access, read and write inquiries and probably will not be stable
when different programs and threads try to access the same location in rapid
succession. A method that does not rely on TCP technology or file transfers
uses the computer's main random access memory (RAM). One thread will
write a message at a designated position in the RAM and set a semaphore, a

generalized mutex, and then the other thread can access this RAM location and vice versa. This process is not easily explained in a few sentences. Therefore, a detailed description of this process will be provided in Section 4.2.1.

Right above the *backend* there is the 'Deep Learning Interface', its job is to convert EEG data and other inputs so the Tensorflow deep learning framework can process it. Without the interface handling data types, loading AI networks, and reporting back results, no machine learning-based data processing would be possible.

To execute pre-defined protocols, we need a method to interact and store the procedures of the protocols. The easiest method for this would be pre-defined routines. That would be static and limit scientific studies' progress as every change in the experiment would need a change in the code-base. Therefore, to have better generalizability of the software, we introduce a file-based protocol format. This could, for example, be achieved using Comma Separated Value (CSV) files. The extendability of those, however, is very limited. Therefore, JSON, a modern hierarchical structure, is used instead [44].

One last important component we have not talked about is the graphical user interface, the 'GUI'. It is not directly implemented into the backend, nor is it one step away, but it uses a controller to access the model stored in the backend. The paradigm behind this will be covered in the next section. The GUI is the central piece of software with which the research scientist has direct interaction. It is used for the definition and creation of protocols and to supervise running experiments. Next to the renderings from Unity, it is the only visible part of the software. However, in contrast to the visual stimulation, it is generally not viewed by the subject but by the experiment supervisor or research scientist. This visible part is programmed in the open-source software Qt, which is also platform-independent and utterly compatible with the 'backend' as it is also written in C++.

Now that we covered all significant components of the software, we will further analyze the following subsections some used design patterns, the overall software architecture, the user interface, the creation of stimuli, and how the protocols work.

4.2.1 Software Design Patterns

Design patterns are an important matter to consider when writing good software. The implementation of patterns helps to solve problems that are common in computer science. However, the use of patterns should never compensate for the complex thought processes which are behind a successful software architecture. Initially, we will discuss the model view controller paradigm and its advantages. Then we will explore the exchange between two disjunct programs using shared memory and semaphores.

The ultimate goal of the model view controller (MVC) design pattern is that it enables the reusability and interchangeability of individual components [KP+88]. It is composed of three individual modules. The first is the *model*, which encapsulates data and the logic behind the software. In the case of Metis Videre, this is the *backend*. The second component is the *view*, which is the part that is seen by a user interacting with the software. Therefore the *view* hosts all buttons, labels, edit-blocks, et cetera. The third component is the *controller*, every interaction with the GUI is directly routed through the controller. The controller, on the other hand, changes parameters, dataflow, and settings in the model. One might ask: Why is there so much effort put into the structuring of the GUI? With extensive or longer continuous developed software projects, parts of the software are always subject to change. The GUI, in particular, can see many iterations as the codebase changes. The MVC design pattern provides easy interchangeability of components. The model can change without the need to exchange the view or the controller. The same is true for the view and the controller. Especially the interchangeability of the view is interesting from a design perspective. The design can easily be changed or adapted for use on other platforms like smartphones. A graphical presentation of the MVC paradigm is presented in Figure 4.7.

During the lifetime of the Metis Videre project, the GUI was regularly updated and adapted and even had a complete design overhaul (compare Figure 4.8 to Figure 4.12). Thanks to the MVC design pattern, changes were performed efficiently without interfering with other software backend parts.

Another prominent part of the software architecture is the shared memory communication handle. A visualization of the data handling process between Metis Videre C++ and Unity is presented in Figure 4.9. The interface itself is written in C++ and packed as a dynamic-link library (DLL). This DLL can be
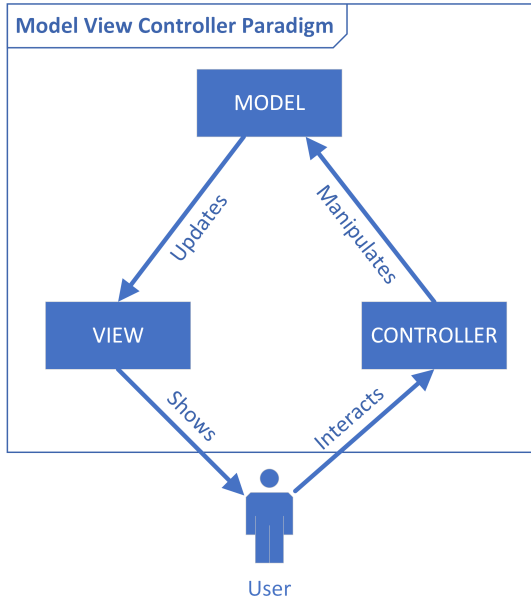
Figure 4.7: The Model View Controller Paradigm. The user interacts with the controller, which manipulates the model, which in turn updates the view, which is presented to the user.

included just normally from the C++ side, and from C# (Unity) side, it can be used as an 'unmanaged' package. Unmanaged in terms of memory handling. C# is a managed programming language, meaning it has commodities like a garbage collector. However, it is possible to access C++ DLLs from special unmanaged C# code. The two streams of data are the stimulation parameters from the Metis Videre C++ software to Unity and a stream of feedback information in the other direction. The stimulation parameters include all information needed to plot a stimulus (e.g., type of stimulus, frequency, number of checkers, size, et cetera). On the other side of the equation, the feedback structure contains the label of the currently displayed stimulus for reference and a timestamp for synchronization. The shared memory can be accessed by the Metis Videre Unity and the Metis Videre main thread, potentially simultaneously. To tackle this problem, we need to control the access to the stored information so that there are no access violations which in the worst case could
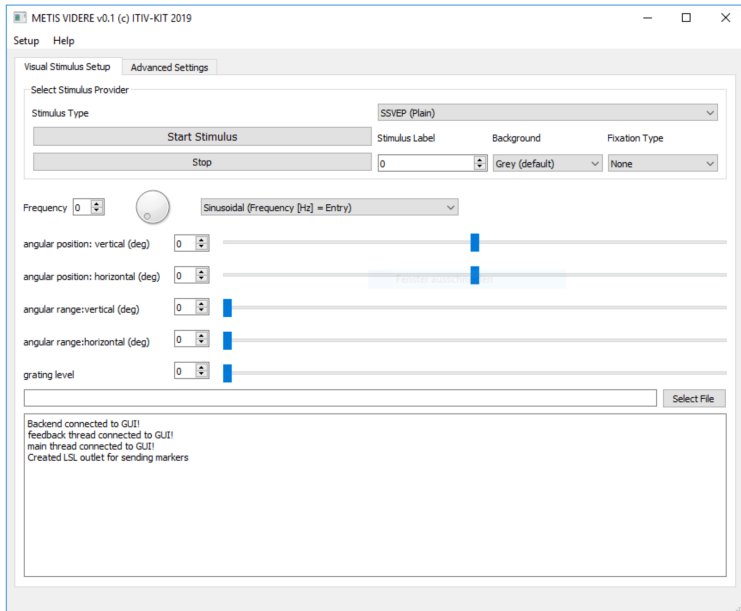
Figure 4.8: The figure shows the old version of the Metis Videre user interface. Thanks to the MVC design pattern, GUI updates could be performed without a complete software overhaul.

lead to a critical system failure. We achieve this by introducing semaphores to protect the shared memory. Both streams of information are equipped with a semaphore which can be increased. The increment indicates that new information was written to the shared memory and can be accessed/read to the Unity application. After reading the parameters, Metis Videre Unity will save the content to its process-specific memory and decrement the semaphore. Multiple parameters can be written to and accessed from the shared memory at the same time. The path from Unity to the main application works the same way. To conclude: We use the direct input from the user on the UI or, better, protocol files for the definition of the experiment. This experiment definition is posted through the shared memory interface utilizing the interface DLL on the C++ side. The semaphore is incremented. The unity application notices the increment, indicating new data within the shared memory. The Unity application is
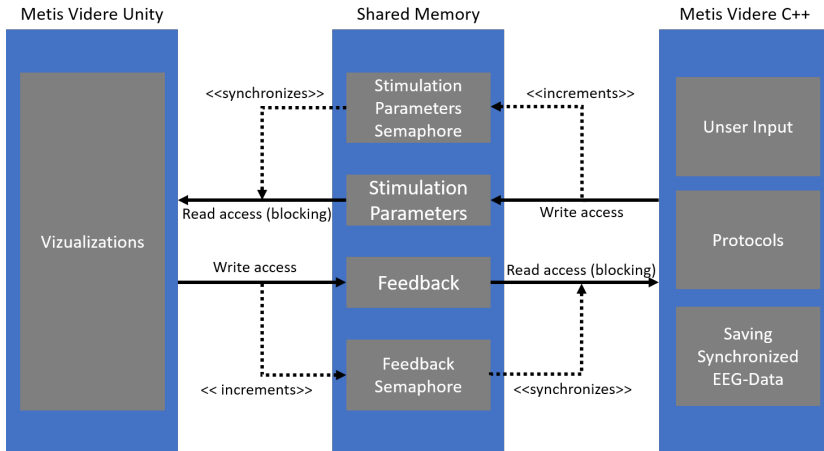
Figure 4.9: The figure shows the communication between the C# based Unity and the C++ based main project of Metis Videre. The main project is writing instructions to the shared memory while the Unity application is reading them. Semaphores are used to synchronize memory access and prevent memory access violations. The same is happening the other way round, in a feedback stream from Unity to the main application, including information about the currently displayed stimulus.

accessing (and therefore blocking) the most recent data package. After it was read, it is internally processed in the Unity thread. Then the visualizations are presented. The visualization happens with the speed of RAM-access in real-time. Unity again accesses the shared memory to give feedback. It provides the label and timestamp back to the main application. The stream of EEG data is synchronized with the feedback stream, and the stimuli labels are added to the LSL stream, ultimately saving annotated EEG files.

4.2.2 Software Architecture

We already discussed two very central elements of the Metis Videre Platform, the 'Shared Memory Interface' and the MVC GUI. To achieve a better overview of the overall software architecture, we will discuss the used languages and formats of the Metis Videre Platform (compare Figure 4.10). The Metis Videre Main application is written in C++, while Unity is primarily written in C# with
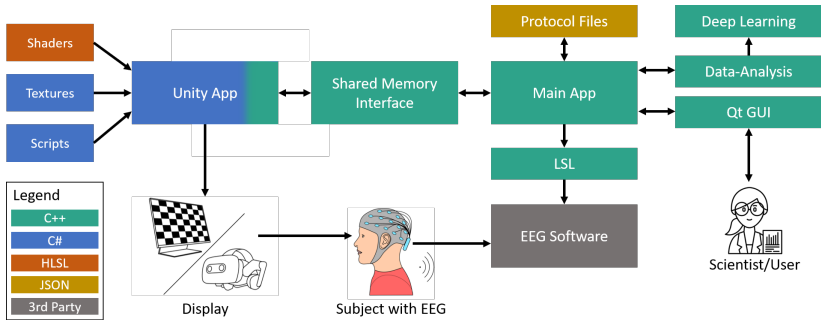
Figure 4.10: The graph shows an overview of different software components and their languages. The core components of the software are programmed in C++, while visual stimulation relies on the Unity platform, which utilizes C# and the DirectX-based High-Level Shading Language (HLSL).

only the possibility to interface with C++ using unmanaged code. When writing high-performance graphical representations of, for example, checkerboards, we do not want to write those in C# but in a language closer to the graphics processing unit (GPU). This language is the High-Level Shader Language (HLSL), used in Unity to describe a shader's mathematical properties.

The EEG software provides initial data handling of the wifi-transmitted EEG data streams. This software is proprietary of the EEG-device manufacturer and can interface with the LSL[2]. The experiment protocols are written in JSON (Javascript Object Notation), a hierarchical markup language.

In the following, we will discuss the architecture of the Metis Videre main application in the form of an abstract class diagram like shown in Figure 4.11. When comparing Figure 4.11 with Figure 4.10, the unified modeling language (UML) class diagram is displaying the software objects within the *Main App* and the other C++ components. The *BCI_Manager* is created with the startup of the application. It creates and destroys (composition) the *Controller* which is responsible for the interactions with the Qt GUI (compare Figure 4.7). Furthermore it composes the *EEG_Record* class. This class is the component

---

[2] There are LSL implementations for many languages, such as C, C++, Python, Java, C#, MAT-LAB, however, Metis Videre only uses LSL from a C++ perspective.
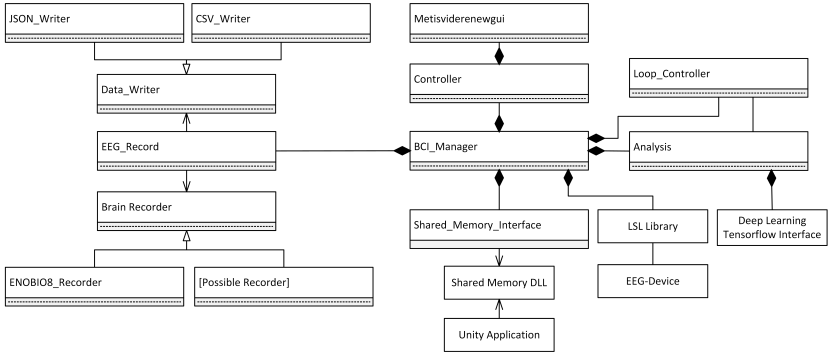
Figure 4.11: The Metis Videre main application class diagram. The *BCI_Manager* class handles the data and control flow of the application and is created at the start of the program. It creates other important objects which handle the graphical user interface, algorithms, communication to Unity and the EEG device. For better readability, methods and attributes are not included in the UML diagram. Simple white boxes represent used packages and libraries.

of the software that is responsible for reading and writing EEG data. For better generalization the *Data_Writer* and *Brain_Recorder* are built as parent classes. Thanks to this architecture, it was easy to exchange, for example, CSV-based file formats with JSON-based file formats during project runtime. On the other side of the equation, multiple EEG devices can be included without changing the inner mechanics of the software as the manager class will always perform the access via the *EEG_Record* class which accesses the *Brain_Recoreder*. In the displayed UML, there is a specific child recorder, the *Enobio8_Recorder* which enables the connection to the Neuroelectrics Enobio 8 EEG device through LSL. Further *Possible Recorder*s can be added on-demand, making the platform very versatile for the use of different EEG recording devices.

The class *Analysis* is responsible for simple algorithmic calculations like a live plotting of a CCA (compare Section 2.5.2) from incoming EEG data. It is connected with the *Loop_Controller* enabling closed-loop functionalities within the software. Finally, the *Deep Learning Tensorflow Interface* is a library implemented for this project which is connected to a Google Tensorflow backend. With this backend, it is possible to load and execute machine learning models (usually H5-files).

The *LSL Library* helps to connect to EEG devices which support the LSL-protocol. The *Shared_Memory_Interface* interfaces with the for this project built *Shared Memory DLL* which connects to the *Unity Application* through a Shared Memory space (compare Figure 4.9).

The implementation outlined in Figure 4.11 is the high-level abstraction of the underlying C++ of the Metis Videre main application. However, it shall be noted that an EEG-based BCI research platform can be built upon different principles and that the chosen methodologies were the ones that were chosen for this project based on the state of the art software architectures and design principles.

4.2.3 User Interface

In modern IT and information systems, almost all modern solutions use a Graphical User Interface (GUI). This kind of interaction method is more intuitive than sole text-based command prompt input. Furthermore, new software users get an easier and better understanding of the functionalities of the software through visual representations. The Metis Videre platform is intended for research purposes, and a researcher should be able to administrate and control the program and experiments easily. For this, the GUI is a key component of the software. As already mentioned, it is based on the model view controller principle. The Figures 4.12 and 4.13 show the landing page and advanced settings, respectively. The most prominent buttons are the Start and Stop/Reset buttons. The Start button is used to send the selected stimuli settings to the Unity application. The Stop button is self-explanatory as well, as it stops the current stimulus and returns the screen to a neutral background. The most important selection option is the *Stimulus type*. In Figure 4.12, the SSVEP (Circular) stimulus is selected. Depending on the chosen stimulus, the other options on the page change dynamically. For example, a radial frequency is only useful for circular stimulation patterns and not for rectangular. The GUI further allows for advanced customization of the experiments, like the setup of the system. Because the angular position and range are dependent on the distance of the subject and the screen size, those parameters can be set in the *Advanced settings* menu. This advanced settings menu is displayed in Figure 4.13.
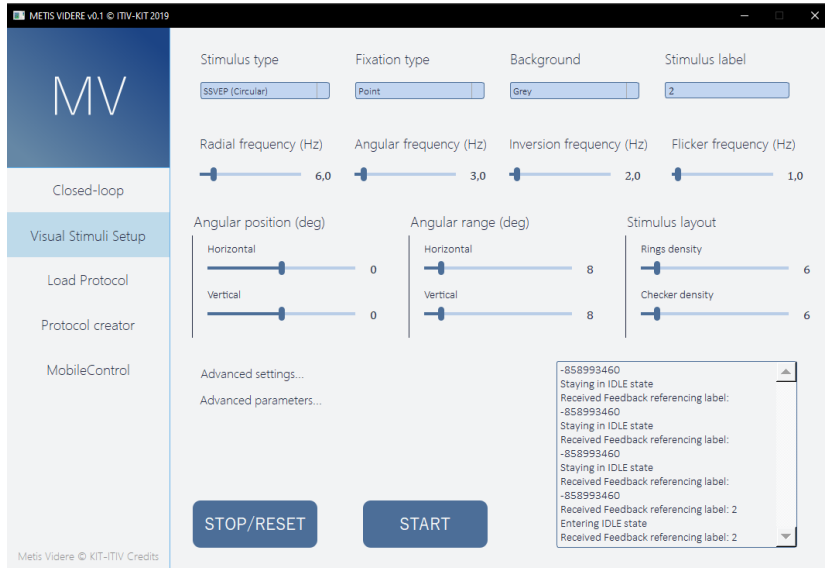
Figure 4.12: A screenshot taken from the latest build of the Metis Videre software. It shows the landing page, which is used for the setup of visual stimuli. The researcher can select and modify stimulus-specific properties like the stimulus type, background, frequencies, size position, and more.

On the left side of the GUI, there is a bar where further sub-menus can be selected. In the closed-loop menu, real-time experiments can be executed, and results from CCA-analysis or even machine learning algorithms can be displayed in real-time. In the protocol creator, experiment protocols can be defined. A protocol is just the concatenation of a multitude of visual stimuli. The protocol creator generates a dynamic list of stimuli that can be created within the *Visual Stimuli Setup* and played at a later point in time as an experiment. These protocols can be saved to the hard drive in a JSON format. This way, it can be ensured that conducted experiments are comparable and repeatable across a study. To load and execute a protocol, we can go to *Load Protocol* tab. The last tab is *MobileControl*, which is an experimental menu for accessing smartphone displays and using those for the display of stimuli.
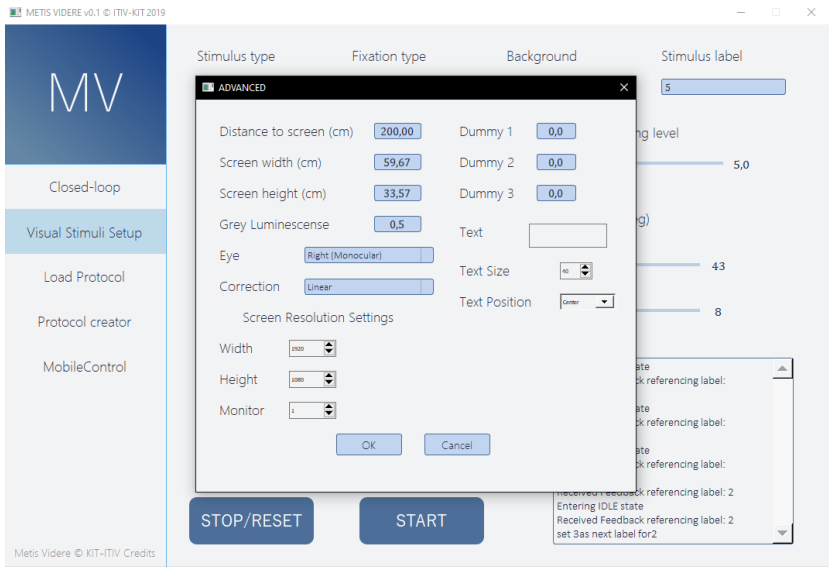
Figure 4.13: This screenshot shows the advanced settings which are possible to be set for the experiments. Important parameters are, for example, the distance of the subject to the screen, the screen size, the luminance, and rendering settings for VR. Furthermore, some dummy variables are available for the modification of complex or experimental stimuli.

Now that we talked in detail about the GUI and how it can be used to create protocols and stimuli, we should also look at the stimuli displayed in Unity. Therefore, the next section will discuss the Metis Videre platform regarding stimulation patterns.

4.2.4 Stimuli

One of the primary aims of the Metis Videre Platform is to deliver precise stimulation patterns in real-time. Real-time in the case of a standard commercial VR-HMD like the HTC Vive means it can render an image at 90 Hz with a resolution of 2880×1600 pixel. Especially for higher frequency stimuli, which in the case of pattern onset can be up to 45 Hz, fast rendering capabil-
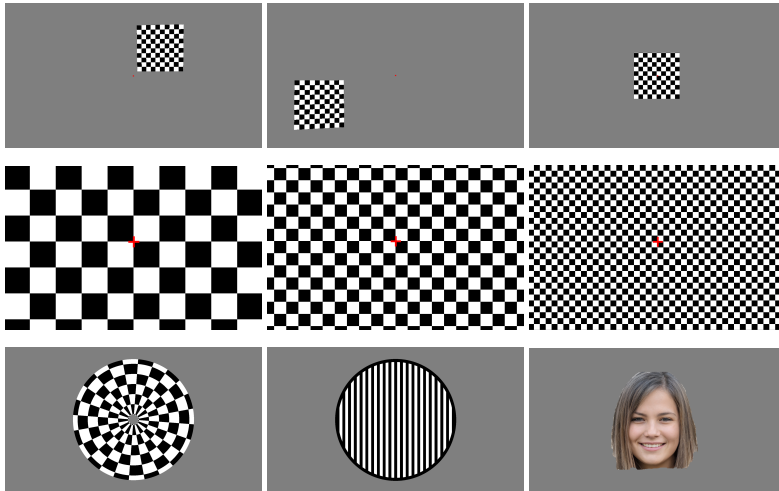
Figure 4.14: The figure shows different showcases of visual stimulation patterns. The top row is displaying a quadratic checkerboard stimulus of a size of 10° times 10°. It is displayed at positions (6°, 6°), (-16, -6°) and (0°, 0°). The middle row shows a sweep VEP. In this case, the spatial frequency of the stimulus is swept from a low frequency to a high frequency. The bottom row has more complex stimulation patterns like circular stimuli and images.

Source: *Parts of the figure were already by the author at [5]. The face is GAN-generated from www.thispersondoesnotexist.com (accessed December 2020).*

ities are crucial. Fortunately, a fast render queue is one of the advantages of using the Unity platform. In the Fundamentals of Section 2.3.3, we already introduced how visual stimulation for SSVEP generation works in general. Figure 4.14 shows how those stimulation patterns look like as taken directly from the screen. We can produce stimulation patterns at any given position on the screen, stimulating different positions of the retina. Furthermore, we can display a red fixation dot (upper row), target (middle row), or choose not to use fixation at all (bottom row).

We can notice that the background is a special kind of grey, that grey level is luminance adjusted so that the mean value of white and black pixels equals the grey value. That value is specific to a specific monitor, and Section 4.3.2

will explain how to perform this so called gamma correction. However, why is gamma correction necessary? When sweeping a stimulus (compare middle row) in spatial frequency, at some point, the grey level of the background and the combination of black and white will blend into one color. If the luminance is corrected, the screen appears to be static. If it is not corrected, a flashing of the overall screen is perceivable. Gamma correction is especially cruicial when designing experiments regarding visual acuity, where we want to experimentally find the smallest checker-size the eye and brain can resolve. In the bottom row, on the left, there is a circular stimulation pattern analog to Figure 2.15. In the center, a circular stimulation pattern with bars instead of checkers is presented. This form is of interest when the orientation of the edges of the stimuli is of importance for the researcher, for example, when exploring the influences of astigmatism on perception. An example experiment would be to correlate the brain-response to different orientations. By this it would be possible to reconstruct the astigmatism angle [41]. The last stimulus is an image. Such complex stimuli, especially faces, might induce responses in higher cognition when presented to subjects [4, 40]. Therefore, the research platform must be able to render complex imagery at different frequencies. All presented stimulation can perform pattern onset, reversal, movement, and gradual sinusoidal stimulation, inducing SSVEP responses in the brain. At the current point in time, the platform focuses on stimulation patterns that induce SSVEPs. However, there are other methods for visual stimulation like the P300 ERP response, which might be interesting for ophthalmological experiments in the future. Those are currently no part of the Metis Videre software.

4.2.5 Experiment Protocols

For a smooth and easy reproducibility of an experiment, protocols are a great idea to implement. In early iterations of the Metis Videre software stack, protocols were just CSV files where each entry relates to a specific value like frequency or stimulus type. Yet, old CSV files need to be converted to the new format whenever a new parameter was added. This was not very practical, especially with an ever-growing and evolving platform.

We can harness the power of a hierarchical structure like XML or JSON. Compared to CSV, the major advantage of those data formats are that they are easily human-readable. Furthermore, as already mentioned, they are hierarchical;
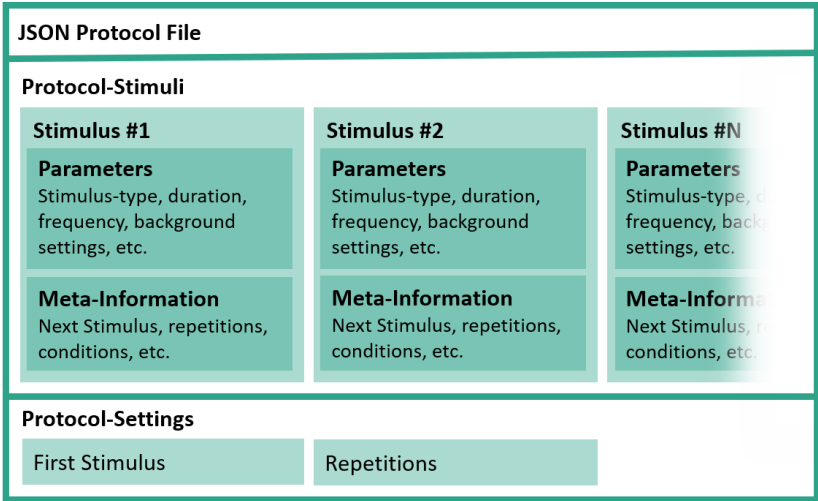
Figure 4.15: Graphical representation of the JSON protocol files used in Metis Videre. The general structure is divided into two main sections, the protocol-stimuli information and the protocol-settings information. Thanks to detailed meta-information, it is possible to create non-linear procedures of stimuli using this file format.

therefore, more complex structures can be defined and easily parseable. In the specific case of Metis Videre, the JSON file format got used because its annotation is shorter than XML, making it quicker to read and write, and it allows the use of arrays. Figure 4.15 shows a graphical representation including the hierarchy of a JSON protocol file as implemented in Metis Videre.

4.3 Hardware of 'Metis Videre'

Metis Videre, as a platform project, includes a high amount of hardware and setup implementations. This section about hardware will cover the experimental setup, including the design of a simple chin-rest with an integrated phoropter. Furthermore, we will explore how gamma correction works on a display, and finally, a novel type of 3D-printed electrode, the Circuittrode, will be introduced.
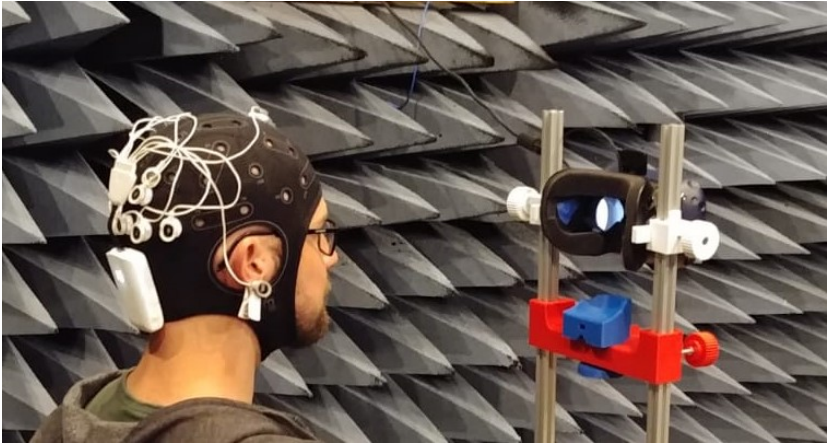
Figure 4.16: The author is sitting in front of a VR device which is mounted on a chin-rest. He is wearing the Neuroelectrics Enobio 8 EEG. In the background the antenna measurement chamber is seen. However, there were no improvements in signal quality in comparison to a normal lab setup.

4.3.1 Experimental Setup

First, we will look at the general setup of an experiment in the Metis Videre platform. Figure 4.16 shows the setup, including the EEG device from Neuroelectrics and a chin-rest with mounted HTC Vive Pro VR HMD. The photo was taken in an antenna measurement chamber to reduce external electrical interferences on the EEG signal. However, there was no significant effect on signal quality. The chin-rest used to holster the VR HMD is very versatile and was developed for the Metis Videre platform. Its most important feature is that it allows the subject to put their head on a support structure. Due to the decreased mobility, there are inherently fewer movement artifacts. Furthermore, less concentration and muscle tension are needed for the subject to keep his position stable. The complete chin-rest setup was modeled in a Computer-Aided Design (CAD) tool and was 3D printed on a fuse deposition molding (FDM) printer. The design of the chin-rest is presented in Figure 4.17 and was published at SPIE Neurophotonics [5]. The chin-rest platform further allows the fixation of the forehead and to mount a simple phoropter. Using this phoropter, different lenses or occlusions can be mounted in front of the eyes.
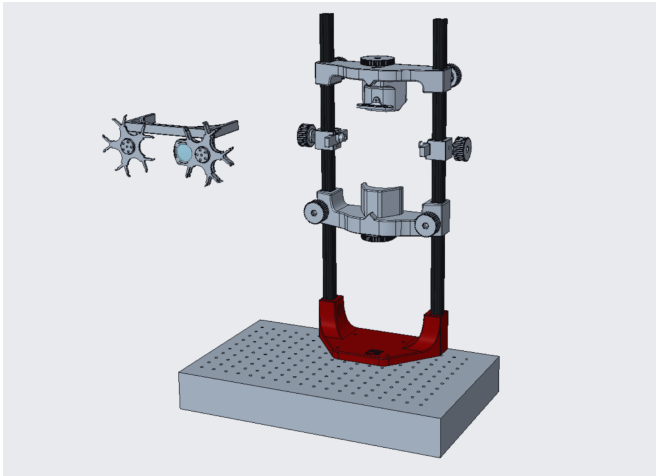
Figure 4.17: The rendered CAD model of the chin-rest modules in Metis Videre. It is composed of a standing foot, which can be mounted to a optical table, a forehead rest, suspensions for a VR-device and a simple phoropter. We open sourced the complete design which is available at [5].

Source: *Already published by the Author in [5]*

Most of the time, the system was not set up in the antenna lab like seen in Figure 4.16 but rather at a laboratory setting at the Institute for Information Processing Technologies (ITIV). The ITIV laboratory setup is presented in the Figures 4.18 and 4.19. Thanks to mounted occlusion items, monocular measurements can be performed with both eyes open. In addition, the lenses in the phoropter can be used to increase or decrease the distance of the monitor virtually. Notice that on the same chin-rest, the VR headset can also be mounted.

4.3.2 Gamma Correction

We as humans experience luminance, not in a linear but an exponential manner. Displays often do not attribute to human perception. If a white (RGB (255,255,255)) and a black (RGB (0,0,0)) pixel are optically mixed it is not the same perception as a grey (127,127,127) pixel. To achieve this, we need to *gamma correct* the screen with the help of a gamma correction value γ.

Figure 4.18: Side view on the Metis Videre laboratory setup. On the left, a high-frequency monitor is installed. On the right, the modular chin-rest is presented. The subject can sit on a chair which is provided. For reduced perceptual distraction, the rest of the optical table is covered in white linen.
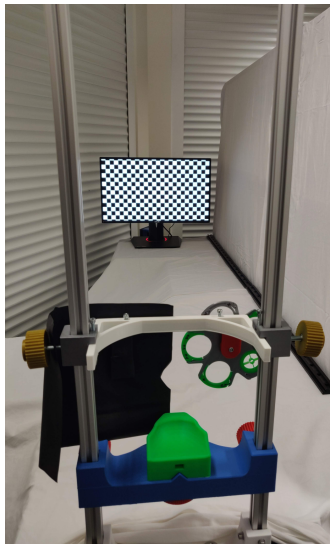

Figure 4.19: The point of view of the subject in a Metis Videre experiment. They can put their chin on the chin-rest and watch the stimuli on the monitor. In this photo, the setup measures visual perception properties in the right eye (OD). The lenses are used to modulate the virtual distance of the screen.

109

Figure 4.20: The screenshot shows the gamma calibration screen, which was implemented in unity. Half of the screen displays a grey value, the other half small black and white checkers. The screen is calibrated for luminance γ when the subject cannot perceive the difference between grey and checkerboard from the point of measurement.

When displaying a grey area on the left and a fine checker of black and white pixels on the right, the monitor is defined as calibrated if a subject cannot perceive the border between the two representations. A tool was implemented into the Metis Videre software framework to perform the screen calibration. The calibration-screen is displayed in Figure 4.20. Within the screen, the checker size and the luminance value of the grey-value are modifiable. For the calibration, the subject is positioned at the desired measurement location (compare Figure 4.19). Then the correct γ value is calibrated so that the border in the center is not perceivable anymore. For some screens, it is necessary to tilt the monitor. This is due to manufacturing differences in the background illumination of the screen, especially in TN (twisted nematic) displays, and less so in IPS (in-plane switching) and OLED (organic light emitting diode) displays. When transforming the RGB color space to HSV representation, the representation for the grey value is:

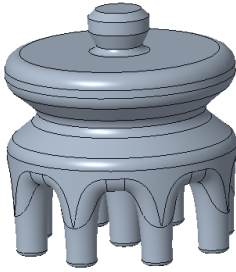$$(H, S, V) = (0, 0, v), \quad v \in [0, 1]. \tag{4.11}$$

Figure 4.21: An early design of a dry EEG electrode. The electrode has legs to penetrate through hair. After printing it could have been coated with a silver-layer for conducting properties.
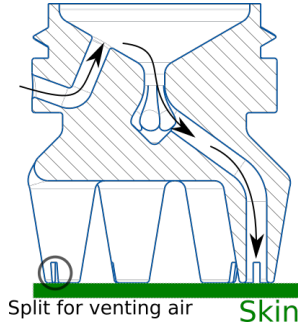

Split for venting air   Skin

Figure 4.22: By introducing a canal system, electrolyte can be transported directly to the scalp, minimizing exposure.

We want to achieve a V value so that

$$V_{out} = V_{in}^{\gamma} \qquad (4.12)$$

with V_{out} being the gamma corrected pixel value in the HSV representation. The gamma value can be then calculated as

$$\gamma = ln(V_{in} - V_{out}). \qquad (4.13)$$

For some perceptual experiments, it is of utmost importance to correct the screen's luminance, especially for sweep stimulation patterns with variant spatial frequency. Such stimulation is essential when researching visual acuity and perceptional resolutions. Furthermore, a calibrated screen improves the reproducibility of BCI experiments.

4.3.3 EEG-Electrodes - 'Circuittrodes'

In Section 3.5.3 a concept for the design of novel EEG electrodes was described. This section will explain the developed electrode in-depth, look at the individual development steps and the electrodes' features.
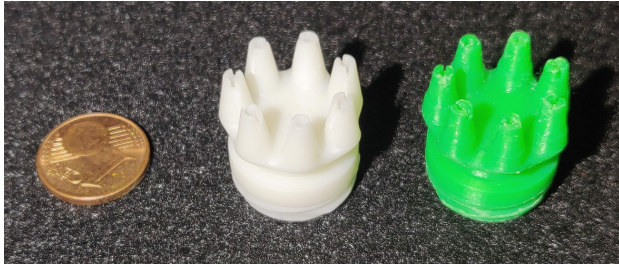
Figure 4.23: The figure shows a comparison between a FDM printed electrode (right) and a SLA printed electrode (center). A euro-cent piece is presented on the left for size-reference. The SLA print has a higher quality and resolution, especially the internal canal system is completly intact.
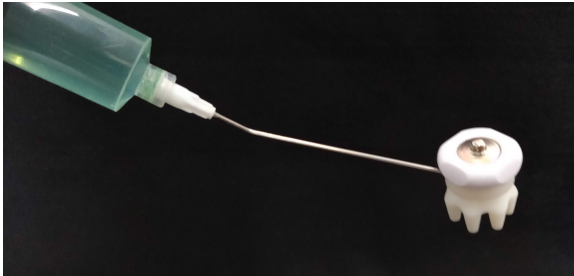

Figure 4.24: The assembled electrode with attached electrode cap being filled with electrolyte using a syringe. The electrode cap has a connector for attaching wires from the EEG device.

It is important to notice, that figures and descriptions and figures of the electrodes might have already appeared in [3] and [39].

A concept CAD-rendering of a dry electrode was created and is presented in Figure 4.21. It was designed to test the printing properties of the 3D printer. Initially, the idea was to coat the 3D printed electrode with a silver coating. However, as the production of such a coating is challenging and even dry electrodes available on the market did not perform well, approaches towards a dry electrode were stopped.

Wet electrodes are great; however, the typical state-of-the-art cylindrical wet electrodes expose the subject to a lot of electrolyte fluid. Therefore, a new

design for wet electrodes is proposed in Figure 4.22. This new design introduces canals through the legs of the electrode. For the manufacturing of the electrode, 3D printing was chosen. However, FDM printing technologies were only feasible to some degree, as it is hard to resolve the delicate structures inside the electrode. Another 3d printing technology, called Stereolithography (SLA), uses UV light exposure to extract 3D designs from a resin polymer. Figure 4.23 shows the direct comparison of the two printing technologies on the test case of the electrode design presented in 4.22. Prusa Research a.s manufactured both used printers. Namely, the FDM printer *Original Prusa i3 MK3* and the SLA printer *Original Prusa SL1*. We will not go into much detail about the backgrounds of said technologies. The most important thing to know is that the FDM printer usually print from a 0.4 mm nozzle and a layer height of up to 0.05 mm. Due to the viscosity of the melted polymer, the resolution is limited. The SLA printer, however, reaches a layer resolution of up to 0.01 mm, signifying a 40-times superior layer resolution than the FDM printer. Furthermore, due to the different technological approach, a maximum resolution in X-Y dimension of up to 0.047 mm is achieveable in the SLA printer[3].

Thus, the SLA electrode (center, Figure 4.23) has a better resolution in comparison to the FDM printed counterpart (right, Figure 4.23). Furthermore, the medical-grade resin can be used in the SLA printer to ensure biocompatibility. Said resin is also being used, for example, in dental applications [FYR+18].

By attaching an electrode cap with an integrated transformation-electrode made from Ag-AgCl from Neuroelectrics, we achieve a closed electrode design like shown in Figure 4.24. Thanks to the opening on the side, it can be manually filled and prepared with a syringe.

This kind of design still has a major flaw. The electrolyte can be transported to the skin. However, it can hardly be controlled how much electrolyte is leaking onto the head when the electrode is filled while installed in an electrode cap. So far, there have been ideas to implement a return canal from which electrolyte can be pumped back to a reservoir [Mea19]. The novel approach of this work is to use the head itself as a seal for the creation of a closed-loop canal

---

[3] Specifications were extracted from the technical documentation of the 3D printers. Further information can be obtained at: https://www.prusa3d.com/.
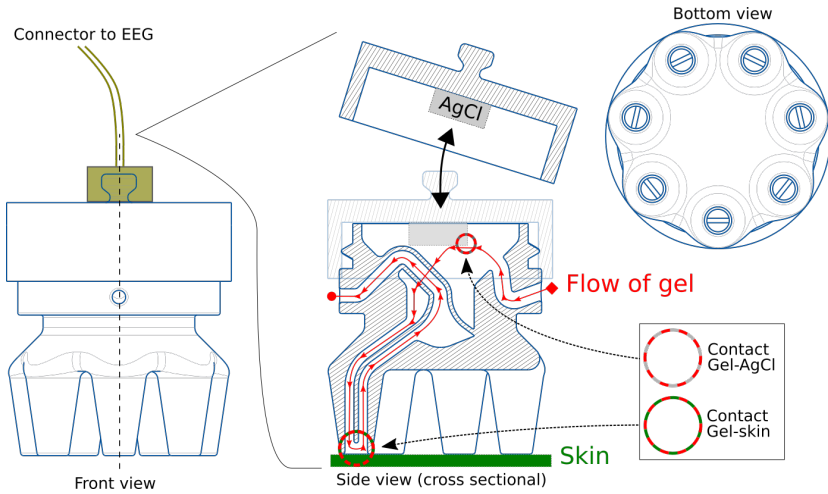
Figure 4.25: The figure shows a technical drawing from different perspectives and a cross section of the circuittrode. The skin of the scalp is used as a seal to close the canal. By using a divert module, the electrolyte is transported up the electrode again. The electrolyte is introduced using a syringe and capillary forces help with the distribution of electrolyte. Inside the electrode cap, the transformation-electrode made from silver chloride (AgCl) is located.

Figure 4.26: Top view of the circuittrode. A 1-Euro coin is displayed for scale. The entry-point for the syringe is on the left side of the electrode and the outlet for excess electrolyte is on the right.

Figure 4.27: Side and bottom view of the circuittrode. The bottom view nicely displays the canals and cavities.

system. Using a closed system, the electrolyte cannot easily exit the circuit if the contact pressure of the electrode is high enough. The ionic charge carrier circuit is still closed via the electrolyte in the electrode. Furthermore, the newly added riser canal adds an indicator when the electrode is perfectly filled, as some electrolyte will be exiting through the other side of the electrode. The final design of the electrode is presented in Figure 4.25. The design of this novel electrode has been filed for patent [3]. The Figures 4.26 and 4.27 provide further perspectives on the finished SLA printed electrode. The exiting and rising canals are seen from below-view, while the entry and exit canals are seen from an above-view. A twist-lock is printed in place at the top of the electrode to allow for an easy attachment of an electrode cap with a transformation-electrode.

The final setup using a VR device, the headrest, and the 3D printed electrodes is displayed in Figure 4.28.
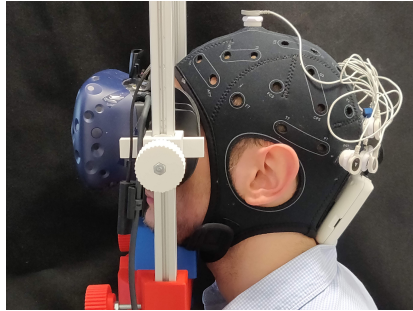
Figure 4.28: The photograph shows the hardware setup of the Metis Videre hardware platform. The circuittrodes are mounted in a neoprene cap and are connected to a Neuroelectrics Enobio 8 EEG device. The subject is viewing stimuli in a HTC Vive Pro VR HMD, whilst resting their head on the chin-rest.

4.4 Data Science of 'Metis Videre'

This last implementation section is about data science. Data science is the umbrella term for statistical methods and machine learning approaches. For machine learning approaches, the QUACK process, which was introduced in Section 3.3 is a good starting point. As we already covered machine learning concepts, now we will learn how it is used in the Metis Videre platform.

This section will discuss how to implement AI structures to perform perimetry using the EEG brain-computer interfaces objectively. For that, stimuli at different positions in the visual field were displayed, and data was recorded using the hardware and software setup of Metis Videre.

4.4.1 Preprocessing

It is quite common practice when working with machine learning approaches to standardize the data [GBC16]. A standardized signal is a signal whose expected value is 0 ($E(X) = \mu = 0$) and whose variance is 1 ($Var(X) = \sigma^2 = 1$). Mathematically this standardization can be expressed as a Z-transformation as

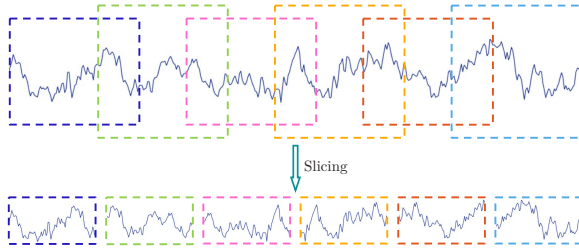$$Z = \frac{X - \mu}{\sigma}. \tag{4.14}$$

Figure 4.29: For augment EEG data, slicing can be used to get from one stream of EEG data multiple training slices.

In comparison with image recognition, the major downside of pattern recognition in EEG signals is that there is not much data to be found online. Especially not with respect to a particular problem, like stimulation with a specific frequency at a particular position, which is not the fovea. Therefore, labeled data is sparse. Thus data augmentation is of utmost importance to increase the number of training data. One of the easiest and most efficient ways to augment data is a procedure called *slicing*. As the name suggests, it is the procedure of cutting a single measurement into multiple slices, compare figure 4.29. Something unique about SSVEP EEG recordings is that the signal should be present equally distributed over the whole stimulation period. Therefore, slices may interlap to allow for even more training data. This form of augmentation is also a regularization method as it further forces the network to recognize inherent signal properties instead of local variations like artifacts [SSF+17].

4.4.2 Architecture

When working algorithmically with EEG signals, the aim is often to discover structures and features in the data. For example, in an EEG-data frame, just like in an image, values close to each other have a higher correlation than points far away from each other. Therefore, CNNs are used preferably to MLP neural network architectures. The creation of a well-performing network architecture is a long and tedious process. Especially the number of layers and the dimensions of their kernels are crucial.

EEG signals are more comparable to videos than images as their alteration over time is an essential feature. Therefore, a CNN for the analysis of EEG signals should not have standard 2D convolutions, as those only allow the convolution over the temporal or over the spatial dimension. A 3D convolution allows for flexible use in spatial as well as in temporal dimension or even in both simultaneously. When remembering Figure 3.15, temporal dimension is usually larger for EEG recordings than spatial dimensions. The Enobio 8 EEG device only has up to 8 channels. In layman's terms, one could say that such an EEG signal is comparable to a video signal consisting of only up to 8 pixels.

The job of the CNN is the same as the job of the CCA. It has to transform the measured values from the time domain to the frequency domain and allow for the best fit throughout different channels for a selected frequency. We can assume two scientific hypotheses now: The first one is that CNNs can reproduce the results of complicated statistical functions like the CCA algorithm. The second is that if the frequencies of the SSVEP and its harmonics are the only features of relevance for detecting SSVEPs, the CNN approach will not be better but equal to the CCA approach.

The complete network structure, which is explained in the following, is provided in Figure 4.30. The initial step is to transform the data to the frequency domain. To get a frequency domain representation, we need to convolve the signal with a reference signal. The dimensions width W, height H and depth D are chosen to be

$$(W, H, D) = (1, 1, 10) \tag{4.15}$$

for the initial layer. This means that all channels are being convolved with a ten units long kernel. Assuming a sampling frequency of 500 Hz, that corresponds to a kernel size of 0,02 seconds. Four of those kernels will be used at a stride of one in every direction. After that, a convolution of the spatial dimensions is performed. For this, the kernel is chosen to be

$$(W, H, D) = (2, 5, 1) \tag{4.16}$$

therefore, the filter size is equal to the input size, reducing the output feature vector to only having depth and no spatial information. A max-pooling layer is performing a reduction of dimensionality of the factor of two in the next layer. For regularization purposes batch-normalization and dropout [GBC16, GZM20] have been chosen. The amount of dropout was estimated by hyperparame-

ter optimization, which is explained in the next section (Section 4.4.3). The chosen activation function is ELU. The ELU function is defined as

$$ELU = \begin{cases} x & x \geq 0 \\ \alpha\left(e^x - 1\right) & x < 0. \end{cases} \tag{4.17}$$

Accordingly, the ELU function is a mapping of the identity for values greater than zero, and values less than zero are considered in contrast to the ReLU, but not as strongly.

The described procedure is performed three times with an increasing number of filters. The more often the signal is convolved, the larger the receptive field of the neural structure is until it has enough information extracted from the signal to perform the classification task. Said classification task is implemented in the form of a fully connected network analog to an MLP. The flattened feature vector is fully connected with 256 dense neural nodes, which are again connected to 128 dense nodes until finally a binary pathology classification is performed.

Another interesting thing to notice is the second input and output of the network structure. As the closed-loop system always knows where a stimulus is presented, this information can be further considered in the fully connected structure. If there is a diverging influence of a stimulus to different areas of the retina, this input will help distinguish it. The position data is fed in as a vector tuple of distance to the center and the quadrant where the stimulus was presented. The second output is a CCA regression. It is not used for the actual classification of the signal, as performed in the classification output. However, it is used to push the network towards the right direction for convergence. Thus, the regression output is used as a form of regularization. Own work like the experiment performed in DeepBees (compare Section 3.4) and other works [GBC16] also show that multi-task learning is a powerful tool to regulate neural networks. The CCA value can be easily calculated for the training snipped, and be used to train the multi-task model.
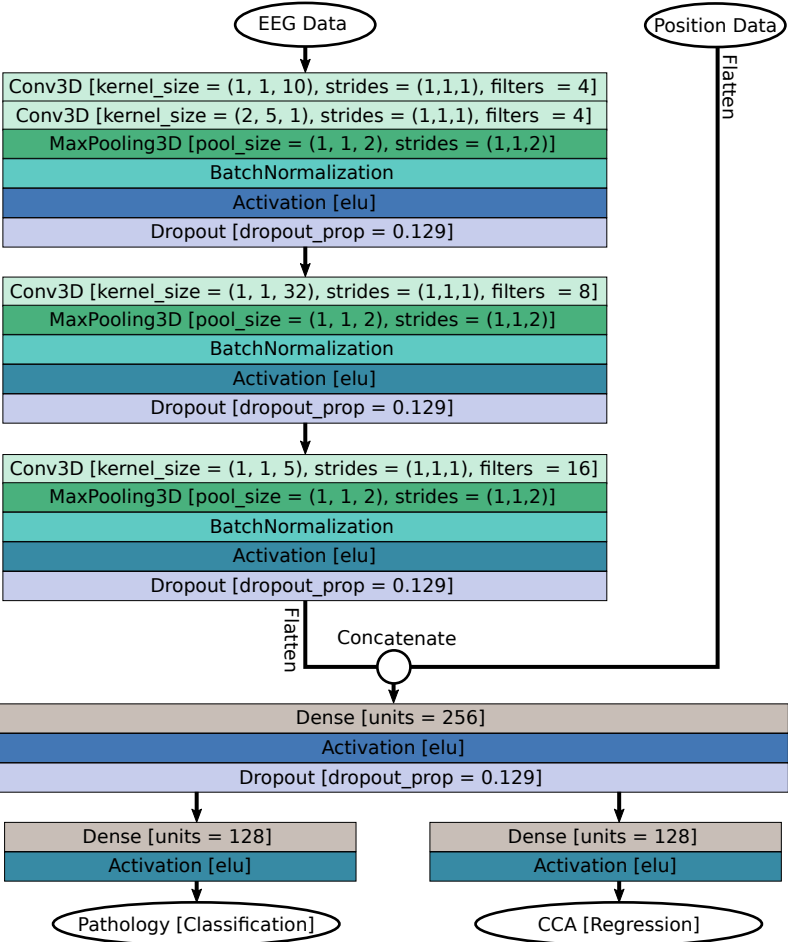
Figure 4.30: The final implemented neural structure used in the objective perimetry experiments with the Metis Videre platform is shown. It uses 3D-convolution layers to enable convolutions in time and spatial domains. Furthermore, multi-task learning is used as a form of regularization.

Source: *Previously published by the author in [5].*

4.4.3 Hyperparameter Optimization

In neural networks, various parameters can be set. Those that are not directly weights and biases of the network itself are called hyperparameters. Such hyperparameters are, for example, the amount of dropout during training, the number of neurons per layer, the learning rate, and regularizers like L1- and L2-regularization. In general, well-chosen hyperparameters help improve the performance of the network and, at the same time, generalize it further, which protects against overfitting. There are several ways to choose these complex parameters. Mostly they are set by trial and error or by expert knowledge. However, this expert knowledge is a subjective feeling and may be far from the optimum. Therefore, these hyperparameters should be optimized by machine learning procedures as well. Three methods are of particular interest here: *Grid Search, Random Search, and Gaussian Hyperparameter Optimization.* Grid search is probably the most careful procedure since it tries all combinatorial possibilities. Because of this, the runtime of the method is also the longest and thus often cannot be executed due to exploding search space. A remedy for this is the Random Search algorithm, which does not proceed systematically but randomly to find a good set of hyperparameters. Accordingly, one can be lucky or unlucky with this approach. A systematic approach, which is more spread out and only slightly dependent on luck, is the Gaussian Hyperparameter Optimization algorithm. Here, after testing a set of randomly chosen values, an underlying normal distribution is assumed. This normal distribution has a theoretical maximum, which is used as the value for the next test, the value after training is determined, and the process is iterated. Hyperparameters are highly co-dependent, and the dependencies and correlations are usually not directly evident. A visualization of a hyperparameter optimization process can be seen in Figure 4.31.

In procedures like this, the split of training, validation, and test-set is crucial. We split training and validation sets in order to prevent overfitting to the training set. Suppose the optimization process would only be done for the training set, and hyperparameters would be optimized towards the said set. In that case, the resulting network will perform tremendously well on the set. However, it probably will underperform on the test set and any other data. This is because the test set aims to mimic never-seen data. Therefore, the test set is only touched *once* after the whole optimization process is finished. One might think that the optimization process increases the chances of overfitting not towards the
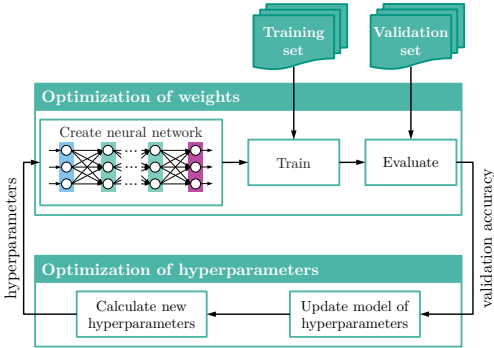
Figure 4.31: The Figure shows how hyperparameters are optimized in a closed loop approach. The training set is used for training, the validation set is used to evaluate the model. According to that evaluation the model of hyperparameter is updated, generating new hyperparameters, ultimately creating a new neural network.

Source: *Previously published by the author in [5].*

training but the validation set. This is true to some extend, however thanks to the split of the sets and the actual training being disjunct from validation, we obtain a better-generalized machine learning model after training.

4.4.4 Software implementation

The implementation of the presented network was performed in Keras. Keras is a high-level library to interface with different deep learning libraries. In this case, the underlying library is Google's Tensor Flow. All implementations were initially performed in Python and then translated to C++ by exporting the H5-file containing weights and biases network structure. The C++ version of the neural network enables high-performance real-time inferences.

The networks were trained at the Institute for Information Processing Technologies (ITIV) on up to 20 computers at a time. The computers were equipped with an NVIDIA RTX 2080 to perform the training process on the GPU. Jobs were manually distributed and collected, and only the best network (on the validation set) was evaluated on the test set.

5 Studies and Evaluation

> All we have to decide is what to do
> with the time that is given us.
>
> *Gandalf (J.R.R. Tolkien)*

In the previous chapters, concepts of visual perception as well as a broad research platform were introduced. This chapter will focus on the conducted studies and their evaluation. Therefore, it will also contain the results of said studies. The studies are designed to inherently expand the horizon of knowledge of humanity regarding visual perception. The Fundamentals, Concepts, and Implementation are only means to make sense of the data collected from these studies. It is important to note that the study does not necessarily have to be about visual perception in terms of what a human subjectivity sees and how we can interpret vision objectively. The study about the circuittrodes is a study that pushes the understanding of EEG technology and, therefore, automatically our ability to perceive how vision is processed in the brain objectively. These studies are not intended to enable clinical decision-making or provide statistical validation for the general population. However, they provide insights into the brain's perception. This chapter will cover four different studies, all involving visual perception.

The first one will focus on Virtual Reality and the simulation of vision. To be precise, the simulation of diseased vision with a focus on visual field loss. The study was conducted in a clinical setting at the University Clinic of Heidelberg at the David J. Apple Laboratory. A total of 22 participants were immersed in the VR vision experience and asked a questionary. The results of this clinical case study are presented in Section 5.1. The second study was a small laboratory study at the EEG lab at the Institute of Information Processing Technologies at KIT. It investigates the effects of periodical stimulation patterns on the human visual cortex. In a laboratory setting, steady-state visually evoked

potentials were induced in a subject by providing stimuli on a high-frequency refresh rate monitor. The occipital lobe, where the primary visual cortex is located, was observed using electroencephalography. The results allow insights into visual information processing within the brain. The first and second study results are combined to perform a third study on visual field loss. Visual stimuli were presented throughout the subject's visual field, and the pathology of the stimulation positions was evaluated. The study was conducted in an international laboratory setting at KIT, Karlsruhe, Germany, and MIT, Cambridge, MA, USA. A final study was conducted on the newly developed electrodes. The study was designed to compare the new electrodes with the state-of-the-art electrodes and work out their advantages. Those electrodes might provide a better starting point for perceptional and brain researchers in the future.

5.1 Clinical Case Study: Virtual Reality for Vision Science

The system used for the experiment was introduced in Concept Section 3.1 and Implementation Section 4.1. Furthermore, details of implementation and the study have been published at the Virtual Reality Software and Technology Conference [10], and at the Association for Research in Vision and Ophthalmology annual meeting [9].

This initial study aims to show that it is possible to use perimeter data and Virtual Reality (VR) to simulate visual perception in a fully immersive environment. The focus of the study was on the simulation of individual glaucomatous visual field loss. Individuals were presented with a realistic simulation of a glaucoma experience in virtual reality. For evaluation, they were presented with a questionnaire.

5.1.1 Experiment Design

The clinical evaluation study was designed and conducted closely with physicians and researchers from the David J. Apple Laboratory at the Heidelberg University Hospital. The David J. Apple Laboratory is a renowned research

lab within the ophthalmology department and is led by Prof. Dr. med. Gerd U. Auffarth. Three groups of participants were recruited:

- Healthy individuals (N = 6) with no scientific background in perceptional research or ophthalmology. In the following, this group will be referred to as *normals*.

- Healthcare *professionals* (N = 11) who have background in perceptional research or ophthalmology. These persons had an academic education in vision science or practiced at the Heidelberg University eye clinic.

- Patients (N = 5) who have glaucoma themselves.

It is important to note that the study participants were not affiliated with the Author or the Author's institution. The patients who participated in the study were observed for glaucoma progression over several years. During this study, we used the software presented in this dissertation to import their perimeter data into the VR software and present them with their own data. With a click of a button, the patients could view the progression of their individual glaucoma situation. The visual field test was performed using an Octopus 900 device and standard glaucoma (G) pattern. Measured sensitivity thresholds were converted and extrapolated to match the Oculus Rift VR field of view and presented to the patient. Due to data protection, generic progressional glaucomatous vision field loss was simulated and presented for healthcare professionals and the normal group. Furthermore, the participants were able to choose from monocular field loss and binocular field loss and were further able to freely move around in the immersive virtual environment and interact with the environment. The environment included a living room, kitchen, bathroom, an Amsler grid, and a standard vision Landolt ring chart on the wall. The simulation's starting point with activated binocular visual field loss is presented in Figure 5.1. The top images show an early state of binocular glaucoma, and the bottom images illustrate the progressed glaucomatous experience. Notice that black was used to occlude the areas which were affected by glaucoma. This is because black is the abstinence of information. Interestingly, when presented with a fixation target, participants tended to suppress black areas' perception, and only rendered areas were perceived after some time of acclimation. To better comprehend individual subjective experiences, a questionnaire was designed to map the participants' reports to a quantitative scale. The next
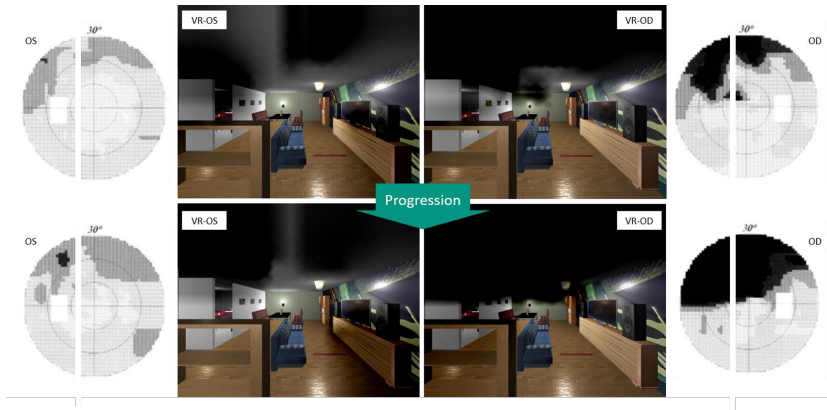
Figure 5.1: The figure shows screenshots from the vision simulator developed for this dissertation. The top row shows on the left the perimetry scan of a left eye. The center-left view is the respective VR experience. The right side is the same for the right eye. The two images in the center form a stereoscopic 3D experience when viewed in a VR HMD. Within the software, a button can be pressed to simulate the disease's progression, which displays the next level of severity (bottom row).

Source: *Previously published by the Author in [9].*

subsections will discuss the participants' experiences and what we might learn from the experiment.

5.1.2 Results

The questions of the questionnaire were the following:

- **I:** The simulation is suitable for the visualization of the course of (my) vision disease.

- **II:** I found the disease effects in the simulation to be realistic.

- **III:** I found the course of the different progression stages to be realistic.

- **IV:** I see a benefit in this form of visualization.

Answer in survey	Numerical representation
strongly disagree	-2
disagree	-1
no answer	0
agree	1
strongly agree	2

Table 5.1: Conversion of strings to numerical ratings.

- **V:** For a better understanding of the condition, the system should be recommended to relatives.

- **VI:** The use of simulation was helpful in assessing the effects of (my) visual impairment.

Participants had to choose one of 4 possible answers or chose not to answer a question. The quantitative representatives of the answers are denoted in Table 5.1.

The final results are presented in Figure 5.2 as a bar graph including a confidence value of one sigma. The smaller the confidence bar, the more reliable the result over all participants. Most agreed that this VR technique is better understandable for a layperson than grayscale data representation created by a perimeter. Participants reported that the VR platform is well suited for the simulation of vision diseases. The glaucoma group agreed that such data representation was helping them to evaluate their individual vision progression. Furthermore, they reported that the simulation can be considered realistic; however not fit their comprehension of real perception.

5.1.3 Conclusion

According to our study involving N = 22 participants, we conclude that VR is suited for understandably visualizing the progression of individual glaucoma cases. This study shows that it is possible to display reduced visual field performance and progression due to glaucoma in a realistic manner in VR.
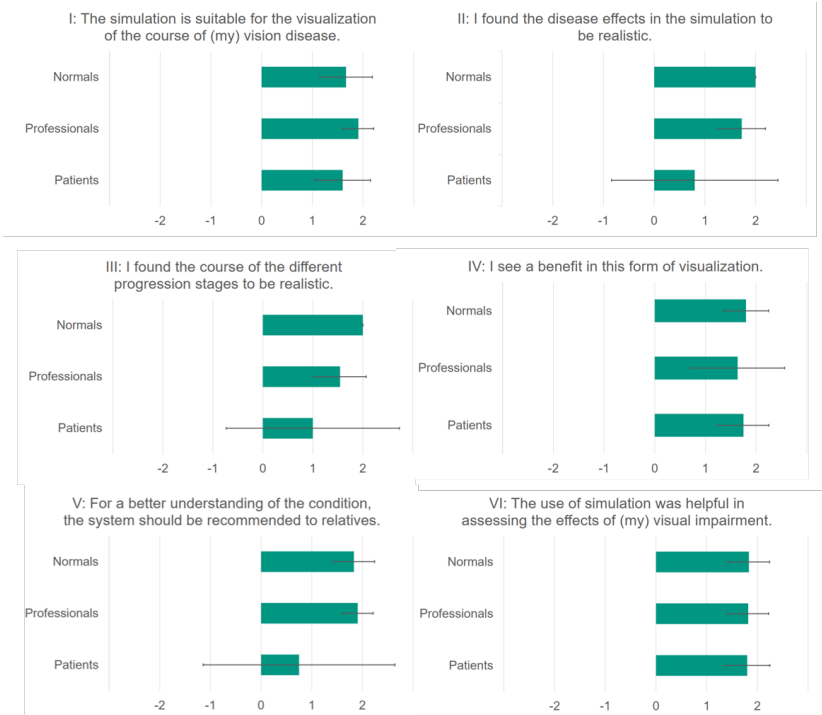
Figure 5.2: The figure shows all questions of the questionnaire and their respective answer (*strongly disagree (-2)* to *strongly agree (2)*). The three participant groups: *normals, professionals*, and *patients*, are displayed separately for each question. It can be observed that the opinions of normals and professionals match almost always, while patients' perceptions differ.

Source: *Previously published by the Author in [9].*

Furthermore, we learned that the brain automatically erases black areas from the active field of view if those black areas stay constant. This is very similar to the optical nerve head's perception in the visual field, where no perception is happening. To recreate real perception with vision diseases, pupil tracking could be used to automatically update effects according to the user's gaze, enabling a more realistic immersion. Generally, platforms for the simulation of vision could be used to educate professionals and relatives and provide a better understandable data representation of the status of vision loss in patients compared with grayscale or numeric perimetry maps. Some patients reported that they would not want their individual data to be seen this way by relatives. We do not have precise data on why this is the case because the same individuals saw a benefit in the novel visualization of the data. However, we can assume that individuals who are not generally accepting their individual state of vision might have an even harder time sharing such an immersive and, therefore, intrusive way to their bio-medical data. Hence, the patient should always be consulted before showing medical simulations like this to relatives who might be present at the hospital. On the other hand, the representation of the data in an immersive VR environment could help physicians and even content creators to better understand visually impaired people's individual needs.

5.1.4 Recent Developments

After the publication of our work, other researchers further improved and developed perceptional vision experiences in VR and cited the presented work. Dittrich et al. [DRT19] researched vision and aging using virtual reality. They were particularly interested in Presbyopia, Cataract, AMD, and Glaucoma. In 2020 Jones et al. released the openVision Sim [JSCWBC20], a particular interesting open source project, to simulate visual field loss also including glaucoma. They implemented 'real-world tasks of daily living' into their application, enabling further research regarding perception with vision diseases. The vision simulation system presented was also used for the international laboratory study (Section 5.3) regarding objective glaucoma detection using EEG devices.

5.2 A study in Human Vision on Steady-State Visually Evoked Potentials

Now that we learned and simulated diseased vision and the subjective perception of the visual field, we will focus on a more objective way to view visual perception. In this study, we will examine the visual cortex with the help of the Metis Videre platform utilizing an EEG BCI. Its primary objective is to better understand stimulation frequency and pattern-sizes for the subsequent international laboratory study regarding visual fields (Section 5.3).

All responses which will be discussed in this section are based on SSVEPs (compare Section 2.3.3) and their magnitude is determined using the CCA algorithm (compare Section 2.5.2).

5.2.1 Experiment Design

The most prominent features of an (SSVEP) stimulus are its size and frequency. Other factors like pattern onset, reversal or even a moving stimulus also play a role. However, this study will focus on size and frequency. The experiment was performed at the EEG laboratory of the Institute for Information Processing Technologies at KIT. In total, 10 participants in the age range between 20 and 30 were recruited to participate in the study. All participants were screened for photosensitive epilepsy beforehand in a questionnaire to reduce the experiment's risks as some people react sensitively to certain frequencies. 30 different frequencies were evaluated for each of the 10 participants. The experiment protocol iterated through all frequencies specified from lowest to highest. A checkerboard stimulus at a size of ten degrees was displayed in a VR device for each frequency. Each stimulus was displayed for 7 seconds with a 3 seconds break between stimulations. The available pattern onset frequencies can be calculated on a 90 Hz VR display as follows:

$$f_{Stim} = \frac{90\,\text{Hz}}{N}, N \in \mathbb{N} > 1 \tag{5.1}$$

The highest frequency which is possible to be displayed according to Nyquist is, therefore, 45.0 Hz. For the lowest frequency, there is no lower bound. However, the brain is at some point unable to produce a steady response to a

signal. Therefore, this study will start at 3 Hz and test all possible frequencies up to 45 Hz. Furthermore, it is noted that the higher the frequency of a stimulus, the more pleasant it is to look at, especially high frequencies above the flicker-fusion rate are hard to distinguish from a static image [AUS18].

For stimulus size, the limiting factors are the resolution of the VR device and its maximum field of view, which is around 90 degrees for a HTC Vive VR device. We can expect that the brain's response to a small stimulus is also diminutive and increases with stimulus size up to a certain amount.

Finally, we need to perform an analysis to compare the results obtained from the experiment with one another. For that, the CCA is calculated with a set of matching reference frequencies. Then the mean value of all CCA results (window size of two seconds and a shift of 60 ms (30 samples at a sampling rate of 500 Hz)) is calculated for each simulated window. Finally, the results are presented in a Boxplot including, mean, quartiles and outliers [HKW11].

5.2.2 Results and Conclusion

The results of the study are presented in the Figures 5.3 and 5.4. In the frequency analysis, we can see that the response of the visual cortex to lower frequencies in the range between 3 to 10 Hertz was rather strong. In contrast, for higher frequencies, the response is noticeably smaller. For very high frequencies above 40 Hz the stimulus is not easily perceivable anymore (compare with [AUS18]). Due to the small excitement of the signal, those high frequencies are very suitable for a medical application where we generally want a good signal-to-noise ratio. The explanation why the signal is so bad at those high frequencies is that those frequencies are at or even above the *flicker-fusion-rate*, meaning that stimuli are perceived as continuous images by conscious thinking. Also, we want to stay away from frequencies in the spectrum of the natural alpha range of the brain, which could evoke an epileptic seizure in people with epilepsy. Therefore the frequencies between 8 and 12 Hertz should be avoided. Thus, 7.5 Hz can be considered a safe but still high and pleasant frequency for photo-stimulation of the retina and the brain.

In Figure 5.4 we can see the comparison of the effect of increasing stimulus size on the strength of the SSVEP CCA. For small patterns, the response is small and increases with the increment of the stimulus size. It reaches saturation
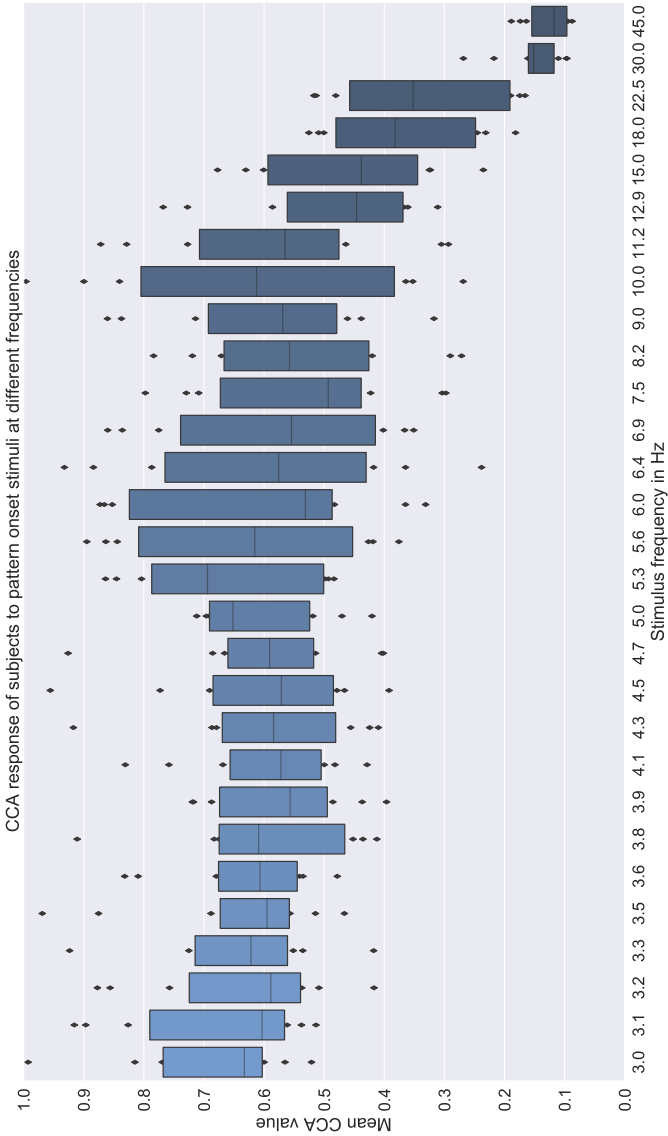
Figure 5.3: Comparison of the CCA analysed responses of the human visual cortex to different stimulation frequencies. The stimuli presented where performed in a pattern onset fashion with N=10 subjects. On the 90 Hz HTC Vive VR device, only certain frequencies were displayable. For a better visualization, the chart is displayed in landscape.
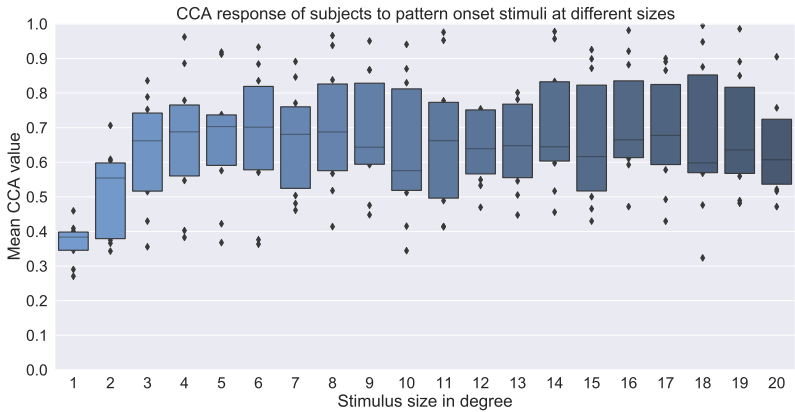
Figure 5.4: Comparison of the CCA analyzed the response of the human visual cortex to different stimulation sizes. The experiment was conducted in VR with a pattern onset stimulus at 7.5 Hz. The response rises with increasing stimulus size up to a size of around 5 degrees. After that, the increment of stimulus size has little to no effect on the CCA.

around 5 degrees of visual angle. After this, additional increments to the size of the stimulus do not result in a reliable increment of signal in the ten tested participants.

For further experiments, we have to consider that the stimulation in this experiment was performed with a central stimulus. However, when testing for the *what is where* - the perimetry of vision - it is crucial also to stimulate the periphery. What can we expect in the periphery? There are fewer photosensitive receptors than on the macula therefore a larger stimulus is needed in the periphery than on the macula.

Furthermore, while central visual perception is processed relatively close to the skull, peripheral vision is processed further within the brain. Thus, peripheral signals are harder to detect with an EEG. To take this into account, we should not orient ourselves at the lower border for signal saturation for central stimuli, which is at 5 degrees, but increase the size to a stimulus size of 10 degrees, which should be suitable for peripheral stimulation as well.

In summary, it can be concluded that for future ophthalmological experiments to investigate the *what is where*, one should use stimuli of 7.5 Hz at 10 degrees.

5.3 International Laboratory Case Study: Objective Perimetry using Brain Computer Interfaces

In this study, we combine the results of the studies presented in Sections 5.1 and 5.2 to produce a challenging new study regarding visual perception. This study aims at performing objective perimetry and utilize a CNN machine learning algorithm and compare it to the state-of-the-art CCA algorithm. As there is much faith in the use of CNN algorithms to analyze brain data, this study is an important contribution to ophthalmology, brain-computer interfaces, and deep learning alike. The results of this study were presented in SPIE Neurophotonics [5].

5.3.1 Experiment Design

The experiment was structure-oriented on the Metis Videre platform. Meaning all features of the platform were incorporated to perform the study. The software itself was shared from KIT to MIT, where a similar setup was built at an EEG laboratory. Headrests were used to ensure minimal head movement of subjects. At both sites, a Neuroelectrics Enobio 8 EEG device was used for the recordings. The experiment ran fully automated once started, thanks to the closed-loop experiment design using the Metis Videre software protocols. As mentioned beforehand, the study aimed at utilizing a BCI to perform perimetry, meaning testing of the performance of the visual field of a subject. The fundamentals of perimetry were introduced in Section 2.2.2 and the state-of-the-art in objective perimetry was analyzed in Section 2.2.4.

We can note here, that the stimuli used in this study were at 10 degrees in comparison with 4 degrees at [Göt10], who only used four stimulus positions at 5 degrees eccentricity. The Metis Videre setup, in contrast, uses a total of 13 positions:

- *Central vision* at the position (0°, 0°)

- *Close peripheral vision* at the positions (6°, 6°), (6°, -6°), (-6°, 6°) and (-6°, -6°)

- *Far peripheral vision* at the positions (6°, 16°), (16°, 6°), (-6°, 16°), (-16°, 6°), (-6°, -16°), (-16°, -6°), (6°, -16°) and (16°, -6°)

Based on the conclusion from the previous laboratory study 7.5 Hz and pattern onset checkerboards at a size ten degrees were chosen for the SSVEP stimuli. In addition, the relatively large grating size (one degree per checker) was chosen due to the reduced visual acuity in the periphery and to also compensate for the aberration effects of VR lenses.

The study was designed as a laboratory study at two major international technical universities. During this international study there was no access to diseased subjects. However, by using the vision simulation software introduced in Sections 3.1 and 4.1, which was evaluated in Section 5.1, we can perform masking on stimuli. This masking is inspired by the effects of glaucoma on a real patient. For each subject, five measurements were performed. One without any masking. One with advanced glaucoma in the right visual field. One with severe vision loss in the upper field. One with peripheral vision loss and the last one with central vision loss. The tested positions and the vision loss masks are presented in Figure 5.5. Furthermore, an exemplary stimulation is presented in Figure 5.6.

It was defined that if a stimulus is occluded by more than 50%, it is said to be pathological; otherwise, it was considered a healthy data point. Each stimulus was presented for a total of seven seconds. Ten subjects were recorded on-site at KIT, while five subjects were recorded at MIT. Subject age ranged from 22 to 33 years ($\mu = 26.5$, $\sigma = 3.2$) and 20 % of the subjects were female. Per individual $15 \times 5 \times 13 = 975$ seven-second snippets were collected. As the neural network only needs 2.5 seconds as input data, by using a step size of 536 ms, we were able to obtain nine samples per interval, ultimately adding to a total of 8775 samples. This is still a small dataset for a machine learning approach, as the available training data was further reduced due to the mandatory train-validation-test split.

The neural network in use was the CNN implemented in Metis Videre, which is shown in Figure 4.30. The network was optimized using *scikit-optimize* to
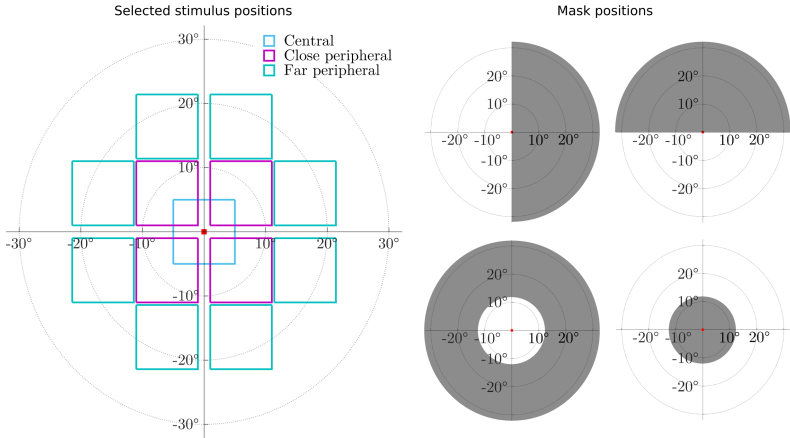
Figure 5.5: The left graph displays a human visual field. 13 positions were selected for individual testing of visual performance. The experiment examined central vision and peripheral vision, covering up to 41° of the visual field. On the right, mask positions are displayed. Those were used to augment visual impairment and reduced visual performance in healthy subjects.

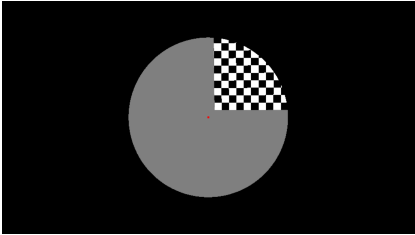**Source:** *Previously published by the Author in [5]*

Figure 5.6: A screenshot from the VR environment with a visualization of a visual mask and a visual stimulation checkerboard pattern (near-periphery). In the specific case, more than 50% of the stimulus is presented, therefore it is assumed to be non-pathological.

find a good set of hyperparameters by utilizing the bayesian hyperparameter optimization. However, an independent test set was identified and separated so that the results described in the next section allow for a transferable and reliable interpretation of results. Additionally, we performed a CCA-threshold analysis on the data (the CCA algorithm was introduced in Section 2.5.2). For the estimation of the threshold level, the same subjects used as training and validation data for the CNN approach were utilized. Furthermore, for each position in the visual field, an individual threshold was estimated, where only one Neural Network was used. Suppose the machine learning method can understand the underlying principles of the data correctly, it should get at least the same performance compared to the CCA data analytics approach.

5.3.2 Results and Conclusion

After the training and optimization process for the CNN approach and the threshold estimations for the CCA approach, the resulting models were tested. The test set was composed of three subjects that were not included in training and validation. However, the resulting dataset has a bias towards the healthy case. Therefore, we tested on a random standardized subset of healthy cases and all pathological cases in the test set. In theory, we could just get lucky and get a good random subset of healthy cases. To counter this, we performed the test five times. The mean and one sigma of the accuracies of the CCA and CNN approaches are presented in Figure 5.7 respectively for the cases of binocular stimulation and monocular stimulation.

The algorithms performed better for binocular stimulation, especially for central stimulation (1). For the peripheral stimulation locations (2-5), the algorithms performed slightly worse. We can see an apparent dip in performance at location (2). This can be due to regularities in the testing data. For a more reliable interpretation, more data has to be collected and evaluated. However, as this significant drop for (2) is only present in the binocular case, which is generally better performing, we can assume a systematic error in the data, which might be due to blinking or other distractions. The far peripheral stimulation locations (6-9) were hard to recognize for both algorithms[1]. The results

---

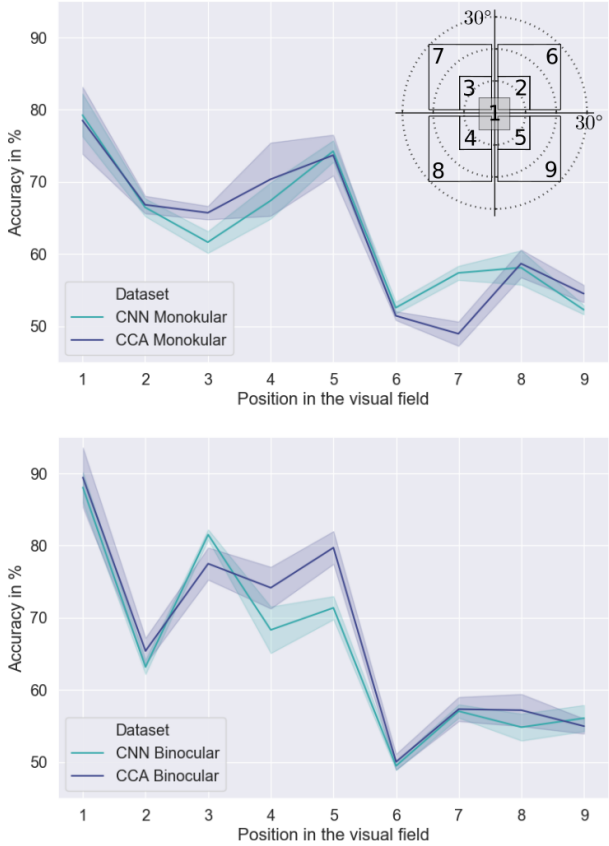[1] The far peripheral stimulation positions of each quadrant were summarized.

Figure 5.7: The graphs display the final accuracies of the tested system. The bottom graph compares the CNN and CCA approach for the binocular case, the top for the monocular case. The crosshair-like structure at the top right visually displays all tested positions. An accuracy of 50 % is considered random for a balanced test-set. Therefore, the algorithms work better than random if results of over 50 % are achieved.

Source: *Previously published by the author in [5].*

Position	CNN (Bi)	CCA (Bi)	CNN (Mono)	CCA (Mono)
(1)	88.1 ($\sigma = 2.1$)	89.4 ($\sigma = 4.1$)	79.3 ($\sigma = 3.0$)	78.5 ($\sigma = 4.8$)
(2)	63.19 ($\sigma = 1.0$)	65.42 ($\sigma = 1.8$)	66.5 ($\sigma = 1.8$)	66.9 ($\sigma = 1.2$)
(3)	81.5 ($\sigma = 0.7$)	77.5 ($\sigma = 2.2$)	61.7 ($\sigma = 1.5$)	65.7 ($\sigma = 0.9$)
(4)	68.3 ($\sigma = 3.2$)	74.2 ($\sigma = 2.9$)	67.4 ($\sigma = 2.5$)	70.4 ($\sigma = 5.1$)
(5)	71.4 ($\sigma = 1.6$)	79.7 ($\sigma = 2.3$)	74.3 ($\sigma = 1.5$)	73.7 ($\sigma = 2.8$)
(6)	49.5 ($\sigma = 0.6$)	50.1 ($\sigma = 1.1$)	52.6 ($\sigma = 0.8$)	51.5 ($\sigma = 0.6$)
(7)	57.1 ($\sigma = 0.9$)	57.4 ($\sigma = 1.7$)	57.4 ($\sigma = 1.0$)	49.0 ($\sigma = 1.7$)
(8)	54.9 ($\sigma = 1.9$)	57.2 ($\sigma = 2.2$)	58.1 ($\sigma = 2.4$)	58.7 ($\sigma = 1.9$)
(9)	56.1 ($\sigma = 1.8$)	55.0 ($\sigma = 1.0$)	52.3 ($\sigma = 0.7$)	54.5 ($\sigma = 1.2$)

Table 5.2: Comparison of the mean accuracies of the machine learning CNN (Convolutional Neural Network) and CCA (Canconical Correlation Analysis) statistical threshhold approaches. Bi means, that binocular stimulation of the eyes were performed (both displays of the VR device active). Mono on the other hand means, that one display of the VR device was deactivated. The numerical positions are mapped according to the visualization provided in Figure 5.7.

Source: *Previously published by the author in [5].*

were, better than random (which is 50 % for binary classification). However, the data shows that it is hard to correctly classify the pathology from only 2.5 s of data. Generally, SSVEPs are harder to detect in monocular measurements.

As we can also see in Table 5.2, which displays all numerical results, there is only little difference between the state-of-the-art CCA approach and the CNN approach. To put it in numerical: the CCA approach was 1.75 % better in the binocular approach but 0.07 % worse in the monocular case. For the size of the test set, this is a small difference and not statistically relevant as it is within or even below one sigma.

We can conclude that the machine learning CNN approach understood the underlying patterns of the data and correctly classified the information from end-to-end learning. Therefore, the trained machine learning algorithm was en-par with the carefully crafted mathematical approach, which in itself is fascinating.

The next question is, why, if the machine learning works, is it as good as the state-of-the-art and not better? The first and probably most realistic explanation is that the stimulation frequency and its harmonics are the only important features in SSVEP analysis. Furthermore, we can assume that either artifacts do not play a huge role in this kind of signal analysis, or, if artifacts are present, the CNN approach has no way to work around those. The other explanation is that if there are other features present than the stimulation frequency and its harmonics, might vary over multiple subjects or are hard to learn. In both cases, a larger dataset would be needed to identify relationships in the brain signals.

This research showed that it is indeed possible to perform perimetry only using SSVEPs, especially for central and close peripheral vision. However, to further improve the system and make it more reliable, a lot more data has to be collected, and not only over a rather young test population with simulated disease but on a larger cross-section of society, especially including visually impaired. Through the mentioned collection of data, the CNN approach might even perform better in the future, as machine learning learns from examples, and the more examples there are, the better the possibility of finding correlations and causalities in the data.

Furthermore, the stimulation pattern (checkerboard) is a rather simple stimulation. The human brain, on the other hand, is a highly complex visual processing engine. Therefore, if the stimulus is made more complex (e.g., faces, complex patterns, or objects), the response of the visual cortex could be more robust.

5.4   Circuittrodes

This section in the *Studies and Evaluation* chapter is focused on the novel introduced electrodes. In all experiments of this dissertation, either wet electrodes or the circuittrodes were used for signal acquisition. These circuittrodes might improve the way we utilize EEG-based BCIs by maintaining the easy-to-use principles of dry electrodes while simultaneously improving the signal-to-noise ratio. The concept and implementation of said electrodes were described in Sections 3.5.3 and 4.3.3 respectively. In the following, we will look at objective and subjective features of this new kind of electrode design and discuss

Figure 5.8: Comparison of state-of-the-art wet electrodes (center) and the circuittrodes (right). Both photos were taken directly after dismounting the EEG cap. The circuittrode leaves visibly fewer residues in the hair of the subject, making immediate showering obsolete.

its advantages and disadvantages compared to other state-of-the-art electrodes like wet, dry, and adhesive electrodes.

In comparison to adhesive electrodes; for dry, wet, and also the novel circuittrodes, an EEG-cap is needed for the mounting (compare Figure 4.28). The most obvious advantage of this new kind of electrode compared with wet electrodes becomes easily noticeable as soon as the electrode cap is dismounted. Figure 5.8 shows the direct comparison of gel-filled wet electrodes (center) and the circuittrodes (right). Both photographs were taken exactly after the cap came off without any clean-up. The residues are strongly noticeable in the center image, and the subject has to take a shower before they can resume their normal schedule. However, the novel circuittrode only relies on a minimal amount of conductive gel, which is barely perceptible after the cap is removed. After just combing through the hair, it is not noticeable that an EEG measurement was performed. Especially for subjects with long hair, this proved to be a highly effective method for performing EEG experiments.

Figure 5.9 shows a direct comparison of dry electrodes, wet electrodes, and the circuittrodes with active stimulation. The subject was presented with an SSVEP checkerboard stimulus at 7.5 Hz. A higher CCA while stimulating and a lower base-line-level while stimulation is off indicates better SNR. In

the presented experiment, the direct comparison between wet, dry, and the novel circuittrodes was performed on the same individual in succession so that different day-forms can be ruled out.

Generally, calculating the SNR for brain experiments is a tough problem as it is difficult to distinguish signal parts from noise in the recorded brain signal. For the sake of this evaluation, the SNR is defined as:

$$SNR_{\text{Electrode}} = 20 \cdot \log \left(\frac{\mu_{\text{Stimulated}}}{\mu_{\text{Not Stimulated}}} \right) \qquad (5.2)$$

with μ being the mean CCA value calculated for all stimulated and non-stimulated sections of Figure 5.9. This results in the follwing values for the SNR of each electrode: In the direct comparison, the dry electrodes performed

Electrode Type	Signal Quality (SNR)
Circuittrode	4.54 dB
Adhesive	5.48 dB
Wet	3.92 dB
Dry	3.19 dB

Table 5.3: The comparison of different electrodes SNR calculated as defined in Equation 5.2.

the worst of all tested electrodes, whereas the adhesive electrodes perfomed the best. This however is no surprise as the contact with dry electrodes is usually worse than with wet counter-types. If electrodes are glued to a shaved scalp, usually a really good singal to noise ratio can be expected, however only few individuals are voluntarily shaving their heads for EEG measurements. Wet electrodes and the novel circuittrodes have only a distance of 0.62 dB and therefore have fairly similar signal quality properties with a slight advantage for the circuittrodes. Graphically the resulting SNR values for all four electrodes is presented in Figure 5.10.

However, a disadvantage of the novel circuittrode is that they have to be prepared before application, which means that they have to be filled using a syringe before they can be applied. However, wet electrodes have the same or even worse disadvantage, as they need to be filled directly on the head.
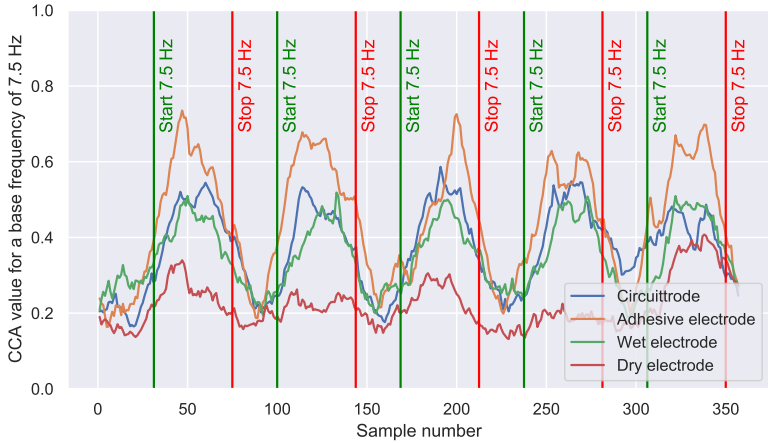
Figure 5.9: A subject with a shaved head was recruited to perform the electrode evaluation. The evaluation was performed with the help of a CCA analysis at 7.5 Hz.
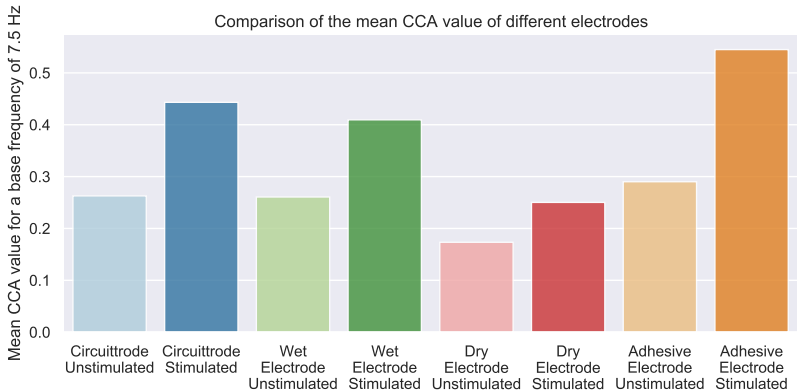

Figure 5.10: Mean values of the CCA for the stimulated and unstimulated parts of Figure 5.9. The larger the difference between the unstimulated and stimulated bars for one electrode, the better.

143

	Circuittrode	Wet	Dry	Adhesive
Signal quality (SNR)	+	+	-	+ +
Usable with hair	+ +	+	+ +	- -
Residues on head	+	- -	+ +	+
Preparation time	+	0	+ +	+

Table 5.4: Quantitative comparison of the novel circuittrode with other state-of-the-art electrode types.

Table 5.4 shows a composed comparison of the state-of-the-art and the novel circuittrode.

Also, according to Dr. Dag Scherfeld, a German Neurologist from Düsseldorf, the EEG measurement is usually one of the most unpleasant examinations for patients. Often they would even prefer a painful examination instead of having messed up hair.

Concluding, the novel circuittrode holds specifications like none of the other compared state-of-the-art electrodes. Especially the ability to be easily useable with hair is especially interesting for use in an outpatient setting while not compromising signal quality.

5.5 Towards more complex stimuli

The most crucial parameter, which is relatively easily changeable, is the stimulus used in neuro-ophthalmologic experiments. Therefore, we will look at variations of the conventional checkerboard patterns[2]. This determines how onset/reversal, superimposed noise, or pictures of faces might improve the elicited signals' magnitude (CCA).

The resulting comparisons may be used to improve existing SSVEP-based procedures using novel types of stimuli. The stimuli and the resulting average

---

[2] Data and description of the experiment were already published by the author in [4].

CCA over all subjects are presented in Figures 5.11 and 5.12. Set 1 uses checker and faces, and set 2 uses imposed noise on a circular checkerboard stimulus. The sets were shown to groups of 5 participants each (age range of 20-30)[3]. The resulting reactions to the SSVEP were recorded using an 8-channel EEG (500 Hz) from Neuroelectrics. Electrodes were positioned in analogy to the other experiments on the occiput. The data was analyzed using the Canonical Correlation Analysis (CCA) calculated over the stimulation window using the stimulation frequency and harmonics. Set 1 shows a significant increase in evocation when using pattern onset instead of reversal. Showing faces instead of checkers lead to a slightly increased response. Set 2 indicates that the addition of noise helps to improve the response at low levels but decreases again at higher levels. As we already know from the previous experiment, SSVEP-based assessment of neuro-visual functions has immense potential in ophthalmology. However, checkerboard stimulations do not necessarily provide the best responses. The addition of low-level noise to the stimuli might improve the generation of SSVEPs in participants. Higher-level stimulation patterns like faces could improve the responses even further. Therefore, we can conclude that future research on the visual field and simulation-based neuro-ophthalmology should consider using more complex stimulation methods.

---

[3] Due to the ongoing Corona-crisis, fewer participants than in previous studies were used in this study.
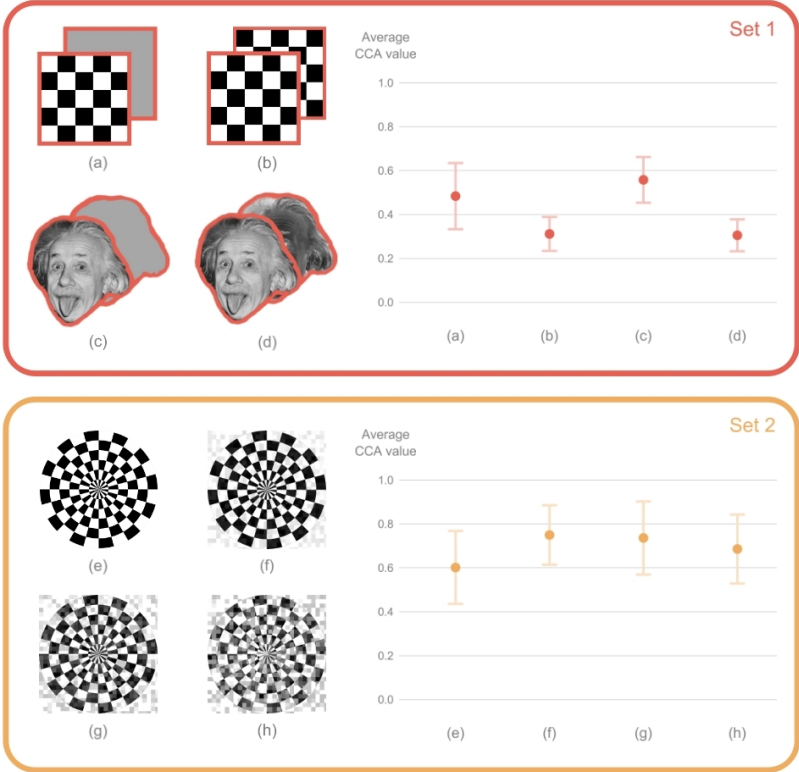
Figure 5.11: More complex stimuli might prove to be beneficial for neuro-ophthalmologic experiments. The upper graph shows the average CCA-value for (a) pattern onset checkerboard stimulation, (b) pattern-reversal checkerboard, (c) face of Einstein pattern onset, and (d) face of Einstein pattern-reversal. The pattern onset options perform slightly better than the reversal stimuli. Faces might provoke a slightly better signal than checkerboards. The graph below shows a comparison of a circular checkerboard pattern onset without (e), and with low (f), medium (g), and strong (h) noise added. Some noise to the stimulus might be better than no noise.
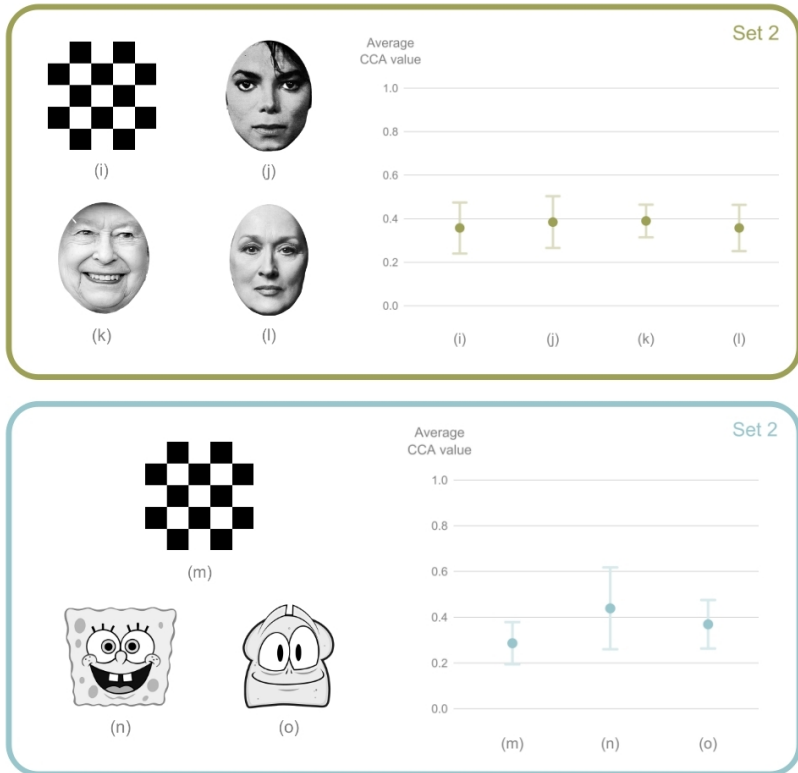
Source: *Previously published by author in [4].*

Figure 5.12: By using the faces of famous people, the CCA response might be increased slightly (j,k,l) in comparison to plain checkerboard stimulation patterns (i). Even cartoon characters will elicit SSVEP responses when flickered (n,o). It is interesting to notice that the well-known image of Spongebob (n) is tendentially eliciting higher responses than an unknown comic figure (o).

Source: *Previously published by author in [4].*

6 Conclusion, Discussion, and Future

> What we see is the solution to a computational problem, our brains compute the most likely causes for the photon absorptions within our eyes.
>
> *H. Helmholtz*

This chapter contains a summary of this dissertation. In addition, the research questions are answered, a critical discussion is provided, and an outlook for the future is given.

6.1 Summary

This book is built on six chapters. Initially, the motivation and the research questions are defined. Then important fundamentals and the state-of-the-art in neuro-ophthalmology and brain-computer technology are presented. The concepts chapter is introducing thought experiments and conceptual systems, which are then implemented in the implementation chapter. Finally, the Studies and Evaluation chapter is providing real-world insights into the actual use of the systems implemented in the previous chapters. In this last chapter, we circle all topics once again and provide a final overall result, answering the research questions from the first chapter.

In this book, we discussed vision of humans and of computers. Furthermore, we learned that one of the most fundamental concepts of vision is the understanding of the principle *what is where*. The overall goal was to develop

a system to objectively and reliably measure neuro visual function. Visual perception is performed by the human visual system (Section 2.1). Several highly complex individual parts and systems are combined into a complete system to make visual perception possible. In ophthalmology, the human visual field (Section 2.2), which can be examined using perimetry, is a central part of perception. However, vision not only happens in the eye but also in the brain (Sections 2.1.4 and 2.1.5), and therefore to measure vision, we need to measure the brain. The brain can be measured with brain-computer interface technologies (Section 2.3) like most prominently, electroencephalography. Steady-State Visually Evoked Potentials are a reliable tool for measuring visual activity in the primary visual cortex (Section 2.3.3). However, for displaying stimuli over a large field of view, technology more advanced than a standard computer monitor is needed. Therefore VR devices (Section 2.4) can be utilized. Those devices can also be used to simulate vision experiences and improve general understanding in ophthalmology (Section 2.4.1). To be able to understand extracted brain information, Section 2.5 gave an introduction to signal-processing, explaining the Fourier analysis and the state-of-the-art analysis for SSVEPs, the Canonical Correlation Analysis (CCA).

In the third chapter, conceptual ideas and concepts are presented. In Section 3.1 virtual reality is used to bridge between biological and technological vision, enabling the simulation of visual perception in VR. Whereas, the relatively new methodology of machine learning and deep learning for computer vision and pattern recognition is introduced in Section 3.2. For the development of such algorithms, the QUA^{3}CK machine learning process was introduced in Section 3.3. To evaluate and test the process and perform a quick deep dive into machine learning and technological vision processing the Deep Bees project was presented in Section 3.4. Finally, the Metis Videre Platform is introduced in Section 3.5. The platform combines all previous topics: Machine Learning, VR, and BCI for neuro-ophthalmology.

Chapter 4 provides valuable insights into the implementation of the concepts presented in Chapter 3. In detail Section 4.1 states how we can simulate vision in VR by utilizing perimetry scans. Sections 4.2, 4.3 and 4.4 explain the inner workings of the Metis Videre platform's software, hardware and data science modules respectively.

In the Studies and Evaluation Chapter (Chapter 5), five studies were performed and evaluated. In the first clinical case study (Section 5.1), the implemented

VR vision simulation environment was successfully evaluated in collaboration with the University Clinic in Heidelberg and the Team of Prof. Auffarth. In the second laboratory study (Section 5.2), the response of the visual cortex to stimulation in VR in varying sizes and frequencies was evaluated. That study leads to the international laboratory study in collaboration with MIT in the USA, which is described in Section 5.3. In that study, a deep learning-powered neuro-ophthalmological system was used to perform objective perimetry on subjects. The study uses a lot of the insights gained from the previous studies. Usability is an important topic for the real-world application of new technology. To improve the user experience for BCI-based technologies, the novel circuittrode was introduced (Section 3.5.3), implemented (Section 4.3.3), and evaluated in Section 5.4. The final study (Section 5.5) provides an outlook on where neuro-photonic experiments are heading, namely towards more complex stimulation patterns, which incorporate higher-level features like faces, cartoon characters, and the addition of noise to a stimulus.

6.2 Answers to the Research Questions

Utilizing all generated insights and knowledge of the previous chapters, we can answer the research questions postulated in Section 1.2.

Research Question 1: How can we use technology to bridge different visual perception stages from the biological to technological perspective and vice versa?

To answer this question, we need to think back to the very beginning, where we defined the individual steps of vision in technology and biology (Figure 1.1). Usually, both pathways are not interconnected by any means. However, by adding interfaces, we learned that we could achieve functional coupling of stages. The concept drawing 3.1 introduced that a display can interface to a computer-generated world. Especially a VR device enables an immersive experienceable for a human observer. This concept was proven through the implementations in Section 4.1 and the clinical case study in Section 5.1. Therefore it can be concluded that a display is indeed a good mechanism to interface visually to computer-generated worlds. Another way to interconnect biological and technological vision was introduced in the DeepBees study of Section 3.4. It uses a camera module to collect visual information from the real

world and uses AI computer vision algorithms on the collected data. This is enabling AI to see and perform pattern matching in the real world. Anlagogous, the final introduced bridging technology, is an EEG brain-computer interface. As visualized in Figure 3.13, we can access perceptional data directly from the brain. This data can then be utilized to understand what is perceived and what is not.

To conclude the answer to this first research question: By using displays, imaging sensors (cameras), and brain-computer interface, we can interface between biological and technological vision concepts.

Research Question 2: How can we simulate a visual perception experience with the help of computer technology?
As we learned in the answer to *Research Question 1*, we can use virtual reality displays to interface with a computer-generated world. By utilizing subjective individual perimetry data and rendering pipelines, we can simulate impaired visual perception. A wide variety of filters can be used to simulate vision conditions (see Figure 4.5). The implementation of immersive visual conditions can be performed with a 3D engine like Unity. By utilizing a GPU, the experience is finally rendered on a VR device (compare Section 4.1). For the specific case of individual vision loss due to glaucoma, a clinical study was performed at the University Hospital in Heidelberg. The study is described in Section 5.1 and concluded that the simulated experience gets very close to reality for the case of glaucoma. However, it is not necessarily able to exactly mimic the perception of reality according to patient feedback. Therefore, by utilizing VR technology and powerful graphic processing in today's computers, we can simulate visual experiences up to a certain degree. However, the field of view, resolution, dynamic range, and inconsistencies like the vergence-accommodation conflict in VR are still hurdles to a fully immersive visual experience.

Research Question 3: How can Deep Learning/AI technologies be utilized in pattern recognition for vision experiments?
Visual perception is a highly complex problem. It is complex in biology as well as in technology. In technology, we try to mimic the workings of biological vision regarding neurons with activation functions and utilize concepts like convolutions (see Sections 3.2.1 and 3.2.3). The idea of this book was to use the same mathematical principles which we transferred from biological to technological vision to understand what a computer is perceiving and apply

those principles again on the biological vision signals. The visual stimuli utilized in this dissertation were the SSVEPs (see Section 2.3.3). In Section 4.4.2 we learned about a CNN specialized in evaluating neuro-visual EEG data to correctly classify the pathological impairment level of vision in central, peripheral, and far peripheral vision. The result of the international laboratory study conducted with the MIT shows that biology-inspired AI models, like the used CNN, can perform on the same level as hand-crafted algorithms (compare 5.3). Therefore, we can conclude that, by utilizing computer vision AI, we can perform pattern recognition on the brain data recorded directly from the primary visual cortex of a subject. The AI was able to perform a par with the reference method. By increasing data sets and incorporating further information into the data stream of the CNN, it can be assumed that the performance of such AI systems could even exceed the state-of-the-art.

Research Question 4: How can we build an objective system that can perform perimetry using a brain-computer interface? What are the possibilities and limitations of the technology?

The previous research questions are leading towards this fourth question. As discussed, we can extract visual experiences directly from the brain using electroencephalography. *Research question 2* provided the answer on how we can simulate diseased vision with healthy individuals and perform experiments in a laboratory setting without needing a patient group for an initial study. According to *research question 3*, we can utilize AI to analyze brain data and extract information about visual functionality from it. These three steps enable the international laboratory experiment from Section 5.3 to be performed. Using this system, we can perform perimetry without the patient's conscious subjective feedback, enabling an objective system. Such a system is especially interesting for testing the visual fields of impaired young children and animals. Due to the bad performance of EEG technology in an everyday setting, novel electrodes for a complete system were implemented in Section 4.3.3 and evaluated in Section 5.4. These electrodes have a better SNR than dry electrodes. They do not leave a mess in the hair like wet electrodes, making them a very feasible alternative for an EEG-measurement setup regarding neuro-visual function testing. However, a downside of this objective perimetry is that for peripheral stimulation, the recorded response of the brain is weaker than for central stimulation. Therefore the system is especially interesting when the general working condition of the visual field is to be checked. More complex stimulation might improve the SNR, and, therefore, the signal in the

periphery might be detected better. Also, peripheral spots can be checked multiple times to increase the system's likelihood of a correct diagnosis. We can conclude that it is possible to build an objective perimetry device utilizing brain signals. The system could excel for specific applications, such as when used for disabled or unconscious patients. However, it has certain downsides, which will be critically discussed in the next section.

6.3 Critical Discussion

Through this book, which answered the research questions of the previous section, we learned that we can design a digital, objective brain-computer interface-based perimeter by combining a virtual reality display, EEG BCI and artificial intelligence algorithms. This perimeter is especially interesting for detecting visual field defects in patients who cannot communicate whether they have seen a stimulus in classical perimetry. The research presented in this book proves that an AI-powered system is fully functioning and is providing comparable results with other statistical methods. The data obtained in the presented laboratory study of Section 5.3 was focused on proving the general functionality of such a system. In the real world, however, there are several aggravating factors: In its current form without pupil tracking, the patient needs to look at a fixation target. Furthermore, the study is not designed to achieve the highest possible reliability for a single stimulus evaluation. The data indicates that by performing multiple measurements and incorporating majority voting, the validity of a perimetry measurement with AI and EEG can be increased further, as the system is significantly better than guessing. For peripheral stimulation, the signal is small, resulting in lengthy measurements if the far periphery is to be tested. With a stimulus size of 10°, the resolution of said objective perimetry needs to be improved further. Right now, the system is trained on binary decision-making (between pathological and non-pathological vision). The conversion to a decibel scale for each stimulus position is theoretically possible. However, that would need a rather large set of data from actual clinical perimetry and a SSVEP BCI like the one presented in this book. Additionally, the performance on peripheral stimuli has to improve. Concluding a perimetry system based on AI, VR, and a BCI has a strong reason for existence as it enables us to learn a lot about visual perception in humans. Furthermore, it is already possible to apply to a subset of medical practices in

its current state. However, more research and development that would exceed this dissertation's scope is needed to bring findings into clinical practice.

6.4 Future Prospects

If we want to bring a system like the ones proposed in this book into clinical practice, a lot of data must be collected through a large cohort of healthy and diseased patients. More complex simulations, which could also be gamified, could lead to novel ways to examine visual perception in general. There is no limitation only to perform perimetry. Several studies performed alongside this dissertation show strong evidence that SSVEPs can be used to detect contrast sensitivity, visual acuity, and even astigmatism through neuro-visual parameters. Furthermore, neuro-visual-based examination methods might provide valuable insights into the inner functioning, physiology, and psychology of the human brain. It is of utmost importance to look at such problems from a system perspective, incorporating the essential aspects of BCI experiments, like stimulus-synchronization, correct labeling, and stimulus design. In the age of artificial intelligence, collecting vast amounts of data will ultimately lead to further discoveries in this field. We can be excited about which aspects of visual perception and the brain humanity will be discovered in the following decades.

It can be expected that advances in brain-computer interface technology will further improve the signal-to-noise quality of brain measurements. Whether this BCI technology of the future is EEG-based or else, is highly debatable. However, there is a need for higher spatial mapping of the brain's inner workings to understand ourselves better. By using technologies like the ones introduced in this dissertation and the many amazing technologies which will be introduced in the future, the eyes can genuinely become the window to the human mind.

Acknowledgments

Doing a Ph.D. is a long process, and it is rarely a straight line from start to finish. The process has many ups and downs; my life was turned upside down multiple times in the last couple of years. I would not have been able to complete this dissertation without the help and support of many people. Collaboration and human connection became even more important in the covid-19 corona crisis and through several lock-downs. In the following lines, I want to thank some people who accompanied and helped me on my journey.

First, I want to thank Prof. Dr. Wilhelm Stork (Willy) for making this dissertation possible. Furthermore, I would like to thank him for his huge support, the countless highly interesting discussions about physics, the eye, entrepreneurship, and everyday life.

Furthermore, I want to thank Prof. Dr. med. Gerd U. Auffarth, for the successful collaboration in the VR studies at the Heidelberg eye clinic and for being the second reviewer of this dissertation.

I also want to thank my family, my mom Christine, my dad Georg, and my brother Raphael. Whenever life was throwing bricks at me, they were a safe harbor to return to. Without them, this work would have been impossible.

A huge thank you goes out to my girlfriend Katharina for the great support.

I also want to thank every student I was lucky to work with during my time at KIT. It was always refreshing to see new perspectives on problems. A special acknowledgment goes to Marius, Balint, and Heiko, whose Master Theses in the field of EEGs accelerated the progression of this work.

Furthermore, I want to thank Alexandre Urpi from MIT for the fantastic collaboration in the lab study regarding visual fields and further collaborative work.

I also want to thank my colleagues at ITIV and FZI. Especially the LAMA-Team, Gabriela, Daniel, Marco, Tim, and Leo, who helped me stay motivated and excellent in teaching.

Acronyms

ADAM	Adaptive Momentum Estimation
ADC	Analog-Digital Converter
AI	Artificial Intelligence
AMD	Age-Related Macular Degeneration
AR	Augmented Reality
BCI	Brain-Computer Interface
CAD	Computer Aided Design
CCA	Canonical Correlation Analysis
CNN	Convolutional Neural Network
CNS	Central Nervous System
CSV	Comma Seperated Value
CRISP-DM	Cross-Industry Standard Process for Data Mining
DLL	Dynamic Link Library
DRL	Driven Right Leg
ECD	Electrocardioagram
EEG	Electroencephalography
ELU	Exponential Linear Unit
EOG	Electrooculogram
FDM	Fused Deposition Modeling
FFT	Fast Fourier Transformation
fMRI	functional Magnetic Resonance Imaging
GPU	Graphics Processing Unit

GUI	Graphical User Interface
HMD	Head Mounted Display
ITIV	Institute for Information Processing Technologies
JSON	JavaScript Object Notation
KIT	Karlsruhe Institute of Technology
KDD	Knowledge Discovery in Databases
LGN	Lateral Geniculate Nucleus
LSL	Lab Streaming Layer
MAR	Minimal Angular Resolution
MEG	Magnetoencephalography
MIT	Massachusetts Institute of Technology
ML	Machine Learning
MLP	Multi Layer Perceptron
MSE	Mean Squared Error
MVC	Model View Controller
NIRS	Near-InfraRed Spectoscopy
OCT	Optical Coherence Tomography
OD	Oculus Dexter - Right eye
OS	Oculus Sinister - Left eye
QUA^{3}CK	The Question Understanding Algorithm Conclude Knowledge Transfer ML-Process
QUACK	Other way of writing QUA^{3}CK
RAM	Random Access Memory
ReLU	Rectified Linear Unit
REP	Retinal Evoked Potentials
RPE	Retinal Pigment Epithelium
SAP	Standard Automated Perimetry

SDK	Software Development Kit
SLA	Stereolithography Apparatus
SNR	Signal-To-Noise Ratio
SSVEP	Steady State Visually Evoked Potential
SVD	Singular Value Decomposition
TCP	Transmission Control Protocol
TE	Transformation Electrode
UML	Unified Modeling Language
VEP	Visually Evoked Potential
VR	Virtual Reality

List of Figures

List of Tables

Bibliography

[ACG⁺09]   Frederico A.C. Azevedo, Ludmila R.B. Carvalho, Lea T. Grinberg, José Marcelo Farfel, Renata E.L. Ferretti, Renata E.P. Leite, Wilson Jacob Filho, Roberto Lent, and Suzana Herculano-Houzel. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *The Journal of Comparative Neurology*, 513(5):532–541, apr 2009.

[AFF15]   Halim Cagri Ates, Alexander Fiannaca, and Eelke Folmer. Immersive Simulation of Visual Impairments Using a Wearable See-through Display. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '14*, pages 225–228, New York, New York, USA, 2015. ACM Press.

[AJM⁺11]   Silke Agte, Stephan Junek, Sabrina Matthias, Elke Ulbricht, Ines Erdmann, Antje Wurm, Detlev Schild, Josef A. Käs, and Andreas Reichenbach. Müller glial cell-provided cellular light guidance through the vital guinea-pig retina. *Biophysical Journal*, 101(11):2611–2619, 2011.

[AS08]   Ana Isabel Rojão Lourenço Azevedo and Manuel Filipe Santos. KDD, SEMMA and CRISP-DM: a parallel overview. *IADS-DM*, 2008.

[AUS18]   Alexandre Armengol-Urpi and Sanjay E. Sarma. Sublime. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pages 1–8, New York, NY, USA, nov 2018. ACM.

[Ayo96]   George Ayoub. On the design of the vertebrate retina. *Origins & Design*, 17(1):19–22, 1996.

[BAC96]   MICHAEL BACH. The Freiburg Visual Acuity Test — Automatic Measurement of Visual Acuity. *Optometry and Vision Science*, 73(1):49–53, jan 1996.

[Bas09]   M. Bass. *Handbook of Optics, Third Edition Volume III: Vision and Vision Optics*. McGraw-Hill Education, 2009.

[BB06]   Wilhelm Burger and Mark James Burge. *Digitale Bildverarbeitung: Eine algorithmische Einführung mit Java*. Springer-Verlag, 2006.

[BD80]   J K Bowmaker and H J Dartnall. Visual pigments of rods and cones in a human retina. *The Journal of Physiology*, 298(1):501–511, jan 1980.

[BE02]   Andreas U Bayer and Carl Erb. Short wavelength automated perimetry, frequency doubling technology perimetry, and pattern electroretinography for prediction of progressive glaucomatous standard visual field defects. *Ophthalmology*, 109(5):1009–1017, may 2002.

[Ber29]   Hans Berger. Über das Elektrenkephalogramm des Menschen - Dritte Mitteilung. *Archiv für Psychiatrie und Nervenkrankheiten*, 94(1):16–60, 1929.

[BGH⁺99]   N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor. A spelling device for the paralysed. *Nature*, 398(6725):297–298, 1999.

[BH98]   B Bengtsson and A Heijl. Sita fast, a new rapid perimetric threshold test. description of methods and evaluation in patients with manifest and suspect glaucoma. *Acta ophthalmologica Scandinavica*, 76(4):431—437, August 1998.

[BJF99]   Harvey J. Burd, Stuart J. Judge, and Mark J. Flavell. Mechanics of accommodation of the human eye. *Vision Research*, 39(9):1591–1595, 1999.

[BM08]   D Banks and Rachel Mccrindle. Visual eye disease simulator. *Proc. of International Conference on Disability, Virtual Reality and Associated Technologies*, 2008.

[Bro82]   Damien Broderick. *The Judas Mandala*. Fantastic Books (25. August 2009), 1982.

[BVB⁺10]   Claire A Brittain, Marco Vighi, Riccardo Bommarco, Josef Settele, and Simon G Potts. Impacts of a pesticide on pollinator species richness at different spatial scales. *Basic and Applied Ecology*, 11(2):106–115, 2010.

[BZR05]   Bei Jin, Zhuming Ai, and M. Rasmussen. Simulation of Eye
            Disease in Virtual Reality. In *2005 IEEE Engineering in
            Medicine and Biology 27th Annual Conference*, pages 5128–
            5131. IEEE, 2005.

[Coh17]   Michael X Cohen. Where does eeg come from and what does
            it mean? *Trends in Neurosciences*, 40(4):208–218, 2017.

[Cra16]   D P Crabb. A view on glaucoma—are we seeing it clearly?
            Eye, 30(2):304–313, feb 2016.

[Csi05]   András Csillag. *Atlas of the Sensory Organs Functional and
            Clinical Anatomy*. Humana Press, 2005.

[CT65]   James W Cooley and John W Tukey. An Algorithm for the
            Machine Calculation of Complex Fourier Series. *Mathematics
            of Computation*, 19(90):297, apr 1965.

[Cyb89]   G. Cybenko. Approximation by superpositions of a sigmoidal
            function. *Mathematics of Control, Signals, and Systems*,
            2(4):303–314, dec 1989.

[DOV03]   Jean Delbeke, Medhy Oozeer, and Claude Veraart. Position,
            size and luminosity of phosphenes generated by direct op-
            tic nerve stimulation. *Vision Research*, 43(9):1091–1102, apr
            2003.

[DRT19]   Frank Dittrich, Danny Rüffert, and Jonas Trezl. Durchblick
            mittels Alterssimulation! In Holger Fischer and Steffen Hess,
            editors, *Mensch und Computer 2019 - Usability Professionals*,
            Bonn, 2019. Gesellschaft für Informatik e.V. Und German UPA
            e.V.

[Erb01]   G. Erba. Preventing seizures from "Pocket Monsters": A way
            to control reflex epilepsy. *Neurology*, 57(10):1747–1748, nov
            2001.

[Ern05]   Ernst Niedermeyer and F. H. Lopes da Silva. *Electroen-
            cephalography: Basic Principles, Clinical Applications, and
            Related Fields*. Philadelphia: Lippincott Williams & Wilkins,
            5th ed. edition, 2005.

[ERW+94]   Stephen A Engel, David E Rumelhart, Brian A Wandell,
            Adrian T Lee, Gary H Glover, Eduardo-Jose Chichilnisky, and
            Michael N Shadlen. fMRI of human visual cortex., 1994.

[FHE⁺05]   Robert S. Fisher, Graham Harding, Giuseppe Erba, Gregory L. Barkley, and Arnold Wilkins. Photic- and Pattern-induced Seizures: A Review for the Epilepsy Foundation of America Working Group. *Epilepsia*, 46(9):1426–1441, sep 2005.

[FKBB82]   Frederick L. Ferris, Aaron Kassoff, George H. Bresnick, and Ian Bailey. New Visual Acuity Charts for Clinical Research. *American Journal of Ophthalmology*, 94(1):91–96, jul 1982.

[FS10]   John P. Frisby and James V. Stone. *Seeing, Second Edition: The Computational Approach to Biological Vision*. The MIT Press, 2nd edition, 2010.

[FYR⁺18]   Luca Fiorenza, Robin Yong, Sarbin Ranjitkar, Toby Hughes, Michelle Quayle, Paul G McMenamin, John Kaidonis, Grant C Townsend, and Justin W Adams. The use of 3d printing in dental anthropology collections. *American journal of physical anthropology*, 167(2):400–406, 2018.

[GAFT⁺14]   G. Gallegos-Ayala, A. Furdea, K. Takano, C. A. Ruf, H. Flor, and N. Birbaumer. Brain communication in a completely locked-in patient using bedside near-infrared spectroscopy. *Neurology*, 82(21):1930–1932, may 2014.

[GBC16]   Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[Gol69]   Gene H. Golub. Matrix Decompositions and Statistical Calculations. *Statistical Computation*, pages 365–397, 1969.

[Göt10]   Andreas Götze. *Methodische und experimentelle Untersuchungen zu multifokaler Stimulation für die objektive Perimetrie*. PhD thesis, aug 2010.

[GZM20]   Christian Garbin, Xingquan Zhu, and Oge Marques. Dropout vs. batch normalization: an empirical study of their impact to deep learning. *Multimedia Tools and Applications*, pages 1–39, 2020.

[HH91]   Jonathan C. Horton and William F. Hoyt. The Representation of the Visual Field in Human Striate Cortex: A Revision of the Classic Holmes Map. *Archives of Ophthalmology*, 109(6):816–824, 1991.

[HKW11]   Heike Hofmann, Karen Kafadar, and Hadley Wickham. Letter-value plots: Boxplots for large data. Technical report, had.co.nz, 2011.

[Hot92]   Harold Hotelling. Relations between two sets of variates. In *Breakthroughs in statistics*, pages 162–190. Springer, 1992.

[HRMB14]   Emma L. Hall, Siân E. Robson, Peter G. Morris, and Matthew J. Brookes. The relationship between meg and fmri. *NeuroImage*, 102:80–91, 2014. Multimodal Data Fusion.

[Hub95]   David H Hubel. *Eye, brain, and vision*. Scientific American Library/Scientific American Books, 1995.

[HXX⁺18]   Chengcheng Han, Guanghua Xu, Jun Xie, Chaoyang Chen, and Sicong Zhang. Highly Interactive Brain–Computer Interface Based on Flicker-Free Steady-State Motion Visual Evoked Potential. *Scientific Reports*, 8(1):5835, dec 2018.

[Jäh05]   Bernd Jähne. *Digitale Bildverarbeitung*. Springer; 6., überarb. u. erw. Aufl. 2005 Edition (7. April 2005), 2005.

[Joh86]   Ray Johnson. For Distinguished Early Career Contribution to Psychophysiology: Award Address, 1985. *Psychophysiology*, 23(4):367–384, jul 1986.

[JSCWBC20]   Pete R. Jones, Tamás Somoskeöy, Hugo Chow-Wing-Bom, and David P. Crabb. Seeing other perspectives: evaluating the use of virtual and augmented reality to simulate visual impairments (OpenVisSim). *npj Digital Medicine*, 3(1):32, dec 2020.

[JSL⁺19]   Linxing Jiang, Andrea Stocco, Darby M. Losey, Justin A. Abernethy, Chantel S. Prat, and Rajesh P. N. Rao. BrainNet: A Multi-Person Brain-to-Brain Interface for Direct Collaboration Between Brains. *Scientific Reports*, 9(1):6115, dec 2019.

[JTD07]   Valer Jurcak, Daisuke Tsuzuki, and Ippeita Dan. 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *NeuroImage*, 34(4):1600–1611, 2007.

[KB17]   Diederik P Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization, 2017.

[KDR14]   Michael Kaschke, Karl-Heinz Donnerhacke, and Michael Stefan Rill. *Optical Devices in Ophthalmology and Optometry*.

Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, jan 2014.

[Kor17]   Daniel Korgel. *Virtual Reality-Spiele entwickeln mit Unity: Grundlagen, Beispielprojekte, Tipps & Tricks*. Carl Hanser Verlag GmbH & Co. KG (11. Dezember 2017), 2017.

[KP$^+$88]   Glenn E Krasner, Stephen T Pope, et al. A description of the model-view-controller user interface paradigm in the smalltalk-80 system. *Journal of object oriented programming*, 1(3):26–49, 1988.

[KP07]   Stéphane Kluser and Pascal Peduzzi. Global pollinator decline: a literature review. 2007.

[Kra16]   Gregory Kramida. Resolving the Vergence-Accommodation Conflict in Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics*, 22(7):1912–1931, jul 2016.

[Lan00]   Gerhard K Lang. *Ophthalmology: A Short Textbook*. Thieme Medical Publishers, 2000.

[LBCM11]   James Lewis, David Brown, Wayne Cranton, and Robert Mason. Simulating visual impairments using the Unreal Engine 3 game engine. In *2011 IEEE 1st International Conference on Serious Games and Applications for Health (SeGAH)*, pages 1–8. IEEE, nov 2011.

[LBD$^+$90]   Yann LeCun, Bernhard E Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne E Hubbard, and Lawrence D Jackel. Handwritten digit recognition with a back-propagation network. In *Advances in neural information processing systems*, pages 396–404, 1990.

[LBS13]   Anat London, Inbal Benhar, and Michal Schwartz. The retina as a window to the brain - From eye research to CNS disorders. *Nature Reviews Neurology*, 9(1):44–53, 2013.

[LGSMV14]   M. Lopez-Gordo, D. Sanchez-Morillo, and F. Valle. Dry EEG Electrodes. *Sensors*, 14(7):12847–12870, jul 2014.

[Li 12]   Li Deng. The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Processing Magazine*, 29(6):141–142, nov 2012.

[LP82]   Harris R. Lieberman and Alex P. Pentland. Microcomputer-based estimation of psychophysical thresholds: The Best PEST. *Behavior Research Methods & Instrumentation*, 14(1):21–25, 1982.

[LZWG06a]   Zhonglin Lin, Changshui Zhang, Wei Wu, and Xiaorong Gao. Frequency recognition based on canonical correlation analysis for ssvep-based bcis. *IEEE transactions on biomedical engineering*, 53(12):2610–2614, 2006.

[LZWG06b]   Zhonglin Lin, Changshui Zhang, Wei Wu, and Xiaorong Gao. Frequency Recognition Based on Canonical Correlation Analysis for SSVEP-Based BCIs. *IEEE Transactions on Biomedical Engineering*, 53(12):2610–2614, dec 2006.

[Mar10]   David Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual*. MIT Press, 2010.

[Mea19]   Kenichi Mimori et al. Patent JP2019072309A - Biological information measurement electrode, biological information acquisition device, and method of manufacturing biological information measurement electrode, 2019.

[MM07]   Richard H. Masland and Paul R. Martin. The unsolved mystery of vision. *Current Biology*, 17(15):R577–R582, aug 2007.

[MMTC18]   Valeria Mondini, Anna Lisa Mangia, Luca Talevi, and Angelo Cappello. Sinc-windowing and multiple correlation coefficients improve ssvep recognition based on canonical correlation analysis. *Computational intelligence and neuroscience*, 2018, 2018.

[MMvdLV16] Sebastiaan Mathôt, Jean-Baptiste Melmi, Lotje van der Linden, and Stefan Van der Stigchel. The Mind-Writing Pupil: A Human-Computer Interface Based on Decoding of Covert Attention through Pupillometry. *PLOS ONE*, 11(2):e0148805, feb 2016.

[MPG90]   A. C. Metting van Rijn, A. Peper, and C. A. Grimbergen. High-quality recording of bioelectric events. *Medical & Biological Engineering & Computing*, 28(5):389–397, sep 1990.

[MPKSEP10] Gernot R. Müller-Putz, Vera Kaiser, Teodoro Solis-Escalante, and Gert Pfurtscheller. Fast set-up asynchronous brain-switch

based on detection of foot motor imagery in 1-channel EEG. *Medical & Biological Engineering & Computing*, 48(3):229–233, mar 2010.

[MPV20]   Neil Miller, Subramanian Prem, and Patel Vivek. *Walsh & Hoyt's Clinical Neuro-Ophthalmology: The Essentials*. Lippincott edition, 2020.

[Mus20]   Elon Musk. Neuralink Corporation Website, 2020.

[Nei11]   Gerhard Neidhöfer. 50-Hz Frequency [History]. *IEEE Power and Energy Magazine*, 9(4):66–81, jul 2011.

[NM14]   Kouros Nouri-Mahdavi. Selecting visual field tests and assessing visual field deterioration in glaucoma. *Canadian Journal of Ophthalmology*, 49(6):497–505, dec 2014.

[NSB85]   R. Navarro, J. Santamaría, and J. Bescós. Accommodation-dependent model of the human eye with aspherics. *J. Opt. Soc. Am. A*, 2(8):1273–1280, Aug 1985.

[NWJ+17]   Masaki Nakanishi, Yu-Te Wang, Tzyy-Ping Jung, John K. Zao, Yu-Yi Chien, Alberto Diniz-Filho, Fabio B. Daga, Yuan-Pin Lin, Yijun Wang, and Felipe A. Medeiros. Detecting Glaucoma With a Portable Brain-Computer Interface for Objective Assessment of Visual Function Loss. *JAMA Ophthalmology*, 135(6):550, jun 2017.

[OBB+16]   J. Vernon Odom, Michael Bach, Mitchell Brigell, Graham E. Holder, Daphne L. McCulloch, Atsushi Mizota, and Alma Patrizia Tormene. ISCEV standard for clinical visual evoked potentials: (2016 update). *Documenta Ophthalmologica*, 133(1):1–9, aug 2016.

[OPB+17]   Mark O'Sullivan, Jonatan Poveda Pena, Andrea Bocchino, Conor O'Mahony, Daniel Costello, Emanuel Popovici, and Andriy Temko. Comparison of electrode technologies for dry and portable EEG acquisition. *Proceedings - 2017 7th International Workshop on Advances in Sensors and Interfaces, IWASI 2017*, pages 15–20, 2017.

[Oys99]   Clyde W Oyster. *The Human Eye: Structure and Function*. Oxford University Press Inc, 1999.

[Pre11]   William H. Press. Canonical Correlation Clarified by Singular Value Decomposition. Technical report, 2011.

[PWK18]   Lukas Paulun, Anne Wendt, and Nikola Kasabov. A Retinotopic Spiking Neural Network System for Accurate Recognition of Moving Objects Using NeuCube and Dynamic Vision Sensors. *Frontiers in Computational Neuroscience*, 12, jun 2018.

[Qui06]   H A Quigley. The number of people with glaucoma worldwide in 2010 and 2020. *British Journal of Ophthalmology*, 90(3):262–267, mar 2006.

[Rad09]   Dean Radin. *Entangled minds: Extrasensory experiences in a quantum reality*. Simon and Schuster, 2009.

[RBG+15]   Peter N. Rosen, Erwin R. Boer, Carolina P. B. Gracitelli, Ricardo Y. Abe, Alberto Diniz-Filho, Amir H. Marvasti, and Felipe A. Medeiros. A Portable Platform for Evaluation of Visual Performance in Glaucoma Patients. *PLOS ONE*, 10(10):e0139426, oct 2015.

[RDF+06]   Giulio Ruffini, Stephen Dunne, Esteve Farres, Paul C. P. Watts, Ernest Mendoza, S. Ravi P. Silva, Carles Grau, Josep Marco-Pallares, Lluis Fuentemilla, and Bjorn Vandecasteele. ENOBIO - First Tests of a Dry Electrophysiology Electrode using Carbon Nanotubes. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1826–1829. IEEE, aug 2006.

[RDF+08]   G. Ruffini, S. Dunne, L. Fuentemilla, C. Grau, E. Farrés, J. Marco-Pallarés, P.C.P. Watts, and S.R.P. Silva. First human trials of a dry electrophysiology sensor using a carbon nanotube array interface. *Sensors and Actuators A: Physical*, 144(2):275–279, jun 2008.

[Rud16]   Sebastian Ruder. An overview of gradient descent optimization algorithms. *Biometrics*, 55(2):591–596, sep 2016.

[She00]   Colin Shearer. The crisp-dm model: the new blueprint for data mining. *Journal of data warehousing*, 5(4):13–22, 2000.

[SHH93]   RA Schachar, TSENG Huang, and X Huang. Mathematic proof of schachar's hypothesis of accommodation. *Annals of ophthalmology*, 25(1):5–9, 1993.

[SSF+17]   Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina

Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping*, 38(11):5391–5420, nov 2017.

[SSS+17]   David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359, oct 2017.

[TF07]   John Tromp and Gunnar Farnebäck. Combinatorics of Go. In H Jaap van den Herik, Paolo Ciancarini, and H H L M (Jeroen) Donkers, editors, *Computers and Games*, pages 84–99, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[Tur50]   A. M. Turing. Computing Machinery and Intelligence. *Mind*, 59(236):433–460, 1950.

[UCW+19]   Lisa A. Urry, Michael L. Cain, Steven A. Wassermann, Peter V. Minorsky, and Jane B. Reece. *Campbell Biologie*. Pearson, 11. editio edition, 2019.

[VB98]   Anne Van Cott and Richard P Brenner. Technical advantages of digital EEG. *Journal of clinical neurophysiology*, 15(6):464–475, 1998.

[VBC+19]   Oriol Vinyals, Igor Babuschkin, Wojciech M. Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H. Choi, Richard Powell, Timo Ewalds, Petko Georgiev, Junhyuk Oh, Dan Horgan, Manuel Kroiss, Ivo Danihelka, Aja Huang, Laurent Sifre, Trevor Cai, John P. Agapiou, Max Jaderberg, Alexander S. Vezhnevets, Rémi Leblond, Tobias Pohlen, Valentin Dalibard, David Budden, Yury Sulsky, James Molloy, Tom L. Paine, Caglar Gulcehre, Ziyu Wang, Tobias Pfaff, Yuhuai Wu, Roman Ring, Dani Yogatama, Dario Wünsch, Katrina McKinney, Oliver Smith, Tom Schaul, Timothy Lillicrap, Koray Kavukcuoglu, Demis Hassabis, Chris Apps, and David Silver. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, 575(7782):350–354, nov 2019.

[VDDP18]   Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience*, 2018:1–13, 2018.

[VO55]   H Von Helmholtz and Others. Über die akkommodation des auges. *Albrecht von Graefes Arch Klin Exp Ophthalmol*, 1(1), 1855.

[WAM14]   Robert N. Weinreb, Tin Aung, and Felipe A. Medeiros. The Pathophysiology and Treatment of Glaucoma. *JAMA*, 311(18):1901, may 2014.

[WBH+00]   J.R. Wolpaw, N. Birbaumer, W.J. Heetderks, D.J. McFarland, P.H. Peckham, G. Schalk, E. Donchin, L.A. Quatrano, C.J. Robinson, and T.M. Vaughan. Brain-computer interface technology: a review of the first international meeting. *IEEE Transactions on Rehabilitation Engineering*, 8(2):164–173, jun 2000.

[WHO19]   WHO. *World report on vision*. WHO, 2019.

[XMV+17]   Jiawei Xu, Srinjoy Mitra, Chris Van Hoof, Refet Firat Yazicioglu, and Kofi A. A. Makinwa. Active Electrodes for Wearable EEG Acquisition: Review and Electronics Design Methodology. *IEEE Reviews in Biomedical Engineering*, 10:187–198, 2017.

[YS19]   Robin E Young and Phillip N Sabes. Device Implantation Using a Cartridge. 2020, 2019.

[YTT+13]   Madoka Yamazaki, Don M. Tucker, Marie Terrill, Ayataka Fujimoto, and Takamichi Yamamoto. Dense Array EEG Source Estimation in Neocortical Epilepsy. *Frontiers in Neurology*, 4, 2013.

[Zha14]   Li Zhaoping. *Understanding vision: theory, models, and data*. Oxford University Press, USA, 2014.

[ZLL15]   Jing Zhao, Wei Li, and Mengfan Li. Comparative Study of SSVEP- and P300-Based Models for the Telepresence Control of Humanoid Robots. *PLOS ONE*, 10(11):e0142168, nov 2015.

Personal Publications

[1] S. Stock, F. Mazura, F. G. De La Torre, M. Gerdes, M. Schinle, and W. Stork, "MMAI - Mobile Moods AI; Electroencephalography Artifact Detection; Towards Objective Assessment of Mental States," in *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*. Institute of Electrical and Electronics Engineers (IEEE), 2021, p. 01–06.

[2] M. Schinle, C. Erler, S. Leenstra, S. Stock, M. Gerdes, and W. Stork, "A Decision Process Model for De-Identification Methods on the Example of Psychometric Data," in *2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICEC-CME)*. Institute of Electrical and Electronics Engineers (IEEE), 2021, p. 1–6.

[3] S. Stock, M. Gerdes, B. Kovacs, H. Maier, and W. Stork, "Elektrode zur nichtinvasiven Messung von elektrischen Körpersignalen auf einer Hautoberfläche," *WO Patent Pending*, 2021.

[4] S. Stock, M. Gerdes, M. Schinle, J. Veloso de Oliveira, L. Hauptmann, L. Martini, and W. Stork, "Evaluation of Innovative SSVEP Stimulation Patterns for Neuro-Ophthalmology," *Investigative Ophthalmology & Visual Science*, vol. 62, no. 8, p. 2391–2391, 2021.

[5] S. C. Stock, A. Armengol-Urpi, B. Kovacs, H. Maier, M. Gerdes, W. Stork, and S. E. Sarma, "A system approach for closed-loop assessment of neuro-visual function based on convolutional neural network analysis of EEG signals," in *Neurophotonics*, vol. 1136008. SPIE, 2020, p. 1136008/ 19 Pages.

[6] D. Grimm, T. Hotfilter, G. Molinar, M. Stang, and S. C. Stock, "Fakultätslehrpreis 2020: KIT-Fakultät für Elektrotechnik und Informationstechnik," 2020.

[7] S. C. Stock, J. Becker, D. Grimm, T. Hotfilter, G. Molinar, M. Stang, and W. Stork, "QUA³CK - A Machine Learning Development Process,"

in *Proceedings of Artificial Intelligence for Science, Industry and Society — PoS(AISIS2019)*, ser. Proceedings of Science, vol. 372.   Scuola Internazionale Superiore di Studi Avanzati (SISSA), 2020.

[8] J. Becker, D. Grimm, T. Hotfilter, C. Meier, G. Molinar, M. Stang, S. Stock, and W. Stork, "The QUACK Machine Learning Development Process and the Laboratory for Applied Machine Learning Approaches (LAMA)," Paper presented at Artificial Intelligence for Science, Industry and Society Symposium (AISIS 2019), Mexico City, Mexico, October 20–December 25, 2019. [Online]. Available: https://indico.cern.ch/event/781223/book-of-abstracts.pdf

[9] S. C. Stock, C. Erler, W. Stork, G. Labuz, H. S. Son, R. Khoramnia, and G. U. Auffarth, "Suitability of virtual reality for vision simulation–a case study using glaucomatous visual fields," *Investigative Ophthalmology & Visual Science*, vol. 60, no. 9, p. 2441, 2019.

[10] S. Stock, C. Erler, and W. Stork, "Realistic simulation of progressive vision diseases in virtual reality," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, ser. VRST '18. New York, NY, USA: Association for Computing Machinery, 2018. [Online]. Available: https://doi.org/10.1145/3281505.3283395

[11] C. T. Vu, S. C. Stock, L. T. Fan, and W. Stork, "Highly parallelized rendering of the retinal image through a computer-simulated human eye for the design of virtual reality head-mounted displays," in *Online SPIE Photonics Europe, 6 - 10 April 2020 - Optics, Photonics and Digital Technologies for Imaging Applications. Vol.: VI. Ed.: P. Schelkens*, ser. Proceedings of SPIE, vol. 11353.   Society of Photo-optical Instrumentation Engineers (SPIE), 2020, p. 1135316 / 21 Pages.

[12] J. Marstaller, F. Tausch, and S. Stock, "Deepbees - Building and scaling convolutional neuronal nets for fast and large-scale visual monitoring of bee hives," in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*.   Institute of Electrical and Electronics Engineers (IEEE), 2020, pp. 271–278.

[13] F. Tausch, S. Stock, J. Fricke, and O. Klein, "Bumblebee re-identification dataset," in *IEEE Winter Conference on Applications of Computer Vision Workshops, WACVW 2020, Snowmass Village, CO, 1 March 2020 through 5 March 2020. Proceedings*.   Institute of Electrical and Electronics Engineers (IEEE), 2020, pp. 35–37.

[14] S. Stock, A. Bertemes, M. Stang, M. Böhme, D. Grimm, and W. Stork, "Feedi - a smart wearable foot-band for navigation and guidance using haptic feedback," in *Human Interaction, Emerging Technologies and Future Applications II : Proceedings of the 2nd International Conference on Human Interaction and Emerging Technologies: Future Applications (IHIET – AI 2020), April 23-25, 2020, Lausanne, Switzerland. Ed.: T. Ahram*, ser. Advances in Intelligent Systems and Computing, vol. 1152. Springer, 2020, pp. 349–355.

[15] A. Fard, H. Bagherinia, S. Stock, and J. Straub, "Automatic Detection of the Optic Nerve Head in Line Scanning Ophthalmoscope Images in CIRRUS™ HD-OCT," *Investigative ophthalmology & visual science*, vol. 58, no. 8, 2017.

[16] S. Stock, S. Kubach, P. Sha, J. Straub, and W. Stork, "The Benefits of Real Time Pupil Tracking on the Quality of the B-Scan," *Investigative ophthalmology & visual science*, vol. 58, no. 8, 2017.

[17] J. Straub, S. Jaward, S. Stock, C. Leahy, C. Burns, and F. Sousa, "3D Visualization of Retinal Vasculature Using Virtual Reality and 3D Printing," in *ARVO imaging in the Eye Conference, Baltimore, May 06*, 2017.

[18] S. Stock, S. Kubach, and J. Straub, "Algorithm Parameter Optimization with Evolutionary Algorithms," in *ARVO imaging in the Eye Conference, Baltimore, May 06*, 2017.

Supervised Theses

[19] C. Erler, "Simulation Progressiver Augenkrankheiten mittels virtueller Realität," Master Thesis, July 2018.

[20] F. Schweikardt, "Objekterkennung mittels Transfer Learning durch Convolutional Neural Networks (CNN): Eine Fallstudie zur automatisierten Mahlzeitenerkennung," Master Thesis, Supervised in Collaboration with Gabriela Molinar, October 2018.

[21] J. Herrmann, "Objekterkennung mittels Transfer Learning durch Convolutional Neural Networks (CNN): Eine Fallstudie zur automatisierten Mahlzeitenerkennung," Master Thesis, Supervised in Collaboration with Gabriela Molinar, October 2018.

[22] L. M. Martini, "Instance segmentation of objects close to the user for improved perception in augmented reality using transfer learning," Bachelor Thesis, October 2018.

[23] M. R. Warnatz, "Design and prototyping of a virtual reality head mounted display with adaptable vergence and accommodation," Bachelor Thesis, January 2019.

[24] L. Pan, "Gaze estimation using convolutional autoencoders," Bachelor Thesis, Febuary 2019.

[25] J. Marstaller, "Deepbees: Using multi-task learning for joint detection, pose estimation and classification of pollen bearing bees," Master Thesis in collaboration with apic.ai, April 2019.

[26] A. Bertemes, "Prototyping and implementation of a smart foot bracelet for navigation and guidance using haptic feedback," Bachelor Thesis, April 2019.

[27] Y. Fuhrmann, "Integration of pupil tracking in an ophthalmologic virtual reality application for improved immersion and pupil movement analysis," Bachelor Thesis, May 2019.

[28] T. Eggers, "Signal processing in ophthamology: How the retina defines what we are seeing," Bachelor Thesis, May 2019.

[29] C. T. Vu, "Highly parallelized rendering of the retinal image through a computer simulated human eye," Master Thesis, May 2019.

[30] N. Penkov, "Rapid prototyping and evaluation of a connected distributed load measurement system," Bachelor Thesis, December 2019.

[31] R. Mokni, "Implementierung einer intuitiven Multimediaoberflaeche fur ein ophthalmologisches VR-Gerät," Bachelor Thesis, December 2019.

[32] B. Kóvacs, "Examining the human neuro-visual system using electroencephalography: A deep learning approach towards objective perimetry," Master Thesis, December 2019.

[33] B. Geiger, "Echtzeit Überwachungssystem zur Laserbechriftung von polymeren Materialien; Real Time Monitoring System for Laser Marking of Polymeric Materials," Master Thesis, December 2019.

[34] H. Maier, "An electroencephalography based system for objective refraction and visual acuity assessment using convolutional neural networks," Master Thesis, January 2020.

[35] S. Mueller, "Adaptive gesture control for automotive applications with few-shot-learning," Master Thesis, Supervised in Collaboration with Marco Stang, April 2020.

[36] S. Chen, "Ein Ansatz für Maschinelles Lernen in der Produktion: Ein Software-Framework zur Ausfuehrung von kuenstlichen Neuronalen Netzen in C++," Master Thesis, February 2020.

[37] A. Noshadi, "Correction of choroidal and retinal thickness maps using widefield optical coherence tomography," Master Thesis, In Collaboration with Zeiss Vision Science Lab in Tübingen, February 2020.

[38] M. Rees, "Artifact detection in electroencephalography signals for human brain computer interfaces," Master Seminar Thesis, March 2020.

[39] M. Gerdes, "A rapid prototyping approach to eeg electrodes for use in machine learning supported brain computer interfaces," Master Thesis, July 2020.

[40] J. Veloso de Oliveira, "Exploring visual stimulation for brain-computer interfaces: A comparison of stimulus designs for ssvep-based applications," Bachelor Thesis, September 2020.

[41] L. Hauptmann, "Measurement of astigmatism using a ssvep-based brain-computer-interface," Bachelor Thesis, November 2020.

[42] L. Loh, "Comparison of convolutional neural network and canonical correlation analysis in ssvep-based brain-computer interfaces," Bachelor Thesis, December 2020.

[43] F. Tausch, "Visual re-identification of bumblebees returning from foraging trips," Bachelor Thesis, In Collaboration with apic.ai, December 2020.

[44] J. Helmond, "Closed-loop user-experience in a brain-computer interface for neuro-visual experiments," Bachelor Thesis, September 2020.

[45] L. M. Martini, "A cross-platform approach to brain-computer interfaces for monitoring and device control using augmented reality," Master Thesis, November 2020.

[46] F. Tausch, "Visual re-identification of bumblebees returning from foraging trips," Bachelor Thesis, December 2020.

[47] H. Scheidt, "Development of a low-cost open-source electroencephalograph focused on scientific research," Bachelor Thesis, January 2021.

[48] L. Rothmann, "Development of an automatic laser diode lifetime testing setup," Master Thesis, In Collaboration with Carl Zeiss Meditec Inc., January 2021.

[49] S. Baumann, "Key classification in music recordings using deep convolutional neural networks," Bachelor Thesis, March 2021.

[50] P. Ahrendt, "Detection and classification of higher cognitive functions with ssvep using a brain-computer interface," Bachelor Thesis, March 2021.

[51] F. Mazura, "Artifact detection in an eeg-based bci for the assessment of neuro-visual function," Bachelor Thesis, April 2021.

[52] F. Gómez de la Torre Chujon, "Using brain computer interface for the recognition and classification of various mental states using neural networks," Bachelor Thesis, May 2021.

[53] R. Göbel, "Smart Home Raum-Sensor zur Personenverhaltenserkennung," Master Thesis, May 2021.

[54] T. Weyer, "On steady-state visual evoked potential detection in short response time for human brain computer interfaces," Master Seminar Thesis, Mai 2021.

[55] C. Kaiser, "Quality assessment and noise estimation of eeg data," Master Seminar Thesis, Mai 2021.

[56] M. Ruehle, "Development of a firmware for an ophthalmologic vr-device. towards the treatment of presbyopia by vergence and accommodation training," Bachelor Thesis, June 2021.

[57] N. Mohammadi, "Improving user experience and user interface by designing an optimized model for convoaid app," Master Thesis, June 2021.

[58] J. Hausberg, "Eeg-based objective adhd diagnosis using convolutional neural networks," Master Thesis, September 2021.

[59] S. Gross, "Investigating cognitive enhancement through binual beats utilising brain-computer interfaces," Bachelor Thesis, November 2021.