UNIVERSITÄT BIELEFELD

DOCTORAL THESIS

Combined analysis of Electroencephalography and eyetracking to create new windows into cognitive human-machine interaction

Author:
Dennis WOBROCK

Supervisor:
Prof. Dr. Helge RITTER

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Engineering

in the

AG Neuroinformatik
CITEC

January 18, 2022

Declaration of Authorship

I, Dennis WOBROCK, declare that this thesis titled, “Combined analysis of Electroencephalography and eyetracking to create new windows into cognitive human-machine interaction” and the work presented in it are my own. I hereby confirm that:

- The present Doctoral Thesis has been independently realized by myself.
- Where the supplied data sets, drawings sketches an graphical representations were not created by myself, this is clearly attributed.
- This work was done wholly or mainly while in candidature for a research degree at Bielefeld University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at Bielefeld University or any other institution, this has been clearly stated.
- I have not used any sources other than those indicated in the present work.
- Where I have consulted the published work of others, this is always clearly attributed (including Tables and Graphics).
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: 12.04.2021

Gedruckt auf alterungsbeständigem, holz- und säurefreiem Papier entsprenchend ISO 9706

UNIVERSITÄT BIELEFELD

Abstract

Technische Fakultät
CITEC

Doctor of Engineering

Combined analysis of Electroencephalography and eyetracking to create new windows into cognitive human-machine interaction

by Dennis WOBROCK

Since their inception, machines, computers and robots have steadily grown in complexity to solve ever more complicated problems. For these systems of ever-growing complexity to be usable by the largest number of people, they need to be made affordant through tests evaluating system interactions. These tests have however shortcomings, leaving users sometimes lost and frustrated with these systems.

In this work, we improve these interface evaluation test by relying on a Brain-Computer Interface combining Electroencephalography with Eyetracking. We use this bi-modal setup to provide complementary insights about a user's perception which can be gathered from any interaction scenario. To achieve this, we have created a set of methods which allow our system to be applicable and informative in a variety of situations. For scenario transposability we developed the Fixation-based Component Synchronization method, allowing to reestablish synchronous recordings even when markers are lacking. Using both recording modalities and the Fixation-related Potentials observable thanks to them, we propose four different methods which provide insight into how user's perceive the considered interaction. These four methods are the General Difficulty via Eyetracking (GDET) method, the Steady Peak Property Quantification (SPPQ) method, the Segment Frequency Bands Analysis (SFBA) method and the User-dependent Potential Variation (UdPV) method. These four methods provide respectively information about difficulties relating to the explored environment as a whole, specific elements in the environment, the task with which the environment is explored and specificities about the strategy with which the user explores the environment. We discuss and test the extent of all four of these methods in a series of three laboratory studies presenting artificial and natural interaction scenarios. The three scenarios presents different tasks and levels of difficulty allowing to establish the utility of these methods and verify their transposability between situations. All proposed methods are simple to implement and offer a new way to approach the analysis of interaction, both in a design environment and as a promising way to create adaptive interfaces.

Acknowledgements

Working towards this thesis during five years and finally writing up and presenting it was interesting and exciting but also sometimes a stressful experience.

I would like to express gratitude to Prof. Helge Ritter for giving me the opportunity to do a PhD Thesis on this fascinating topic. I learned a lot about my research topic and even more about research in general while working in the CITEC Neuroinformatics group. My thanks go to my thesis co-supervisor Andrea Finke from whom I received valuable feedback concerning by work and drafts of publication. I would also like to thank the people having taken part in the IP-19 project which served as the launching point for this PhD. Namely, I would like to thank Prof. Thomas Schack, Prof. Dirk Koester, Dr. Kai Essig and Shirley Mey for the help they offered during this work. I would like to stress my gratitude towards Shirley Mey who shared some of the experimental data with me which are analyzed within this work.

Additional acknowledgments go towards my colleagues in the Neuroinformatics group for interesting and valuable discussions as well as helpful advice with my work. Specifically, I would like to thank Nils Hachmeister, Lukas Twardon, Alexandra Moringen and Sasha Fleer with whom I shared an office space during this time. For his help to adapt to the differences between the German and French lifestyles, I also express gratitude towards Guillaume Walk.

A big part of this work would not have been possible if it were not for the students assisting me with the experiments performed during my time at CITEC. Namely, my thanks go to Raphael Dickfelder, Felix Friese and Philip von Neumann-Cosel for their help with creating some of the software used within the experiment. Most of all I would however like to thank Stefan Genster for his invaluable help and rigor during the execution of the various experimental procedures detailed in this work.

I would like to extend my thanks again to Andrea Finke for allowing me to help her with teaching tasks. The teaching aspects and interaction with students, both in the context of supervising bachelor thesis and group projects has been an enlightening experience and helped me with the formalization of my own material.

Last but not least, I would like to thank my family and friends who helped and supported me during these last five years.

Contents

List of Figures

List of Tables

List of Abbreviations

aMLRd	**a**daptive **M**ulti**L**inear **R**egression with **d**ispersion term
AUC	**A**rea **U**nder **C**urve
BCI	**B**rain-**C**omputer **I**nterface
BMI	**B**rain-**M**achine **I**nterface
CVeP	**C**odebook **V**isually **e**voked **P**otentials
ECG	**E**lectro**C**ardio**G**raphy
ECoG	**E**lectro**Co**rtico**G**raphy
EEG	**E**lectro**E**ncephalo**G**raphy
EMG	**E**lectro**M**yo**G**raphy
EOG	**E**lectro**O**culo**G**raphy
ERD	**E**vent **R**elated **D**esynchronization
ERP	**E**vent **R**elated **P**otential
ErrP	**Err**or-related **P**otentials
ET	**E**ye**T**racking
FCSync	**F**ixation-based **C**omponent Re**Sync**hronization
FFT	**F**ast **F**ourier **T**ransform
fMRI	**f**unctional **M**agnetic **R**esonance **I**maging
fNIRS	**f**unctional **N**ear-**I**nfra**R**ed **S**pectroscopy
FRP	**F**ixation-**R**elated **P**otential
GBVS	**G**raph **B**ased **V**isual **S**aliency
GDET	**G**eneral **D**ifficulty via **E**ye**T**racking
GSR	**G**alvanic **S**kin **R**esponse
HMI	**H**uman **M**achine **I**nteraction
HR	**H**eart**R**ate
HSI	**H**uman **S**ystem **I**nteraction
ICA	**I**ndependent **C**omponent **A**nalysis
ISO	**I**nternational **O**rganisation of **S**tandards
LDA	**L**inear **D**iscriminant **A**nalysis
MARA	**M**ultiple **A**rtifact **R**ejection **A**lgorithm
MLRd	**M**ulti**L**inear **R**egression with **d**ispersion term
PET	**P**ositron **E**mission **T**omography
PCA	**P**rincipal **C**omponent **A**nalysis
ROC	**R**eciever **O**perating **C**haracteristic
ROI	**R**egion **O**f **I**nterest
SFBA	**S**egmented **F**requency **B**and **A**nalysis
SPPQ	**S**table **P**otential **P**roperty **Q**uantification
SSVeP	**S**teady **S**tate **V**isually **e**voked **P**otentials
SVM	**S**upport **V**ector **M**achine
t-SNE	**t**-**S**toschastic **N**eighbourhood **E**mbedding
UdPV	**U**ser-**d**ependent **P**otential **V**ariation
UI	**U**ser **I**nterface
VR	**V**irtual **R**eality

Chapter 1

Introduction

1.1 Motivation

In every-day life, people use many different devices which help them complete tasks that would otherwise prove to be difficult and arduous. Washing-machines, cars or computers, for example, are some of such devices. All these devices significantly shorten and simplify complex tasks, such as washing clothes or traveling to distant destinations. As technologies progress, people are evermore frequently confronted with new systems created to help them to be more effective. With respect to the previously mentioned devices, this can take the shape of a new energy-efficient washing process, an automatic parking assistance or a software created to solve a previously unconsidered problem. These new technologies are put in hands of users who have little to no knowledge or expertise about them. Consequently, users need time to familiarize themselves with the functionalities these new devices provide. Learning and mastering a system can take a long time depending on its complexity and novelty to its users.

One of the main goals pursued in the field of Human Machine Interaction (HMI) is to reduce this necessary learning time by optimizing the interaction between users and systems. In most situations, improving interaction is achieved by creating an affordant communication interface. An affordant interface is designed in a way to present the system's capabilities in an intuitive and self-describing fashion to the user. In the given examples, this could translate to new buttons or icons on a washing machine or a screen displaying instructions. For the car parking assistant, it could consist of a display showing a planned path superimposed onto environment camera images. An optimal interface can take radically different appearances depending on the studied interaction. Elaborating such an affordant interface therefore requires careful consideration.

The field of Human Machine Interactions has developed series of tests, tools and strategies aimed to help with the creation of affordant user interfaces. Unfortunately, even when an interface is constructed through these techniques, it is still common for users to get confused or lost during manipulation. Emerging difficulties and confusion can lead to "disconnections" between system and user. These disconnections cause demotivation on the user's part to follow-through with the interaction. While this does not mean that the whole interaction is always halted completely, it still leads to a reduced communication quality between user and system, which is not desired. Even small grievances, when occurring often enough, can dissuade a user from engaging with a system further.

When such "disconnections" occur, it implies that the process of designing and testing the interface has failed to anticipate a user's particular grievances in a specific situation. This can, for example, be due to information being poorly presented or the user having mistakenly performed an action which was not in line with his or her goal. The HMI design and testing process aims to identify and apply remedies to such situations before they occur.

For this purpose, the design process is constructed to be iterative and cyclical. It is common to perform multiple iterations of this design cycle process to reduce the number of situations which can cause grievances. Every new design iteration will however also yield less improvements than the previous one. At a certain point, performing a new design cycle iteration is not worth its time and material cost of execution for the designers anymore. Yet, even in situations where seemingly enough iterations have been performed, problematic use-cases can still remain unidentified. Being that this main method of improving interfaces does not eliminate all situations of disconnection, how can disconnections still be avoided or at least be made to occur less frequently ?

Moments of disconnections most often occur due to problems in communication between a user and a system. When improving an interface, it is thus necessary to identify< many possible situations where the user experiences a lack in communication. However, considering all possible situations is overwhelming. To still consider the most common ones, an information gathering process is included in the HMI design process. This information gathering process can however also suffer from a lack of communication between users and the designers. For example, when designers test a new microwave button layout with users, users may notice small issues, such as a wattage selection not seeming intuitive or feedback sounds being too loud. However, as these issues seem inconsequential to the user in the test situation and are not precisely inquired about by the designers, this source of grievance persists and may bother end users. To advance the information gathering process accordingly, new approaches to obtaining information about the user's perception concerning the interaction are needed to improve existing approaches.

Brain Machine Interfaces (BMI) offer a novel approach to Human Machine Interaction. Indeed, as one of their most known use-cases, BMI can allow patients with motor difficulties to circumvent their bodies' own peripheral nervous system and still interact with machines and other systems. In this sense, BMI creates a new way of communication between user and interface. This is accomplished by analyzing a patient's brain activity and translating it into command signals for the system. The command signal is elaborated by inspecting a variety of properties of the recorded brain signal. Brain signals and their properties vary in relation to the interaction context. Brain activity properties also contain information about many different other aspects of an interaction. Notably, they can be employed as an alternative way to gather information about users' decisions during regular interactions. Combining such BMI insights with the classical approaches could serve to improve HMIs.

The analysis of brain activity through BMI provides a novel perspective to Human-Machine Interaction. Specifically, information from brain activity properties could prove to be useful at two different stages during the life-cycle of a User Interface. The first stage concerns the application of BMI within the HMI design cycle. Indeed, while the design cycle is extensive, it often fails to collect all valuable information about areas where the interface needs improvements (see the microwave example above). By providing access to brain activity, Brain-Computer Interfaces (BCI) allow for a new type of analysis of the user's performance and potential grievances. BCI could thus improve the information gathering yield of each iteration of design cycle without lengthening the overall process, as more cycle iterations classically would. Additionally, the observation of brain activity adds the possibility to gather information about aspects neither the designer inquired about nor the user vocalized. Monitoring brain activity offers a way to non-invasively follow the user's perception of the interaction, providing additional data to help for interface optimization.

This aspect of BMI also makes it potentially usable in a second stage of the systems life : the actual deployment of a system. Most machines are static entities, having no representation of their user aside from the direct input received through interface manipulation. While a system could be conceived which tailors its interface to its users by asking questions about their perception of the interaction, this would disrupt the interaction by shifting the user's attention to these questions and away from the core interaction. It would also generally length the interaction as a whole. BMI bypasses this issue by recording information without the attention of the user being pulled away from the main interaction. As such a BMI could make an interface more adaptive by actively responding and changing in relation to its users' needs.

Both approaches could greatly enhance Human-Machine Interaction. However, the application of Brain-Machine Interfaces in the HMI interface evaluation is a generally sparsely explored topic. At present, the application of BMI offers already significant challenges. For one, classically, brain activity is observed as it relates to specific interaction paradigms, where uncommon stimuli, such as flashing lights, are presented to the user. Integrating these paradigms into existing interaction may significantly alter the system and/or the underlying actions performed on the system. For this purpose, information obtained from these interaction paradigms are harder to transfer to interactions where the paradigm is not present. This would limit the generalization of any obtained BMI-related insight as attention would be allocated differently.

Another issue is that most of the research on BCI has been performed on very specific systems, making it difficult to apply this new brain activity modality in a way that is useful and transposable to a wide scope of interfaces and systems. To meet these challenges, additional information from other modalities which have already shown their usefulness in interfaces can be considered. Indeed, certain modalities have been successfully used to enhance Human-Machine Interaction in a similar way, such as heart-rate monitoring or Eyetracking. Taking inspiration and combining such modalities with BMI offers a new information-rich approach which provides new insights into interaction and interface design while circumventing the difficulties posed by a strict BMI approach.

In this work, we explore how such multi-modal Brain-Computer Interface can be reliably implemented to be effectively usable in a wide variety of systems without modifying the underlying interaction. Additionally, we also examine how the properties of this multi-modal BMI can be applied to identify a user's understanding of an interaction and ultimately improve it.

1.2 Focus

The main aim of this work is to propose a Brain-Machine Interface which can be easily integrated into the design process of a variety of systems and provides information relevant to their improvement. From a practical point of view, we approach this topic by employing a bi-modal BMI. More specifically, our system combines the recording of Electroencephalography (EEG) with the logging of eye movements through optical Eyetracking (ET). By combining both of these recordings, we will show that more relevant information for the designers can be extracted from manipulation.

These two modalities were chosen for a variety of reasons. Firstly, both EEG and Eyetracking devices can easily be worn by users without impeding their movements, making them portable to new testing scenarios with new interfaces and not mechanically changing the interaction process. More importantly however, these recording modalities were also chosen to compensate for each-others weaknesses: Brain activity recorded during interaction with a system is rich in information but difficult to accurately extract. To identify which particular information about the system is contained within the brain activity, the actions performed by the user need to be accurately logged and contrasted with the brain activity. As actions and their effects on the user significantly change between systems, a more commonly encountered reference point for brain activity is needed. Eye movements do not change in nature between interaction but are still guided by the considered interaction. Eye movements are present in nearly every interaction and can discretize the manipulation into fixations and saccades. Through this visual segmentation of interaction, a common paradigm can be established with which a great variety of interactions can be inspected. In other words, interaction is evaluated by observing how brain activity changes during the visual exploration of a system.

This approach is novel as it only relies only on eye movements and brain activity to access interaction relevant information for interface improvement. Thus, to achieve the main goal of this work, we focus this work around the elaboration and the testing of a set of methods for acquisition and analysis of ET and EEG data. These methods aim to achieve two objectives : providing information about the user's perception of the interaction and being applicable in a wide range interfaces and applications. For this purpose, our methods are devised to make the data from both recording modalities available in most situations.

We first study how the data of both recording modalities can be gathered reliably for a combined analysis. While mostly relying on existing methods, we elaborate a new method, the fixation-component resynchronization method (FCSync), based on an adaptive version of the Multi-linear Regression with dispersion term (aMLRd) method, to guarantee that synchronous recordings all possible scenarios.

The data from both recordings are then analyzed, first individually and then in combination. In the latter case, the segmentation of EEG data according to eye movements allows for the observation of Fixation-related Potentials (FRPs). In particular, the P100 potential will be observed in FRPs for its utility to obtain relevant information, such as the attentional effort displayed by the users.

Methods are established to examine four different kinds of grievances occurring during interaction : grievances associated with the layout of the system as a whole, grievances with specific system elements, grievances with the task demanded from the user or the grievances due to the users expertise or manipulation strategy. The methods inspecting these difficulties are examined across a series of three studies featuring artificial and naturalistic interactions. The first study offers a controlled environment allowing for the elaboration of these methods and a framework to apply them. The two following studies feature more focused natural interactions. In both of them, a single dimension of interaction is modified to inspect the transposability of the elaborated methods between interaction scenarios. The explanatory power of our methods does provide new and useful insights into interaction and can be applied to new interactions without requiring deeper changes or adaptations otherwise common for BMI (e.g. flashing or hidden stimuli). Specifically, the four methods for analysis which build the foundation of this work are : the General Difficulty analysis via Eyetracking (GDET) method, the Stable Potential Property Quantification (SPPQ) method, the Segmented Frequency Band Analysis (SFBA) and the User-dependent Potential Variation (UdPV) method.

1.3 Contribution

The main contributions of this work first and foremost concern the application of BCI in the context of interface evaluation. BCI has traditionally served as a way to interact with systems and their interfaces through thoughts directly. Most of the research is aimed at allowing medical patients to regain abilities their bodies have lost. These applications are however rarely of use for healthy users and thus uncommon for industrial use. By using passive BCI to gain information about the perception and allocation of cognitive resources for a more common task, a broader appeal of this technology can be reached. This aspect is furthermore strengthened by the development of a portable system which can easily be employed in a wide variety of interactions and does not require adaptations of any kind to be made to the considered system.

Specifically, the domain of interface improvement via the use of biosignals is scarcely explored. Indeed, most of the evaluation performed in this domain is through the usage of eyetracking and heart rate monitoring. Among these two, eyetracking is the most used modality, in particular in the domain of web design where it is used to determine which interface elements grab the attention of the user and help determine if the interface is simple to navigate. Here however, we show that one of the main appeals of eyetracking is also its unique capability to provide information about the users surroundings rather than only about himself, as is the case for other biosignals. This property makes eyetracking particular useful when used in conjunction with richer biosignals such as brain activity.

The combination of eyetracking and brain activity furthermore allows for the observation of Fixation-related Potentials. These potentials have so far been rarely explored, and in most exploration, cases were limited to controlled scenarios. Yet, the information contained within them is particularly useful when applied to a natural environment. This is most notably the case with the P100 potential which reflects the attentional effort a user puts into an eye fixation. Through this metric, particular places of attention can be identified and, when taken on a longer time scale can inform about the general performance of the users. Similarly, through the metrics obtained by each of the modalities on their own, a deeper inspection of the interaction quality can be performed as it relates to general difficulties or task changes. We formalize the gathering and evaluation of these metrics through a series of methods which reliably allow to access these insights.

Overall this domain of research is rather new. Our proposed approach aims to broaden the scope of BCI and biosignal use, potentially also opening the gate to adaptivity in live application of interfaces. Adaptivity would allow to adjust the interface to the user's needs on the fly.

1.4 Outline of the work

In the following chapter, the field of Human-Machine Interaction will be examined. Particular attention is put on how interfaces are designed and, more specifically, how these designs are tested and improved. This will help to identify what kind of issues are encountered when an interface is tested and how they may be solved. The third chapter explains how different biosignals have been employed to support interface investigations and what kind of added value they provide to interface improvement. This chapter also selects the two optimal modalities to evaluate interaction for our objectives : eye movements and brain activity. The fourth chapter pertains to the Brain Computer Interfaces and how to examine biosignals to obtain valuable information about the user's perception of interaction. In that chapter, combining the recording of the chosen modalities (i.e. through FRPs) will be discussed in greater detail. The methods used to extract brain activity properties and FRPs will be discussed at the end of the chapter. In the subsequent chapter, additional methods (i.e. aMLRd and FCSync) are proposed allowing for the observation of both biosignals and their combined FRPs in all situations. As each interface creates new circumstances that designers and users have to confront, methods are necessary to access these biosignal data reliably. The sixth chapter presents the first study. This study was used as an entry point to examine using both modalities to extract information about an interaction. Results are presented as they relate to the considered modalities : first separately and then combined. Each of the performed analyses is discussed in detail. The results are used to create a set of three methods (i.e. the GDET, SPPQ and SBFA methods) which can be employed by designers to reliably gain information about the user's grievances. As the results from this study were however only gained from a very controlled scenario, the stability of the proposed methods has to be verified in more naturalistic interactions, which are closer to real-life interactions.

The two following chapters investigate the transposition of these methods in two distinct natural interaction scenarios. Both scenarios were conducted on a computer screen but presented real world scenes to explore. Both investigate the variation of a single dimension of complexity, which was being identified as a potential origin of user grievances. The first natural interaction study, detailed in chapter seven, concerns the exploration of scenes featuring common and uncommon objects, thus considering element-wise difficulty identification. The second study explores the study of how changes in tasks can be evaluated using our methods. To generate multiple tasks, the scenes presented in this study contained a great number of objects, creating cluttered scenes. This second study, which is laid out in chapter eight, thus examines user task related difficulties. The set of methods proposed in the sixth chapter are tested in these scenarios and then reevaluated. In the study of chapter seven, a new method looking a user strategy grievances is also added to the set of methods (i.e. the UdPV method). The results from both studies are also evaluated between one another. This step allows to validate and expand the effectiveness of the proposed set of methods. The ninth chapter compiles all the results that were obtained through this setup. The tenth chapter present a conclusion about the setup and its utility to interface evaluation and suggests future applications and improvements.

Chapter 2

Inspecting Human-Machine Interaction

2.1 A Brief History

Since the creation of the first tools, humans have interacted with systems present in nature through various interfaces to improve their living conditions. From the use of flint and tinder to light a fire, to controlling robots to scout other planets, the interaction between humans and systems through diverse interfaces has proven to be one of the most important driving factors behind human progress. Interfaces initially helped in simple interactions, such as using baskets to transport foods or shovels to dig more effectively. As time went on and these first interfaces improved living conditions, humans felt comfortable to tackle new and more complex problems. As problems grew in complexity so did the interfaces which were devised and used to solve them. This cycle of technological improvement was perpetuated and resulted in the creation of even more complex interfaces, like airplanes to travel to distant places quickly or telegraphs to transmit information even faster. Initially, this evolutionary cycle came at a cost however : the machines created to interface with the physical world grew so complex that interfacing with them became difficult. Accordingly, to manipulate these more complex interfaces effectively, humans had to adapt their abilities and spend more time learning to adequately operate these interfaces. In the long run, this led to the creation of specialized professions and organizations perfecting the use of a specific interface to a given system designed to complete a known set of tasks.

Around 1945 A.D., explored systems and interfaces reached new heights of complexity with the advent of the first computing machines. These devices were created to handle abstract information. As the first fully-operational *Turing* machines, these devices were theoretically capable of performing any kinds of data manipulation [Minsky, 1967]. This revolutionary scope of possibilities meant that fully mastering their entire range of functions would take longer than ever before. As this became an impossibly difficult task, rather than improving human skill and adaptive ability, the system itself was changed to accommodate for its common use. Concepts surrounding Human-System Interaction (HSI), or more commonly referred to Human-Machine Interaction (HMI), such as User Interface (UI) arose during this time and gained influence ever since [Preece et al., 1994].

Human-Machine Interaction generally designates any communication process that occurs between a human and a system. User interface, on the other hand, is an umbrella term used to describe any space or tool that a user can take advantage of to interact with a specific system, hardware or software level. The goal of HMI translates itself most commonly in the conception, creation and validation of User Interfaces which optimize the interaction between a user and a specific system [Laurel and Mountford, 1990].

As problem and task complexity continued to grow, the requirements placed on computers also grew with it. Initially, this meant new hardware was first put in place to improve the UI of computers. This could consist in additional buttons, but also punch-cards and other peripheral devices expanding on inputs. As memory and processing power became cheaper and more broadly available in computers, adaptations to improve UI transitioned to the software side, changing how the information was presented to the user, by transitioning from terminals to graphical user interfaces [Myers, 1998]. When talking about HMI and UI nowadays, they mostly refer to adaptations done on the software side of the different computer programs. While it is not exclusive to software, most of the terminology and efforts are concentrated around that context [Hackos and Redish, 1998].

Memory and processing power have become a common and abundant commodity. Some systems are now capable of performing a vast array of different actions. The scope of actions of such systems can be so wide that they require extensive documentation to be read to be understood completely. One of the goals of HMI is to create affordant user interfaces. An affordant interface presents relevant information to the user in a way to minimize the necessity of extensive learning and adaption [Norman, 1999]. However, each new and enhanced interface, while coming closer to the goal of eliminating usage difficulties entirely, also still presents many moments where disconnection and disorientation between user and system can occur [Norman, 2013]. To overcome these moments, interfaces need further improvement and their quality needs to be inspected with respects to the goals of Human-Machine Interaction.

2.2 Evaluating Human-machine Interaction

As stated above, one of the objectives of the field of Human-Machine Interactions is to create UIs which enable users to quickly grasp and utilize the whole range of possibilities which the system provides. Minimizing learning time and decreasing the moments where a user may be lost within the interface is the challenge. Properly creating a UI which responds to these goals is a daunting task, as what makes UI functionally appealing is rarely obvious. Depending on the context of use, the user but also the type of the considered system, what the most important with the UI is can be radically different. For example, for navigation software in a car, some user may prefer a general overview of their trip while others only want to see the next turn they are meant to take. Consequently, what constitutes the optimal interface also strongly varies between situations.

One of the earliest methods developed to ensure UI quality came through the implementation of ergonomics norms. The field of ergonomics itself focuses on optimizing a UI around productivity. Its goal thus revolves around increasing the overall output generated by an interaction and can take many different aspects. The norms created by the International Organization of Standards (ISO) serve as a dictionary of best practices. ISO norms have shown their merit thanks to repeated past results and expert consensus [Peach, 1997]. During UI design, ISO ergonomic norms, in particular ISO 9000 norms, can serve as a baseline on which to build the interface. Norms exist on a wide array of topics and can sometimes even be a lawful requirement to create UI in the first place. However, norms are also static notions which cannot always represent the entire scope of interaction. To assess the quality of UI, such as considering an interface's constraints and certain user's specific grievances, more dynamic approaches need to be chosen.

Nowadays, the evaluation of UI mostly occurs during the design phase of a product. Happening at the inception stage of the product, the design phase is cyclically structured [Stone et al., 2005]. During each cycle, the system and its interface are incrementally improved to augment the quality of the product [Nielsen, 1993].

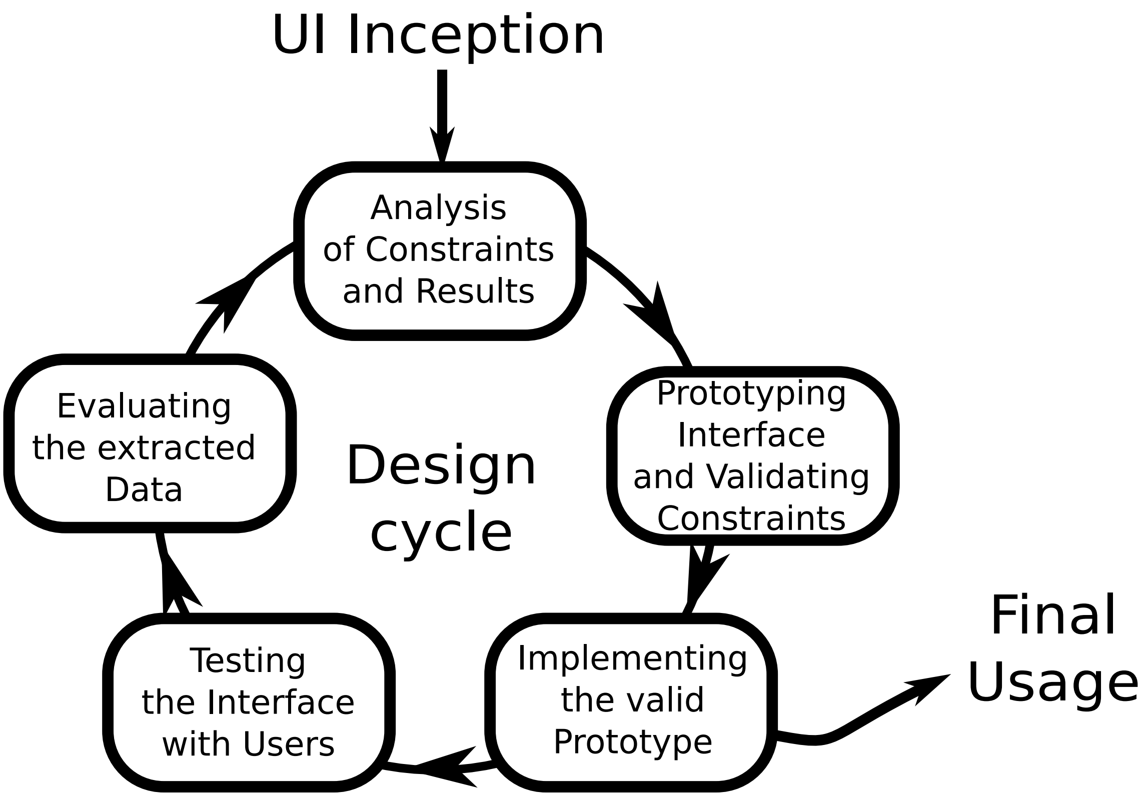


FIGURE 2.1: The iterative cycle for an interface creation before being presentation to the final user.

2.2.1 The Design Cycle

The design cycle begins at the conception of the User Interface of a specific system. When first entering the design cycle, first choices about the design are commonly made thanks to ergonomic norms, production constraints as well as previous experiences of the designers. In subsequent iterations of the cycle, information gained from the previous cycles is used to improve the interface. The general functioning of the cycle can be seen in Figure 2.1 which is described in the rest of this section.

In the first step of the design cycle, the constraints relating to the interface are analyzed. This means that both the system which is associated to the interface and potential users are considered. Manipulation scenarios are developed about how both sides would communicate with one another through this interface. Several objectives are then elaborated, aiming to guide the design toward the desired outcome of the conceived scenarios. As an example, a vending machine can be considered as an interface between a user and a system to exchange specific products, such as a bottle of water. Potential users can range from customers buying products to employees charged to restock the machine. One constraint would be that customers have to receive the product once they paid for it. One scenario to observe would thus consider a user ordering a product through the vending machine. The resulting objective would be for the customer to receive their desired product with least effort.

In the following step, a prototype of the interface is conceived, focused around these objectives. At the same time, several questions are formulated. They aim to verify the correct pursuit of these objectives as well as to examine any emerging behavior around the interface. These questions will help to create the basis for test protocols later. Using the given example, a machine is conceived which allows users to receive a product they have specified. Some of the possible questions are : did the user manage to communicate to the system which product they want ? Was their desired product made accessible to them ? Are they able to retrieve it ?

The third step is a mostly technical process, as it sees the proper implementation of the prototype interface into the system. In some cases, if the considered system is not ready at this point, a temporary system with dummy outputs is used instead.

The fourth step concerns the testing of this interface prototype. Here various users are put in contact with the UI and a variety of scenarios are tested. The scenarios aim to validate the overall quality of the interface and answer the questions elaborated in the second step of the cycle. In this phase many different users are asked to perform many different actions on the interface. All actions and feedback from the users are recorded. Considering the vending machine example : certain users are asked to select different products and retrieve them from the machine once delivered. Performance metrics such as the time required for selecting the desired product as well as the time required for retrieving it once made accessible can be logged.

In the next and last phase of a cycle, the data obtained from the previous tests is examined. Here interest is also put into how the data can be analyze. Indeed, the goal is not only to extract answers on the elaborated questions, but also to gain information about the subjective impressions users had about the interface. The results are then analyzed in the following first step of the next cycle where they serve to establish new constraints with which new objectives can be established. To come back to the selected example, if certain products take more time to be recovered from the vending machine interface, this can be considered undesirable. Consequently, the source of delivery time differences between selected products can be investigated and a constraint to eliminate this source of problem can be established in the next cycle.

The cycle can be repeated multiple times. Most frequently, once enough cycles have been performed the testing phase is ended and the constructed prototype becomes the definitive interface which is associated to the final system. In certain scenarios, a testing batch is still performed but the extracted data reveal no major need for interfaces changes, concluding the design cycle properly.

Optimally, during the fourth step (i.e. the test step), the interface should be handled by multiple different users during different moments of the products life span. Tests should also be done on the type of users who would be confronted with the system in these moments. For example, when talking about the vending machine, this would include testing the setting up of the machine, using it to select various product, using it to restock various products, as well as tests revolving around shutting the device down and dismantling it at end of its life span.

Practically, only a limited amount of resources, and thus cycle iterations, are available. The tested use-case scenarios have thus to be carefully chosen. During each of these scenarios, a series of questions is also presented to the users about their general impressions and satisfaction with their manipulation. These questions also must be chosen carefully as they are key to identify some of the deeper issues users may have with the interface or the system and the information yield of each iteration needs to be maximized.

For this purpose, a way to determine the most informative use-cases and associated questions is necessary. Notably, these aspects would have to be chosen as to best represent real interaction in as many facets as possible. The resulting data would also require analysis in such a way which is most relevant to the qualities of the interface and the system. One of such ways would be to identify the sources of grievances to better allow for improvement of the interface.

2.2.2 Sources of Grievance

Problems or grievances a user might face during manipulation can take different forms. To use the example of the vending machine again, a user might not be able to properly enter their selection, be that as in localizing the product they desire or communicating this information to the system. Or the product is presented to them in a fashion which makes it hard to retrieve. All these grievances can come from different sources, all interacting with each-other.

One of the simplest sources for grievances relates to issues present with a single interface element level. These are issues relating to one object, button or element of the interface. For example, users can have problems with the money slots on a vending machine. In these cases, the solutions of improvement only relate to one element, meaning that the scope of changes to consider is rather small. Similarly, a grievance may also be related to an ensemble of elements in the interface or the layout of the interface. These issues are more difficult to solve as it may require changes to the interface as a whole. Depending on the size of the considered interface, such changes can be significant and difficult to implement.

In parallel, issues relating to the operation of the interface itself can occur. These difficulties can take more complex appearances during manipulation. Notably, difficulties may arise when considering specific tasks. Indeed, the interface may be well suited for one task but leaves the user completely lost in another. For example, a user may find it easy to select a product but an employee finds restocking the vending machine very difficult. Finally, some issues may mostly depend on the variety of users manipulating the interface. While most users are comfortable with the manipulation, a user from a different background may struggle with the interface due to different expectations.

When evaluating the interface, one of the major difficulties in composing questions and analyzing the data is to properly identify the source of the grievances. Indeed, a single issue a user may experience can have roots in all four of these types of grievances (i.e. element grievances, whole scene grievances, tasks grievances and user strategy grievances). Tools are required to identify the source of the issues and how to remedy them properly. For this purpose, examining them through the concept of usability is often considered.

2.2.3 Usability

The concept of usability was developed to assess the quality of a user interface in regards to the users perception [Sinha, Shahi, and Shankar, 2010]. It is often confused with the term "user experience", which focuses on every aspect and service surrounding a product (including marketing and customer service). Usability concerns itself with only the investigation of the quality of interaction with the final product. What constitutes "quality of interaction" can vary significantly between products. As such, usability can concern many different aspects of an interaction. This is particularly the case in domains such as web design, where usability is commonly employed. Generally, however, the concept of usability can be decomposed into three distinct, but often correlated, categories [Frøkjær, Hertzum, and Hornbæk, 2000] : the *effectiveness*, *efficiency* and *satisfaction* when manipulating the interface. These three dimensions hold information about the strengths and weaknesses that lie within the interface. This is graphically represented in Figure 2.2. Understanding these dimensions helps to unearth some of the root problems with the interaction and fix them in subsequent design iterations.

The first of these dimensions is *effectiveness*. Effectiveness refers to the ability of the interface to give the user access to all the necessary tools to complete a task. To be effective,

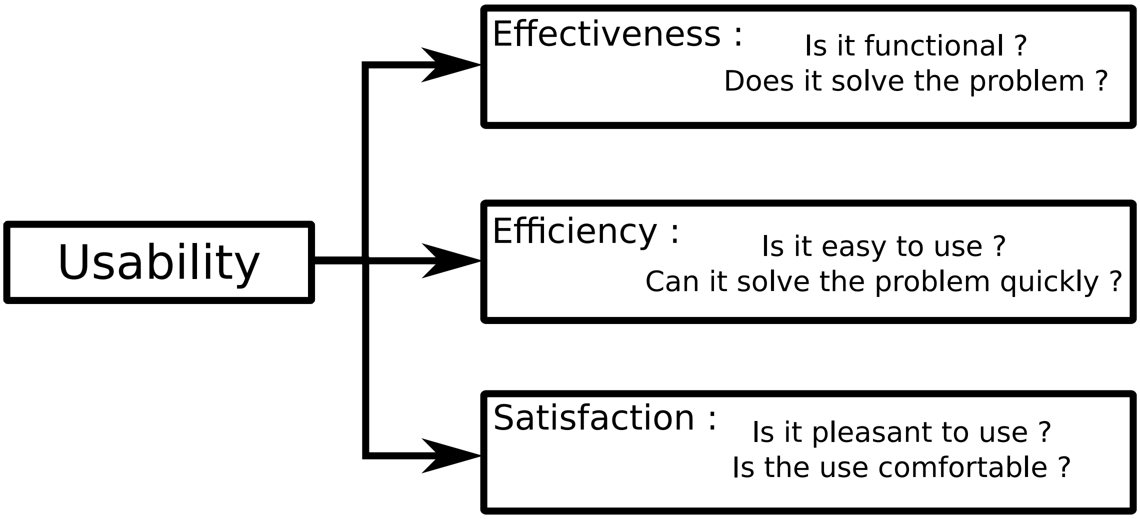


FIGURE 2.2: The main three dimensions which compose the notion of usability, establishing the quality and ease of use of a user interface.

an interface must propose options to users, which allow them to fully reach the goal they had in mind. In practical terms, effectiveness relates to the result of the manipulation and how close it is to the expected result. Commonly, it can be quantified by the number of errors made when pursuing a given task. To come back to the example case of a vending machine, let's consider an interface which separates the user from a selected product. The interface would be deemed effective if the commercial transaction can be achieved. A simple use-case determining effectiveness would be to check if the user manages to select a product, to pay for this product and to finally receive it. While good effectiveness is often the most important aspect of an interaction (did the user receive the product ?), it also does not give much meaningful insight into interaction quality (did the user like the interaction with the machine ?). Indeed, problems in interface effectiveness usually indicate more fundamental problems in the interactions. Finer measures are required to look for more subtle problems and improvements in interaction.

The second dimension is *efficiency*. Efficiency concerns the resources allocated by the user when interacting with the UI. This dimension refers generally to how much effort is dispensed and how much complexity is experienced by the user during manipulation. The reason behind the observation of efficiency is ultimately to minimize difficulty and effort of interaction. This dimension motivates to make interfaces more self-describing and more intuitively structured to users. This can, among other possibilities, be done by building on a design which the user is already familiar with. A generic metric used to quantify efficiency is task completion time, as a faster completion time indicates that the user didn't require too much effort to understand the interface. In the vending machine example, efficiency is notably reflected in the ease to locate and communicate to the system what product the user wants to buy. One can also consider more specific moments of interaction. Notably, what kind of motion the user must perform to put money into the machine and retrieve their product can also be considered as an observable factor of efficiency. If these motions are difficult or tiresome to perform, the efficiency of the interface can be improved.

The last dimension of usability is *satisfaction*. This dimension can also be considered as the "eagerness to engage" with the interface. Indeed, this dimension is more subjective than the other two as it evaluates the user's willingness and desire to manipulate the interface. Satisfaction generally asks the question if the user would engage or engage anew with the interface. Satisfaction is partially related to the aesthetic of the manipulated interface but it also relates to how the procedure of interaction was handled. While satisfaction is difficult to quantify, certain standard questionnaires have been established for its evaluation [Lewis, 1995,Kirakowski and Corbett, 1993]. For the vending machine example, satisfaction would concern the general location of the machine in the environment (where it is situated in a hallway or lobby), the sounds and noise it makes to confirm selection and if a sense of flow [Csikszentmihalyi, 1997], i.e. the demanded workload matches the users abilities, is kept during interaction. Satisfaction covers many more aspects which cannot always be easily expressed by the user. This renders satisfaction one of the more difficult dimensions to quantify.

As the concept of usability has evolved over the years, additional dimensions have sometimes been added. Two of these major dimensions are learnability and memorability [Scholtz, 2004]. They concern the aspect of how easy the interface is learned to manipulate and to reuse intuitively in subsequent manipulations. While both dimensions have become essential parts of studying usability, probing them effectively goes beyond the usability scope investigated in this work. Indeed, these additional dimensions would require multiple studies featuring the same users to be adequately studied. These dimensions are difficult to investigate based on a single study and experiment session.

The combination of the three main dimensions offers already an in-depth view into essential aspects of UI and the underlying Human Machine Interaction. Usability can be used as a framework to help dissect interaction at multiple points during the UI design cycle. These approaches can be applied to create objectives, elaborate meaningful questions and help to adequately analyze the test-users answers. The results offer a useful approach helping to identify where some of the user's grievances might lie. However, both the classical design cycle of UI and the concept of usability have their weaknesses and reach their limit for novel and less common interfaces.

2.3 Limits of HMI

Human-Machine Interaction is still a new and emerging field of research. Whilst methods have been developed to make the inherent complexity of a system more accessible to a broader range of users, problems can always be present. Indeed, due to lacks in concentration from the user or more fundamental omissions of information during the elaboration of a UI, difficulties still appear in the best interfaces across a variety of different users.

2.3.1 Limits of Evaluation

The study of usability during the design cycle of an interface helps to evaluate the interaction. This is done to make sure that the final use of the product is overall more satisfactory. However, during deployment and this final use, it is common to see situations arise which were not covered by the design cycle. This can be users manipulating the interface in situations which are more ambiguous or the user getting lost with very specific parts of the interface. These situations often present unforeseen difficulties for the user, which were not observed during the designing phase. However, as the interface cannot be changed during this final use, these difficulties are more arduous to correct. While far from making it impossible to use, the interaction, and subsequently the interface, is worse off for these issues. Even when objectives, questions and considered use-case scenarios were elaborated carefully during the design cycle, situations of disconnections and disorientation of the user can still arise.

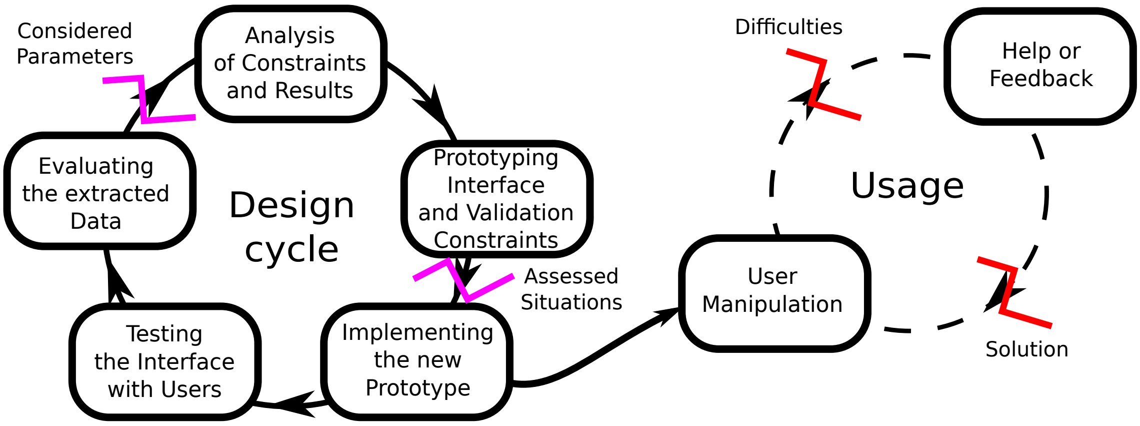


FIGURE 2.3: The iterative design cycle of an interface and the process of using the interface. Colored cracks represent the different moments causing potentials disconnections during interaction.

As displayed in Figure 2.3, this can be due to different failures or omissions which have occurred during the interface's life-cycle, as represented by the cracks. Some of these failures only become apparent after deployment but could have been corrected during the design cycle (represented by purple cracks). Most notably, this can happen when the questions concerning the use of the interface are formulated. Designer and HMI experts are not able to fully envision all the possible use-cases that may occur during the final usage. Design resources are often limited, and the design cycle is only repeated for a limited amount of times. Similarly, after having performed a series of tests, the obtained data may present more information than what is ultimately extracted in the evaluation step. The data are mainly analyzed as they relate to the stated questions and objectives. Additional information from usability tests thus gets lost, and the final interface does not benefit from all possible improvements the data could have yielded.

Once the disconnection occurs during the final manipulation of the interface (represented by red cracks) remedying is difficult. At this point, information about the difficulties is only rarely relayed back to the designer. The system on its own is rarely equipped to aid the user in such situations. While technical helplines and documentation may be present, resorting to them will stifle the interaction. In either of these cases, the user experiences disconnections which are difficult to solve and can be frustrated because of them.

Two main approaches can be envisaged to remedy these disconnections. The first one concerns the improvement of the interface design cycle. More specifically, the aim is to improve the design and evaluation of interaction scenarios proposed to the user. During the creation of use-case scenarios, most of the common scenarios of use are covered. However, as the designer may be anchored in their vision of the interface and the user may not always be able to properly express their grievances, or forget about minor grievances, some seemingly uncommon situations are not explored. Creating ways for designers to efficiently broaden their view of the interaction and to give possibilities for the user to better communicate their issues may help. Additionally, the risk exists that the data obtained are not evaluated to their fullest. Establishing an approach that points out which situations during interaction need further improvements may help to make the design cycle yield more effective changed to the interface, while keeping the resources used the same and increasing efficiency. In other words, designers gain more, and more precise, information about the user from a same design cycle iteration.

The second approach would consist in helping with grievances as they occur during actual manipulation. This would occur long after completion of the interface. For this purpose, the interface would need to adaptively respond to the user's grievances on the fly. This would be possible if the system would collect new kinds of information about its user. Creating such a bi-directional system may prove difficult. Indeed, both approaches are challenging to implement as they need to be correctly integrated in order to not disturb the underlying interaction too much (both on a mechanical and procedural level).

2.3.2 Limits of Interfaces

Most common user interfaces are static entities in regards to their user. The information UIs gather about their users is based on the input which is most commonly collected via the regular operation of the interface, such as button presses. Perceptions of difficulty, grievances, or other more general information about the user are rarely considered by the interface. Indeed, in UI, the user still does most the effort learning and adapting to the interface. To obtain new information which could help the UI adapt, the underlying system would have to be significantly modified, to include new sensors which monitor the user. This could however radically change the underlying interface and its operation by the user, depending on how the sensors and the paradigm used to collect information from them is implemented.

This also poses a series of ethical concerns regarding the kind of data gathered through these sensors. Indeed, the collected data are considered to contain sensitive personal information about the user (e.g. about their physical or mental health) and their mishandling could infringe on the users right to privacy. Even if data security is guaranteed and the data is not used or transmitted beyond the device, many users may still feel reluctant to use the sensors to improve their interaction with the interface. To ensure ethical data usage, users have to be clearly informed about how their data will be recorded, used and handled within a foreseeable time frame [Drennan, Sullivan, and Previte, 2006]. These conditions are appropriately met in the context of the evaluation phase of the design cycle where select users are asked to manipulate the interface in specific and agreed upon ways. In this study, the use of sensors and their data will only be discussed as they relate to the evaluation step of the design cycle phase of a user interface's lifespan.

In order to improve the design cycle phase, more information can be collected about the state of the user during use. One classical approach would consist of questioning test users more frequently and more extensively. Questioning users during manipulation would however distract them from the central operation of the interface. Questioning users just after interaction may result in relaying incomplete information. As mentioned above, users may have already forgotten some small grievances they deemed unimportant. Such grievances could however prove to be more harmful to other users, and their identification may thus be valuable. In either situation, it is still not certain that this additional information can be properly parsed and used by the designer.

The presence of a continuous monitoring during an interface testing phase may help to improve the analysis of the quality of interaction. Indeed, monitoring systems which autonomously look at the user during manipulation and extract information may give reference points to designers to assist them with their assessment of the interface. To implement such monitoring, the data must be gathered without requiring the direct attention of the users, that is not distracting users from performing the main task. This type of information has already been gathered in the past, in other scientific domains, through the monitoring of biosignals [Jung et al., 2005 ; Riseberg et al., 1998]. However, biosignals offer a wide variety of data about the user and their state of mind. To be used in the proposed interface evaluation, a deeper look into biosignals and their applicability to manipulation tasks is necessary. In the following chapter, different biosignals will be discussed in greater detail to identify which is optimal for the identification of user grievances.

Chapter 3

Choosing Biosignals for HMI

3.1 Selection Criteria

Biosignals refer to physiological measurements taken from the human body. More specifically, biosignals are continuous recordings of physiological activity observed in different parts of a living body. The term "biosignal" generally covers a wide scope of signals, varying in nature. Indeed, these signals can range from the contractions of the heart muscles, or movement of the lungs to electrical potentials occurring on the scalp. In contrast to biometrics, biosignals are necessarily measured over time rather than at just a single instance. Historically, the first formal recorded instance of biosignals usage was pulse measurements in the 17th century [Santorio, Keill, and Quincy, 1720]. It is however likely that biosignals were used long before this date as a means to monitor the human body in medical purposes. Biosignals are of particular interest for medicine as they allow to check on a person's health without the need for that patient to be able to communicate or even to be conscious [Lüderitz, 1995].

Heartbeat rate and respiratory rate count among the first biosignals to be monitored, requiring only simple tools such as a stethoscope and a clock. Both of these ways of observing the human body, called modalities from here on, served as a starting point for the field of biosignal monitoring [Kaniusas, 2012]. Indeed, biosignal technologies followed an evolution cycle comparable to interface technologies improving over the years. Systems were created to track biosignals more reliably and enable recording of more complex biosignals. This led to systems allowing for the observation of skin conductivity, the tracking of eye movements and the monitoring of muscle contractions. This evolution now even allows for observations of brain activity. The continuing pursuit of developing methods to record physiological activity is still a highly sought-after goal [Silva, Fred, and Martins, 2014].

The current scope of trackable biosignals is vast. It considers the monitoring of physical tremors of the body (e.g. expansion and contraction of the body during breathing), changing electromagnetic activity (e.g. voltage changes during muscle contraction) as well as observing body changes through other optical methods (e.g. dilation of the pupil). All these different biosignals and approaches require specific setups to be observed properly. A specific subset of these setups relies on intrusive sensors which require surgery and/or strongly restrict the natural motions of the body. These intrusive techniques, also called "invasive" in regards to monitoring brain activity, often yield less artifact prone data [Semmlow, 2008]. In the case of medical applications, the utilization of invasive biosignal recording techniques can sometimes be required for a proper diagnosis.

For medical diagnosis, the patient is made to interact with biosignal technology. This interaction is solely done to gather information about the patient's health. Accordingly, ensuring the quality of the recorded data trumps ensuring the patient's comfort. In the context of this work however, we want to examine biosignals as they relate to regular interactions with other kinds of systems. Ignoring the users comfort for data accuracy runs the risk of significantly altering a regular interaction. For example, all dimensions of usability (see Chapter 2) completely change when operating a vending machine normally and when operating one confined to a small space with little freedom of movement due to sensor cables being in the way. Applying analysis results from one interaction to the other will likely not improve it, due to them having become too different. A similar flow to the interaction can be kept by using non-invasive setups. While non-invasive techniques for recording biosignals, such as Electrocardiography (ECG) or Electromyography (EMG), are more prone to signal artifacts, they are much less constraining. Non-invasive biosignals can be more easily employed in a wider scope of applications.

In this work, we use biosignals to complement the traditional evaluation of interfaces by obtaining new information about a user's perception of a considered interaction. The aim of this work is also to make the extraction of this information applicable to the widest range of applications possible. As stated above, this aspect requires that the observation of biosignals changes the manipulation of the interface as little as possible. Additionally, biosignals must be chosen which provide rich enough information to identify grievances about an interaction. Utilizing biosignals outside of the medical domain to extract meaningful information is a relatively novel use of this technology. Certain applications using biosignals have already been developed which will serve as inspiration for the selection of the most appropriate biosignals.

The use of biosignals in interfaces outside of the medical field include the logging of static biometrics. This is the case when using fingerprint scanners or voice recognition software [Delac and Grgic, 2004]. However, these metrics hold little to no informative value about the users perception during manipulation (with some rare exceptions). In the field of HMI, techniques using more complex biosignals, reflecting changes in the interaction, are necessary. For example, biosignals such as brain activity or heartrate, have been used to extract information about the user's emotions [Van Den Broek et al., 2009] and arousal [Heger, Putze, and Schultz, 2010]. While this demonstrates that new insights can be obtained through biosignals, the specific biosignals to analyze must be chosen with consideration to be useful in as many situations as possible. Constructing biosignal frameworks for analysis of human perception has already been attempted in the past [Silva et al., 2014 ; An, Kim, and Kim, 2013 ; Oliveira, Lopes, and Guimarães, 2009]. In these studies, two of the biosignals which were shown to be the most promising to gain new insights into the interaction, were the observation of eye movements and the recording of brain activity. Indeed, both modalities offer complementary examinations into the interaction and are both relatively simple to implement.

3.2 Eye movements

Eye movements designate any motion that is performed by the eye globe and the eye lid [Holmqvist et al., 2011]. These motions can be differentiated into four major categories. The first are blinks. Each time the user closes and reopens their eye lids is counted as a blink. The second type of motion are saccades. Saccades occur when the user's gaze jumps from one position to another. These motions occur very quickly and can occur over very short distances, known as micro-saccades. The third category of motion occurs in the time separating saccades and are known as fixations. Fixations occur when the gaze stays steadily in a small visual region for an extended period of time. It is usually within this period that the actual perception of the user's environment occurs. The fourth and last category of eye motions are called smooth pursuits. These motions occur when a person's eyes are following a moving object. Smooth pursuits differentiate themselves from saccades by a slower speed of the rotation of the eye as well as the lack of perceptual blindness as it occurs in saccades [Simons and Levin, 1997].

Eye movements are a peculiar type of biosignal as they not only provide information about the person who has produced them, but also provide information about the environment the person has to navigate. This statement relies on the so-called eye-mind hypothesis [Just and Carpenter, 1984], that eye motion is associated with shifts of attention of the user. In this regard, the gaze direction is guided by the task the subject is currently trying to complete [Hayhoe and Ballard, 2005] but also by the structure of the environment. Hence, the gaze informs also which part of the environment shows the most interest.

Eye movements can be recorded in different ways and through different methods. To be best suited for our application, the chosen methods for recording eye movements need to disrupt manipulation as little as possible while reliably delivering data on the orientation of the gaze.

3.2.1 Electrooculography

One of the earliest approaches to record eye motions reliably is through the utilization of electrooculography (EOG). EOG tracks motions of the eye through the observation of the electrical signals related to the muscle contractions around the eyes. This method is often performed alongside Electroencephalography, as their setup is similar. EOG was thus also one of the earliest methods employed to quantify eye movements and to perform gaze tracking [Kaufman, Bandopadhay, and Shaviv, 1993 ; Bulling et al., 2010].

Practically, EOG is often divided into its directional components (as illustrated in Figure 3.1). Indeed, most of the time, EOG is explored as vertical EOG (VEOG), horizontal EOG (HEOG) and, in rare circumstances, radial EOG (REOG). This is done by observing the differences in voltage between two specific electrodes. For VEOG, electrode voltages from above and below one eye are considered. As eye motions are synchronous between the eyes for most users, the choice of eye for VEOG is arbitrary. For HEOG, the compared electrodes are located right of the right eye and left of the left eye. REOG is measured by contrasting the activity of an electrode placed on the forehead with either of those of the HEOG recording. While VEOG and HEOG try to quantify the vertical and horizontal movement of the eye globe respectively, REOG informs about radial movements outward from the center of eye globe. While most other electrical measurement methods require the presence of a dedicated reference and ground electrode (e.g. Electromyography or Electroencephalography), EOG only requires a ground electrode. This electrode is traditionally placed on the mastoids or clipped onto the user's earlobe [Teplan, 2002].

Thanks to these three directions allowing to quantify EOG activity, relative motions of the eye can be observed and used in different kinds of applications. While demonstrated to be possible, eyetracking relying on EOG is however difficult to perform accurately [Kaufman, Bandopadhay, and Shaviv, 1993]. The motions observed through EOG are recorded as voltage difference induced by eye muscles. For example, in HEOG recording, strong positive jumps in the signal may indicate horizontal eye movements in one direction while negative jumps indicate movements in the opposite direction. The overlaying head motions and other artifacts make it difficult to precisely follow the gaze direction on a given interface through this method.

EOG measures mainly three main categories of activity present around the electrical measurement of eye movements. The first category of activity concerns closing of the eye lid during blinks. Blinks present a strong muscle activity which, similar to what is observed in Electromyography (EMG), translates to a related increase of electrical potentials. The second category of measured activity is rotation of the eye globe. Here

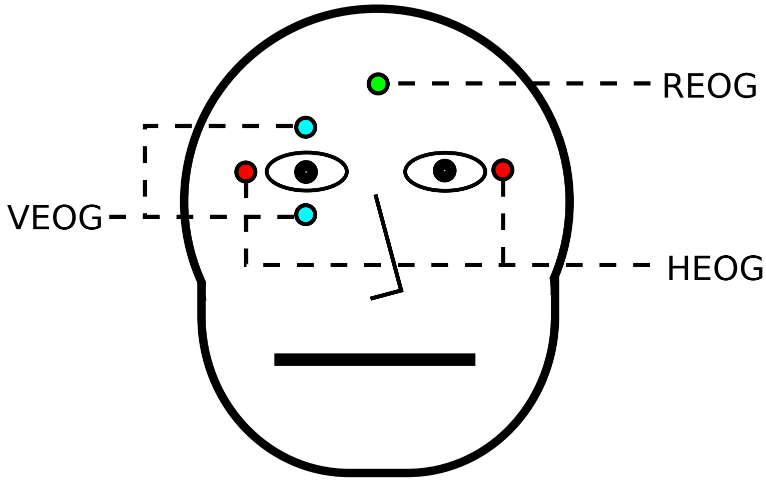


FIGURE 3.1: The standard location of Electroencephalography electrodes on a user's face. EOG is calculated by using the differences between two electrodes. (VEOG : between the blue electrodes, HEOG : between the red electrodes, REOG : between the green and one of the red HEOG electrodes.)

the activity occurs when muscles are used to shift the angular position of the eye globe. This activity is the most relevant when using EOG for gaze tracking, as it informs about angular changes. The last category of activity is observed by considering the dipole nature of the eye. Indeed, when light enters the pupil and hits the retina, the eye acts as a dipole, creating a small but measurable current [Berg and Scherg, 1991]. When employing Electroencephalography (EEG), these three categories of signals can be seen as artifacts which add onto other more relevant brain activity through volume conduction [Mennes et al., 2010]. While imprecise for gaze tracking, EOG can find its application in helping to identify and remove the presence of artifacts in other EEG channels [Croft and Barry, 2000].

The EOG approach gives detailed information about the physical movements performed by the eye. However, it provides little information about how the resulting gaze is orientated in the environment. To adequately obtain information about the viewed environment from eye movements, other techniques are more commonly used.

3.2.2 Optical Eyetracking

The most commonly applied approach for tracking a user's gaze is through the use of camera images. This approach consists of having cameras monitor the eyes, or, more precisely, the pupils of a user. Eyetrackers of this nature commonly require a calibration phase where the user is asked to fixate a series of marked points in the observed space. These points serve as ground truths for training the gaze tracking. While the exact inner workings of the algorithm change between eyetrackers, in most cases the elliptical deformation of the pupil in the camera image is used as the main feature to identify gaze location [Lai et al., 2013].

This technique is more accurate in locating the users gaze in the environment than EOG, however it is also more constrained by the optical conditions of the recording environment. To act independently from the lighting of the user's face, some devices use infrared (IR) light sources and cameras to circumvent changes in the visible light spectrum. This is desirable as the pupil reflects IR light well. This technology is also limited by the camera refresh rate and the movement of the user. The latter means that these eyetrackers need to correctly locate the user's eyes in a larger image. This can be accomplished by directly attaching the eyetracker to its user's head or asking the user to limit their head motions. In studies where the exact gaze location of the user is required, the users head is held fixated thanks to a desk-mounted chin-rest [Duchowski, 2007]. As this can be very restrictive and intrusive to the user, eyetrackers have been conceived in which the camera automatically follows the users head motions by adjusting their own camera orientation on the fly. This allows for the user to perform operations more naturally when monitoring eye movements.

In all these cases, the gaze location is extracted from the camera images for every recorded frame. The obtained information about the eye motion as such is a lot less rich than EOG in terms of physiological eye activity, but gives new information about how this motion relates to the exploration of the environment. This aspect renders this measurement of eye movement valuable. Indeed, it is one of the only approaches which gives not only information about the state of the user (e.g. the orientation of attention) but also how this state exists within the user's environment (e.g. what is looked at). Furthermore, this technique allows for the observation of the dilation of the pupil. Pupil dilation can be associated with different properties such as cognitive workload [Pomplun and Sunkara, 2003] or emotional arousal [Bradley et al., 2008]. As pupil size is also strongly influenced by the lighting conditions during the manipulation, which can change drastically between considered interfaces, this property is not considered further within this work.

The informative value for the users' environment makes optical eyetracking one of the most frequently used biosignal sensors for Human-Machine Interface evaluation [Nielsen and Pernice, 2010]. However, optical eyetracking alone offers little information about the internal state of the user. Indeed, while blinks, saccades, fixations and gaze patterns may inform about which interface elements are important to the user, they inform little about the user's specific grievances relating to them [Irwin, 2004]. Repeated and reoccurring fixations onto the same visual region are an exception to this, as indicating a region of interest (ROI) for the user. However, reliable gaze patterns do not occur in every environment and during every interaction. Therefore, other biosignals which allow to access information more reliably need to be considered. One of the most notable technique which is often displayed in frameworks used to study usability is the observation of brain activity.

With respects to the objectives of this work, optical eyetracking adds a valuable modality to general interface evaluation. Notably, eyetracking's ability to obtain information about the environment independently helps in creating a system which can assist to evaluate in a wide scope of interactions. Interactions and interfaces can significantly vary, making independently acquired reference points more valuable. Optical eyetracking is thus a major component of our system.

3.3 Brain activity

Brain activity is one of the most sought after biosignals. The brain is host to all cognitive processes. Accessing brain activity may subsequently allow to gain a glance at these processes even if only in an abstract and reduced form of recorded data. There have been many different approaches taken to obtain insights into brain activity. Figure 3.2 illustrates different methods and techniques established to extract brain activity and how these techniques perform in terms of temporal and spatial resolution. In the context of the deployment and evaluation of an interface, an approach has to be chosen which does not significantly modify the interaction while providing meaningful data to extract information about a user's perception of an interface.

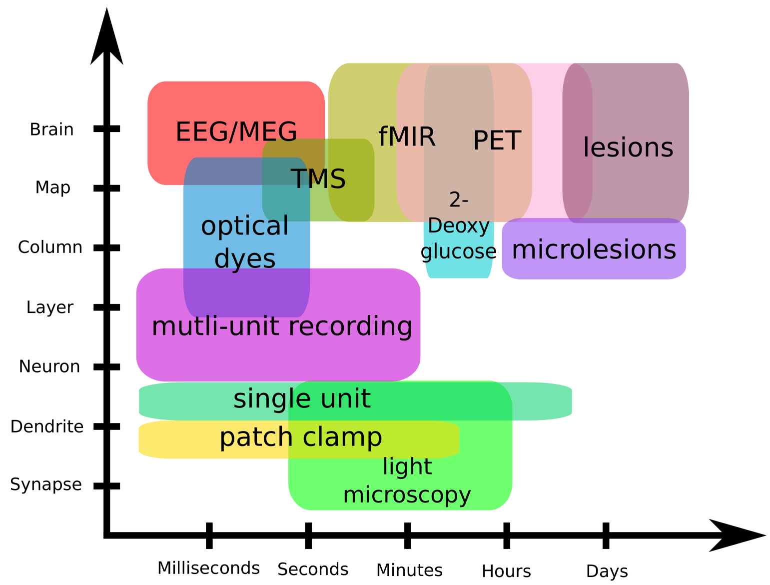


FIGURE 3.2: Spatial and temporal resolution of different methods to acquire information about brain activity and diseases. Redrawn from Cohen and Bookheimer (1994)
.

Existing methods to record brain activity rely on the observation of various physiological phenomena in and around the brain. These include the assessment of oxygenation of different parts of the brain, the recording of electrical potentials measured by electrodes put in and on the brain, or optical methods observing changes in the brain. These methods have all different advantages and drawbacks. One of the main differentiating factors is the invasiveness of these approaches. As stated previously, to be easily transposable between interfaces and users, non-invasive methods are preferable. This excludes patch clamps as well as single or multi unit (i.e. Utah arrays) requiring surgery to install. While such invasive methods record at high temporal resolution optimal for BCI [Lal et al., 2005], their application are only suited for medical patients.

Indeed, another crucial aspect to consider when accessing brain activity is the temporal and spatial resolution of the obtained data. Interaction with systems are processes that occur rapidly and can engage a wide variety of cognitive processes. While remaining in the domain of non-invasive approaches, imaging procedures offer the highest spatial resolution [Wintermark et al., 2005]. Functional Magnetic Resonance Imagery (fMRI) and Positron Emission Tomography (PET) are two imaging methods that generate a full and detailed image of the brain. Their slow temporal updating speed signifies however that fast changes in activity are difficult or sometimes even impossible to detect which is in general a crucial aspect when working with cognitive processes that occur during interaction and interface manipulation [Takeda, Okazaki, and Ushiyama, 2015]. Furthermore, PET and fMRI also require the user to lay still during recording, which is uncommon during regular interface manipulation. The neuroimaging techniques are thus mostly unsuited for our application.

Other methods, relying on the electro-physiological signal are known for their higher temporal resolution, which can be exploited to measure such changes. These acquisition methods will be inspected below to determine which of them is most suitable for the application within the context of human-machine interactions and their evaluation.

3.3.1 Recording Methods

When investigating the operation of a user interface, one of the many facets to consider is user comfort during manipulation. Comfort is an aspect which is mainly a factor for the satisfaction dimension of usability and varies between situations. When integrating additional features into the UI, even just for the sake of interface evaluation, the user performed tasks should remain the same. Indeed, additions which changes the UI operation may make it difficult to transpose the insights obtained to the usage after deployment lacking these changes. This aspect, alongside with invasiveness and recording resolutions (both temporal and spatial) are factors helping to decide which is the most appropriate recording technique. Under these constraints, there are some methods which can be considered for the evaluation of interfaces.

Magnetoencephalography (MEG) measures the magnetic fields at the surface of the scalp. It relays information about the intensity of brain activity at scalp areas during the recording. The observation of MEG is performed at a rather high spatial and, more importantly, a very high temporal resolutions [Singh, 2014]. This method has been used in many different kinds of studies, being a frequently used tool for medical applications [Hämäläinen et al., 1993]. However, this method also has a series of drawbacks. One such drawback is that only the magnetic fields measured tangentially to the user's scalp can be observed. This limits the quality and quantity of brain activity gained through this method. Indeed, due to the electromagnetic properties, brain activity occurring radially from the scalp cannot be measured through this method. More importantly, in terms of comfort, while using MEG does not require long preparation from the user, during operation they find themselves unable to freely move within the chair of the MEG. The MEG device is cumbersome and requires liquid helium and delicate sensors (superconducting quantum interference device (called SQUID)) to operate. These constraints make application of this MEG technique outside of medical applications difficult to justify, both in terms of comfort and cost. A lighter setup is required.

Another possible method to acquire brain activity data is through functional Near-Infrared Spectroscopy (fNIRS). Relying on an optical approach, this method looks at the oxygenation of the blood flowing through the scalp [Ferrari and Quaresima, 2012]. The use of near-infrared light allows to detect and record the oxygen concentration in the blood. This surface measuring of the blood-oxygen-level dependent (BOLD) response can be done using a set of small sensors which can be worn and carried around during manipulation. However, the recorded signal information only updates when new blood arrives to the brain. The temporal resolution is thus dependent on the heartrate, making it impossible to reliably record fast changes in brain activity. Yet, such fast changes are essential when gathering insight in the working of Human-Machine Interactions. Methods which are not limited by slower biological processes would be preferable in this respect.

Some of the highest quality of brain activity data are obtained by measuring the electrical activity of the brain. This is the case for Electrocorticography (ECoG) which relies on placing an array of electrodes on the surface of the brain [Keene, Whiting, and Ventureyra, 2000]. This approach has shown to provide data of high quality allowing to control many different systems [Leuthardt et al., 2006]. The method is less invasive than similar direct brain measurement methods, however it also requires surgery for its implementation. ECoG is thus a recording method reserved for patients and hardly applicable to a broader scope of users. However, similar recordings of brain activity can be done non-invasively through Electroencephalography (EEG), which provides the best compromise for deployment during HMI.

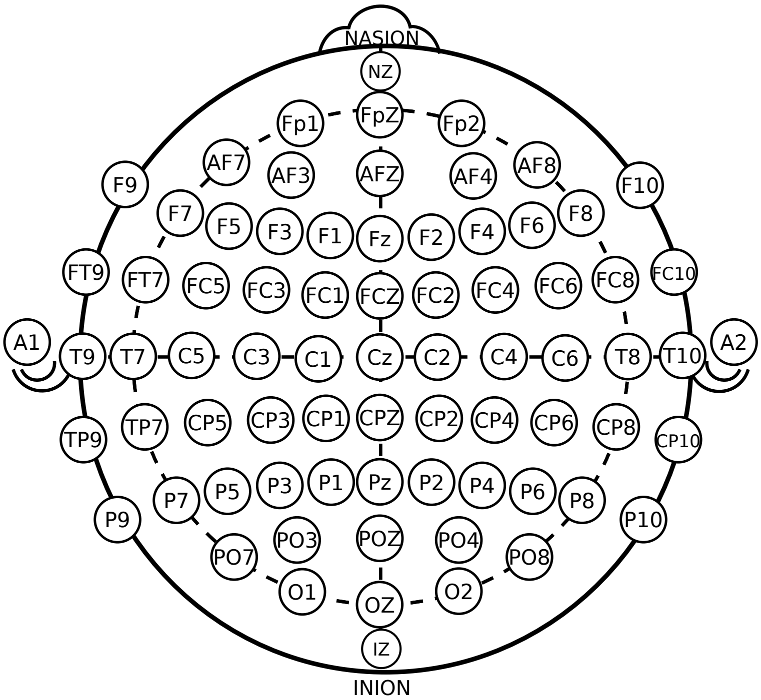


FIGURE 3.3: Mapping of the electrode locations most commonly used in the 10/20 System of Electroencephalography.

3.3.2 Electroencephalography

Electroencephalography (EEG) was invented in 1924 by Hans Berger [Tudor, Tudor, and Tudor, 2005]. It constitutes one of the most used and most accessible way to obtain data on brain activity. EEG recordings rely on a set of electrodes which are positioned on the scalp of its user. The number of electrodes as well as their position can be easily modified depending on the application. EEG offers flexibility and non-invasiveness.

Practically, EEG involves putting a tight EEG cap on the user's scalp. Electrodes are attached onto this cap in accordance with the so-called 10/20 system (see Figure 3.3). This system was created to propose a unified way to place electrodes to the scalp, changing electrode locations according to known angular changes in relation to the central Cz electrode placed equidistantly between the nasion and the inion [Acharya et al., 2016]. In addition to all the electrodes providing information about their recording location, two additional electrodes (marked as A1 and A2 in Figure 3.3) serve as ground and reference electrodes completing the electrical circuit needed for recording. These two last electrodes are often positioned on the ears (via clamps) or on the mastoids. To improve contact, often conductive paste is applied to decrease the impedance on the scalp surface. The data recorded through the electrodes are then relayed and filtered by a dedicated signal amplifier. The amplifier then digitizes the analog signal to make it accessible for numerical analysis. Data acquisition can be performed with temporal resolution rates ranging up to 2000 Hz. Most commonly, recordings are performed between 128 to 512 Hz.

The temporal resolution is thus quite high but the spatial resolution is limited to the selected electrodes. Traditionally, between 16 and 64 electrodes are monitored during EEG studies. As these recordings are performed on the scalp, volume conduction induces strong correlation between adjacent electrodes. Volume conduction is the dissipation of electrical current across neighboring tissue and adds artifacts and signal overlaps to nearby sources. These include, as already detailed above, different kinds of eye artifacts as well as other muscle movement occurring throughout the body, or general electromagnetic noise coming from the electrodes potentially jittering during the recordings. EEG, in contrast to MEG, records both tangential and radial activity at the surface of the scalp, and thus obtains more information about the underlying neural activity [Lew et al., 2013]. Given all these different sources of noise in the recorded data, a consequent amount of analysis and feature extraction has to be performed to obtain useful information from the raw EEG data.

EEG is the main method for acquiring brain activity data that is used in Brain machine interfaces (BCI). Analysis of EEG data has shown to provide new insights into the state of the users. This includes information about the state of arousal of users [Bonnet and Arand, 2007], as well as a series of other information detailing current aspects of their state of being, like emotions [Bos, 2006]. It is easy to setup and to apply in various situations, as well as its market accessibility, makes it a useful tool for helping patients and regular users in a wide variety of tasks. EEG technology has also been extended to Human Machine Interactions. While the method still has some downsides, such as electrode cables potentially hindering movement and the cap reducing user comfort, EEG is the most suitable recording method of brain activity for our purpose.

3.4 Using Biosignals for Applications

Three methods to record useful biosignals have been so far been discussed in detail above. These namely are electrooculography, eyetracking and electroencephalography. To properly integrate them into the design cycle of Human-Machine Interactions to aid with HMI evaluation, these methods have to be incorporated in a way which does not require incumbent changes of the underlying interaction. They also should provide information which is relevant to understanding the perception of the observed Human-Machine interaction. To investigate how these methods can be combined to reach these goals, inspiration can be gained from the applications which have already used biosignals.

3.4.1 General Use of Biosignals

While still rarely used, biosignals have already seen different applications. As stated previously, the first applications of biosignals have been used in diagnosis, gauging the health of subjects. EEG, for example, is often used as a starting point to identify neurological diseases, like epilepsy [Smith, 2005] or to check on the state of consciousness of a patient [Gosseries et al., 2011]. For Eyetracking, medical use is more experimental. It has been employed to quantify the severity of eye movement disorders or, sometimes, in medical cases of concussions [Samadani et al., 2015]. The extracted information through either method is usually integrated into diagnosis and serves to identify the proper treatment of the patient.

Simpler versions of diagnosis have also been implemented in consumer-grade applications. While Eyetracking and EEG are generally not implicated, the application of biosignals is gaining popularity in the sport and fitness domain. Here, heartrate and breathing rhythms are quantified to assess the physical effort and performance [Paul and Garg, 2012]. In these applications, obtaining the biosignal is the goal of the interaction with a system, meaning the main interaction is not impeded by the measurement of biosignals. Most of the interpretation and integration of this information is left to the users to evaluate on their own.

Another type of applications concerns systems which are, at the moment, considered toys or novelty products. These can include applications relying on EEG, in which this biosignal is still the main selling point of the product but is nonetheless used to solve a separate problem. This can be in the context of a game [Ewing, Fairclough, and Gilleade, 2016] or of at-home neurofeedback training for meditation or sleep quality analysis [Krystal and Edinger, 2008]. Still other experimental applications have been envisioned where biosignals are used in art creation [Nagashima, 2003].

Considering most of these applications of biosignals, one remarks that their scope is rather limited. Indeed, biosignal application are either restricted to very specific sets of situations or are used in a dedicated clinical setting. Whereby both situations are not commonly situations by most users. There exists, however, applications that use biosignals in more direct relation to the usability of interfaces.

3.4.2 Application to HMI Improvement

In terms of usability during Human-Machine Interaction, certain biosignals have already found their application. As developed in the previous chapter, HMI can be evaluated through the concept of Usability, which can be split into three main different dimensions : *effectiveness*, *efficiency* and *satisfaction*. Biosignals can be used to enhance any of these three dimensions.

The first dimension, effectiveness, has been integrated into HMI with the creation of Brain-Computer Interfaces (BCI). The first goal of Brain-Machine Interfaces (BMI) was to enable interactions with systems without the necessity for the use of the peripheral nervous system (e.g. limb movements or speech). Robot prosthesis or eyetracking keyboards are prominent examples of this. By implementing such BMI, interfaces become accessible to patients with limited motor functions, which increases the overall effectiveness of the system as well as its interface [Muller-Putz and Pfurtscheller, 2007]. For most applications, effectiveness and proper operability is however already guaranteed for healthy users. When looking for useful information about the interaction, optimization of other dimensions is more desirable to improve their usability.

The study and improvement of efficiency and satisfaction are traditionally confined to the design phase of a product's life cycle. The observation of biosignals has been integrated as added value in the design cycle. Particularly the domain of digital interfaces, such as web-pages, makes use of biosignals. To learn about the quality of the layout of options presented by the interface, eyetracking is often used. Observations of gaze patterns help to locate graphical elements which need improvement [Nielsen and Pernice, 2010]. Similarly, applications of EEG recordings have also been integrated into the design cycle. These signals offer interesting information about general feelings, the overall attention or arousal of the user [Wobrock et al., 2015]. The use of EEG remains much more experimental. However, in all of these cases, the scope of application of or information provided by biosignals is rather limited.

Improvement of usability via biosignals has also gone beyond the design phase. Indeed, as with sport and fitness devices, biosignals can be measured and used to improve interaction during the actual application. Even biosignals such as brain activity have made their way into the improvement of HMI during final usage by the consumer [Kuikkaniemi et al., 2010]. This is done by either adjusting the interaction as a whole or by adapting the accompanying interface [Ewing, Fairclough, and Gilleade, 2016]. Different meditation smartphone applications have notably been conceived, measuring brain activity to estimate the users anxiety. Brain activity has also been used to create adaptive games which change their demand from the players depending on the current difficulties they encounter [Millán, 2003]. All these applications however remain rather superficial in their usage of biosignals and, with a few exceptions, offer little but consumer novelty to human-machine interaction.

As noted above, the properties extracted from the rich EEG data can give new information on the interaction, while Eyetracking provides a way to flexibly associate the users biosignals to the world they are interacting with. The combination of these two biosignals can thus serve as a basis from which interaction may be effectively analyzed and improved. One approach would consist of creating a Brain-Computer Interface to analyze the current shortcomings of interaction. In order to create this type of BCI, it is necessary to understand how both of these recording methods can be combined to provide valuable information.

Chapter 4

Creating A Brain-Computer Interface

4.1 Defining The BCI's Goals

The use of Brain-Machine Interfaces (BMIs) began with the inception of ElectroEncephaloGraphy (EEG). BMIs were the first applications to utilize the recording of brain activity [Tudor, Tudor, and Tudor, 2005] for practical purposes. The term BMI refers to any interface which employs recorded brain activity to alter an interaction. From this definition, the term "Brain-Computer Interface" (BCI) can cover many different types of applications. In this work, we use the terms BCI and BMI interchangeably.

One possible way to group these different types of BMI applications is through the categories of *active* and *passive* BMI [Tan and Nijholt, 2010]. These adjectives denote how brain activity is employed in an interaction. In active BMI, brain signals serve as the main way of communication between the user and the system. The user is explicitly prompted to focus their attention and their thoughts onto specific interface elements. For example, this can consist of imagining motions [Coyle et al., 2011] or focusing on the pulses of a flashing UI element [Chen et al., 2014]. Brain activity relative to these prompts is then acquired and analyzed to determine which interface option the user aimed to select. A common example of an active BCI is the P300 Speller which allows users to select an element from a flashing matrix. It is often used to emulate a computer keyboard [Farwell and Donchin, 1988].

In passive BMI, brain activity is gathered and analyzed without any particular prompt or focus demanded from the user. Here, users are traditionally asked to perform the same series of actions they would normally do in a given interaction featuring no BMI components. Brain activity is gathered continuously and analyzed in the background. It serves to adapt the interface and to shape the interaction in different ways. Passive BMIs generally estimate the cognitive state of the user and give feedback to the user or the system accordingly [Zander and Kothe, 2011]. This feedback is then often used to modify the interface or the manipulation to improve the whole interaction. For example, if an interface presents elements in an unappealing way and reduced attention is detected from the user's brain activity, the interface can change itself to present elements in a more appropriate way [Acı, Kaya, and Mishchenko, 2019].

Both active and passive BMI are useful in specific applications [Lotte et al., 2007] and have their own specific influences on interface usability. Active BMIs deeply change the interaction paradigm of an interface. As stated in the previous chapter, active Brain Computer Interface can extend interface effectiveness and usability to patients with motor difficulties [Guger et al., 1999]. Passive interfaces, on the other hand, can serve as a way to monitor and adapt interfaces discretely during the manipulation. Passive BCIs do not alter the core interaction paradigm inherently. They allow for a potential improvement of satisfaction and efficiency of the manipulation. This second kind of BCI is therefore of greater interest when pursuing our set goal of improving HMI or, more specifically, to evaluate Human-Machine Interaction.

Considering the stated objective of this work, the implementation of a passive BMI is pursued. This BMI aims to help with the evaluation of interfaces, complementing the classical design cycles. To do so, the BMI needs to provide information helping with the assessment of the considered interfaces usability. More specifically, it should provide information about the user's perception of interface efficiency and satisfaction. As detailed in Chapter 3, the modalities of EOG, EEG and optical eyetracking were chosen to be used within this BMI system. Comparable BMIs have already been used in various controlled laboratory settings [Dimigen et al., 2011]. There, they helped to gain new information about the interaction, notably relating to the user's attention. However, they are scarcely explored in broader and more natural contexts, such as those encountered in HMI evaluation. The validity of using BMI in such natural contexts first needs to be verified.

Practically, to verify the validity of this type of setup for HMI evaluation, two challenges have to be met. Firstly, the data recorded from the modalities need to be analyzed to gain information which is relevant and useful to understand the user's perception of the interaction. For this, a series of methods for the preparation and analysis of the data is required. Secondly, HMI evaluation can apply to radically different types of interaction (e.g. from car park assistance, to web-page navigation, to vending machine operation). Accordingly, both the BMI setup and the series of methods used to extract results should to be conceived to be applicable to all these scenarios. In summary, the multi-modal BMI setup thus needs to be easily transposable and should give access to properties which help to understand any interaction.

Valuable insights into interaction are commonly obtained by observing brain signals at the relevant moments and analyzing them in relation to the interaction context present during these moments. Different techniques and algorithms have been proposed for this purpose. The addition of eyetracking provides a way to discretize moments of interaction in an easily transposable fashion (as eye movements are always present). One way to achieve this discretization is by focusing on fixation onsets. When contrasting fixation onsets with brain activity, Event-Related Potentials (ERPs) can be observed which hold information about the interaction. These EEG potentials and their utility will be discussed below. Prior to doing so, the transposability of the BMI setup itself will be reassessed first to ensure the adequate acquisition of these signals without disturbing the interaction.

4.2 BCI Transposability

Both active and passive Brain-Computer Interfaces are commonly implemented into an already existing system. This implementation process requires the modification of the initial system which underpinned the interaction. The process is illustrated by the dotted central rectangle in Figure 4.1. Brain activity is collected from the user through a continuous stream. For our BCI, this stream is obtained via EEG. To extract targeted information about the interaction, brain activity needs to be examined at specific moments in time. Information about the context of the interaction is thus necessary for a meaningful segmentation (and subsequent analysis) of the brain activity data stream. Practically, this context information is commonly provided by the system/machine part of the HMI. For example, if brain activity is examined during button presses, the system provides the timestamp and button code for each press. This is denoted by the red lines labeled "1" in Figure 4.1. However, if the BMI is grafted onto an existing system, adaptations often need to be made to the system so that such specific events are logged and relayed [Rugg and Coles, 1995]. These adaptations commonly take two forms.

The first and more difficult type of adaptation pertains to changes in the functioning of the interface in order to accommodate BMI paradigms. This is frequently encountered in active BMI where such changes are necessary to induce relevant types of brain activity needed for direct interaction with the interface. This is notably the case when working with speller interfaces, such as the P300 speller [Farwell and Donchin, 1988]. The P300 Speller requires letters to be arranged differently and to be flashing at random intervals. Both of these aspects imply significant changes to the traditional digital keyboard. The second type of adaptation is also important but more subtle. It consists of implementing software routines which log and relay context information from the base interaction system to the BMI for biosignal data segmentation. Such information may oftentimes already be logged within the software of the base system. Making it accessible to external devices, such as a BMI, however requires changes in hardware and software, which can often be quite extensive.

In the case of an active BCI, both types of adaptations are required. The first type of adaptations is even a key feature of active BMI. When considering passive BMI, especially for interaction evaluation, which is the case here, only the second type of adaptation is commonly implemented. For the user, the interaction can even seem unchanged. However, such adaptation may still require profound changes and thus impede the application of the passive BMI to a large scope of interfaces. When considering the transposability of our BMI, the adaptation workload required for its use in new interfaces also needs to be considered.

As a workaround to limit adaptation workload, BCI frameworks have been conceived to manage multiple systems at once. In these instances, context information is provided by a user-operated central hub interface or through an external monitoring system surveilling the entire interaction environment [Navarro et al., 2011]. While well suited for patients with motor difficulties who are limited to controlled environments, such frameworks are costly to implement, require additional user training, and remain laborious to expand to new situations. Alternatively, events which provide contextual information about the interaction can be gathered through biosignals rather than the system. This is denoted by the green lines labeled "2" in Figure 4.1. The utilization of eye movement recordings notably allows to indirectly obtain information about important regions of the interface as well as relevant moments during interaction.

The transposability of a BCI between interactions will be examined. For transposability to be optimal, a balance between two aspects need to be found. The first aspect is, as mentioned above, the minimization of adaptation workload. The second aspect concerns the physical ease of integrating the setup into existing interfaces. Indeed, while system adaptation should be minimal, the BCI itself should not become a bigger hurdle to interface evaluation than the interface itself.

4.2.1 Minimizing System Adaptations

Brain-Computer Interfaces collect and relay information both from and to the human and the machine in HMI (see Figure 4.1). From this perspective, BCI is built onto two basal components : A user-side component, which gathers information related to the user, and a system-side component, which gathers information related to the system [Graimann, Allison, and Pfurtscheller, 2010]. The exact role of both components changes depending on the kind of interaction which is observed. The classical approach discussed so far, represented by case 1 in Figure 4.1, combines the user-side gathering biosignals and the system-side gathering context information. By exploiting data provided by both components, biosignals are studied at relevant moments. Practically, this gives us the possibility to study Event-Related Potentials [Rugg and Coles, 1995]. A Steady State Visually-evoked Potential (SSVeP) speller notably works this way by making the interface elements flash and relaying to the BCI which and when each element flashes [Müller-Putz et al., 2005].

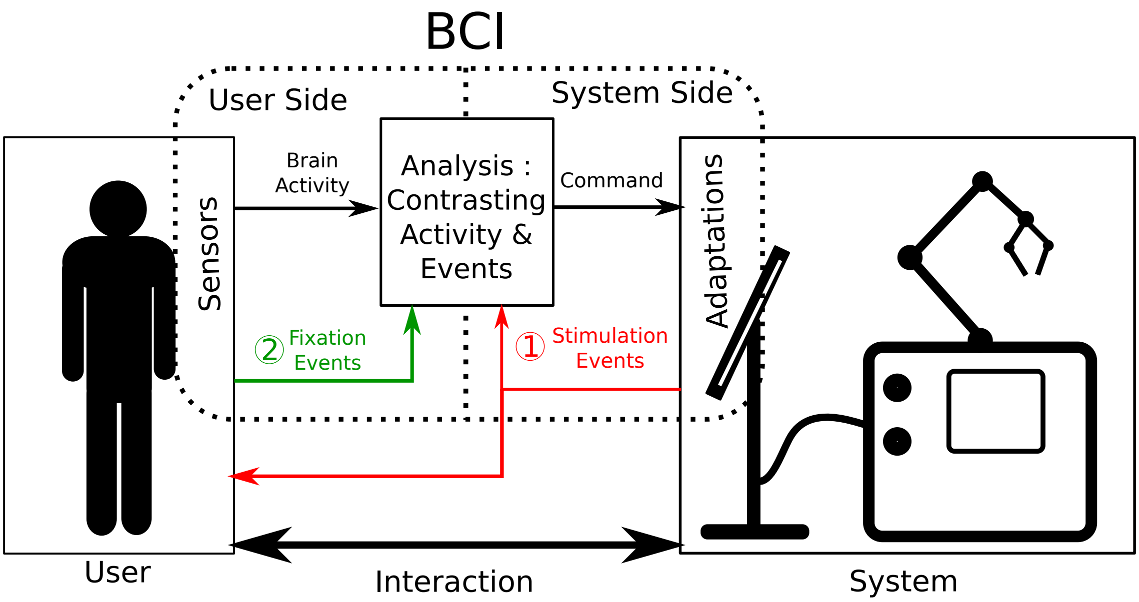


FIGURE 4.1: Functioning of Brain-Computer Interfaces during interaction. Sensors are placed onto the user and adaptations are made to the system to augment interaction. The brain activity of the user is contrasted and analyzed as it relates to events occurring during interaction. Classically (case 1 in red) these events are produced by system adaptations. In this work, we investigate the acquisition of events via sensors such as eye tracking (case 2 in green).

Generally speaking, context information are punctual events defined by a timestamp and a label produced by the system. To reduce adaptation workload, similar information needs to be obtained from biosignals. This is more difficult due to biosignals' continuous nature. In other studies, informational value extracted from the brain activity has been augmented using data from Electromyography (EMG, electrical monitoring of muscle contraction) as well as Galvanic Skin Response (GSR, monitoring of skin conductivity) [Müller-Putz et al., 2015]. These modalities have shown to reliably improve specific interactions [Nijholt, Allison, and Jacob, 2011] and can give clues about the current levels of user arousal or anxiety and further the understanding of the user's current state. However, both of these modalities do not provide the type of discrete context information enabling the observation of ERPs. Accordingly, the results of their combination with brain activity are more difficult to utilize for interface evaluation. A biosignal providing more momentary and precise information is required.

In this work, eyetracking (ET) is utilized to obtain this context-information. As already detailed in the previous chapter, eyetracking itself has already frequently been employed to assess and describe interaction in a wide scope of interfaces [Goldberg and Kotval, 1999] (e.g. for web-pages). Eye movements have been shown to be influenced by user intention and the general task they pursue [Hayhoe and Ballard, 2005]. Their application, while strongly dependent on the use case, can be exploited to estimate certain specificities of the interaction [Henderson, Weeks Jr, and Hollingworth, 1999], giving information about the user (e.g. the speed of interface exploration), as well as their surroundings (e.g. which regions are most explored). As discussed in Chapter 3, eye movements can be decomposed into four discrete events (i.e. fixations, saccades, blinks and smooth pursuits) which help to segment and categorize different moments of importance during manipulation. These events provide context information which is momentary and rather precise.

Relying on ET context information allows to minimize the allocated workload for a further system adaptation (as represented by case 2 in Figure 4.1). Through this mode of operation, the proposed BCI can quickly and easily be applied in new interaction contexts. However, eyetracking information is less precise than system-side context information. Consequently, a set of methods will be needed to analyze the resulting data efficiently. Before doing so however, transposability also needs to be examined in regards to the physical setup of the proposed BCI and how it impacts the interaction. The comfort of each sensor was verified in the previous chapter but their combination and practical application in interaction will be addressed below.

4.2.2 Comfort of our BCI

From a physical point of view, combining Electroencephalography, Electrooculography and Optical Eyetracking is a simple and unconstraining procedure for the user for most applications. Indeed, EEG and EOG typically share the same set of electrodes. In terms of the comfort during recording, an EEG cap requires to fixate electrodes at the correct positions. The cables of the electrodes can, in specific circumstances, hinder motions. Careful placement of these cables does however allow for unhindered and even mobile applications of EEG [De Vos et al., 2014]. Eyetracking, on the other hand, is more difficult to implement without changing user perception at all. The eyetracking sensors (i.e. cameras) require the monitoring of the user's eyes. This means that users will have to adjust their motions to the eyetracker to a certain extent. While positioning the user's head on a chin rest may yield the most accurate tracking, it severely restricts head motions which naturally occur during interaction. Remote eyetracker and head-mounted eyetrackers are more appropriate for our application but still change interaction slightly. In these cases, the user has to act within the constraints of these devices. This means that their liberty of freely moving their body is restricted, to accommodate for the visual range of the cameras or that the user has to view the interface through an eyetracking headset respectively. Both approaches do not modify interaction significantly, but can induce small differences in operation. Despite resulting artifacts, both setups conserve the natural properties and peculiarities of the interaction. Using them in the evaluation process does still allow the resulting interfaces to be applicable in most common situations.

Furthermore, while only passive BMI are explored in this work, the comfort of this setup could be even beneficial to active BCI. Traditionally, attention in active BMI is divided between two parts of the manipulated system. This is represented in the upper scenario of Figure 4.2, where attention is marked by the red arrows. The interface present context information (the green arrows in Figure 4.2) and the operative system performs the action (e.g. in Figure 4.2, this could be the movement of a robot arm). This technique is used in speller applications such as the P300 [Farwell and Donchin, 1988] or Steady State Visually-evoked Potentials spellers [Cecotti, 2010]. When operating such interfaces, the user's attention is first directed towards a digital keyboard allowing for the selection of letters and, subsequently, to the text field in order to verify the correct letter was chosen. By using eyetracking, the necessary context information theoretically is provided by the user-side information. This can be seen in the bottom scenario of Figure 4.2. Consequently, BMI operation does not require an additional selection interface, reducing the

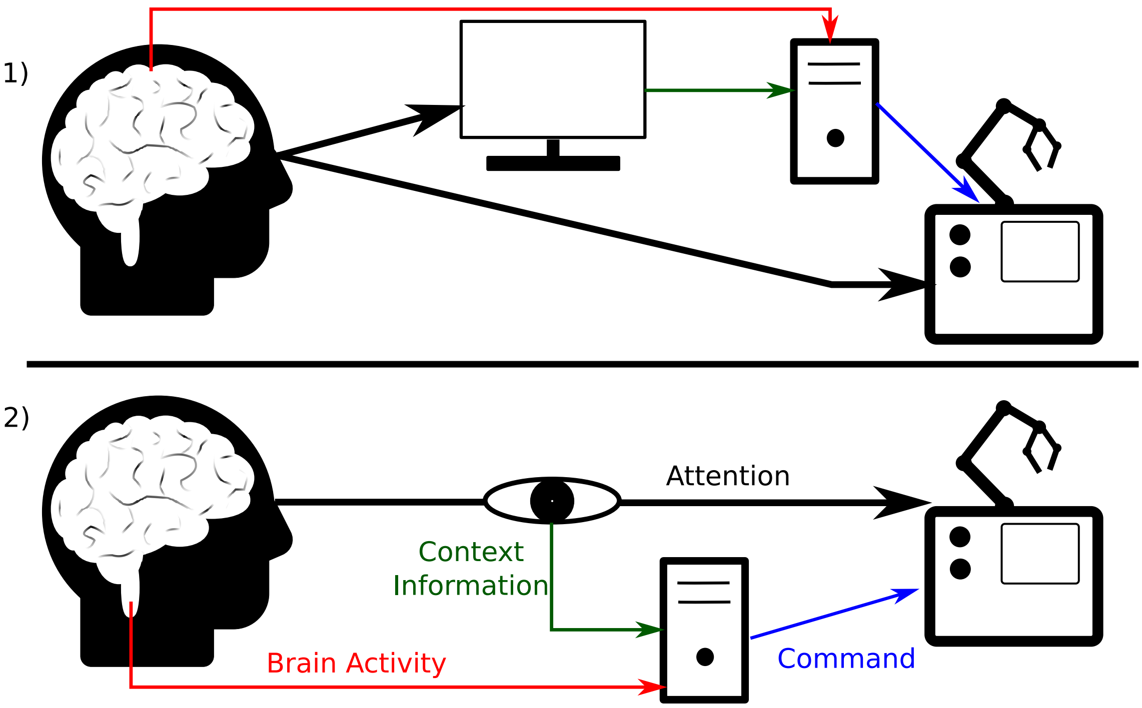


FIGURE 4.2: Information pathways to operate a BCI. In upper case 1, a classical BCI relies on the focalization of the user both on the operated system as well as on an interface used to operate the system. The user's attention is split between both elements. In lower case 2, rather than gaining contextual information from a secondary interface, it is directly extracted from eye movements. Thus, the user's attention is not split.

system-side adaptation requirements and allowing the user to fully focus their attention onto the main interaction. This dimension is not further explored in the present work, as it is not relevant to the stated goal of interface evaluation. But it offers an interesting feature to the portability of the proposed approach to active systems.

These different properties show that the proposed bi-modal (i.e. using brain activity and eyetracking) BMI is quickly applicable and transposable to many different applications. However, to align with the proposed goals, valuable properties must be extracted from biosignals. A series of methods is thus needed to process the collected data.

4.3 Accessing Relevant Information

As mentioned above, the context information provided by eye gaze motion is less informative than classical system-side event logging. Most notably, the labeling of the events does not benefit from the same precision as system events as it is limited to the observed eye movement type (i.e. fixation, saccade, blink ...) and its location. This makes the extraction of informative insights relating to interaction more difficult. Consequently, when combining and contrasting brain activity and eye movement recordings, the resulting data are more difficult to analyze properly in relation to the interaction.

To obtain valuable information about the interaction, methods for the analysis of the gathered data will be proposed. Data from the modalities can be evaluated separately or in a combined way to obtain insights. Before discussing these methods, however, the transposability of the setup needs to be addressed one last time. Since the BMI setup aims to be applicable in many different interaction scenarios, the obtained data are prone to contain a variety of artifacts relative to the interaction circumstances (e.g. user muscle artifacts, sensor tracking difficulties, ...). To make the data usable in all circumstances, an extensive preprocessing of the data must be performed to eliminate or minimize artifacts. Thereafter, the effective extraction of information can be addressed.

4.3.1 Preprocessing

For the initial preprocessing on the data, ET is considered separately from the EEG and EOG data. This is due to the different nature and sensor being used to collect this data. The preprocessing begins with the analysis of EEG and EOG recordings.

As detailed in the previous chapter, EEG is measured through a set of electrodes. These electrodes are placed at different locations on the subject's scalp. Electrodes are connected to an amplifier, which is in turn powered by the main power supply. Accordingly, when no hardware correction is performed, a residual artifactual oscillation is often present at 50 or 60 Hertz (Hz) depending on country voltage standards. More noticeably, recordings also often contain a low frequency voltage drift at around 0-0.1 Hz. This drift originates from biological and electrical processes (e.g. sweating) occurring at the contact points between electrodes and the user's skin [Luck, 2014]. Both high and low frequency artifacts are most often compensated by using appropriate high-pass and notch (i.e. bandcut) Butterworth filters. After these corrections, still other types of noise can persist which may occlude informative properties contained within the data.

Muscle activity and eye movement activity occurs close to the considered locations of the EEG electrodes and will, through volume conduction, be picked up by these electrodes. To compensate for this additive noise, an Independent Component Analysis (ICA) is often used [Mennes et al., 2010]. The effectiveness of this method in removing artifacts is dependent on the quantity of electrodes during EEG. An ICA performs a basis transformation of the data, changing recorded samples from a sensor space into a source space. In our case, this means that if each EEG data-point is defined through N values corresponding to the voltage measured at N electrodes (noted $\overrightarrow{\mathbf{x_e}}$), the ICA transforms these N values into N new values (noted $\overrightarrow{\mathbf{x_s}}$). This is formalized in Equation 4.1, where **A** is the transformation matrix between these spaces. The ICA creates the matrix A by minimizing the kurtosis between dimensions. Through this step, the ICA aims to establish a matrix which dimensions are statistically independant from each other.

This way, the ICA serves as a data driven method separating the data into statistically independent sources of brain activity. Each source (i.e. dimension of $\overrightarrow{\mathbf{x_s}}$) is analyzed and labeled as valid or artifactual. Typically, this labeling is done through expertise. By examining the averaged activity of an ICA source, spatially across electrodes and temporally, artifactual components can be identified. For example, a source dimension presenting unusually strong frontal activity may be identified as a source of eye movement artifacts. There are also methods which autonomously identify artifactual components such as MARA (Multiple Artifact Rejection Algorithm) [Winkler, Haufe, and Tangermann, 2011]. Other autonomous analysis methods also rely on simultaneous recordings of other modalities, such as eyetracking [Dimigen et al., 2011], to identify precise artifactual sources. Once artifactual sources have been detected, these dimensions are rejected (i.e. removed) from the basis change matrix. Practically, the different source dimensions of $\overrightarrow{\mathbf{x_s}}$ are multiplied by 0 if they are artifactual and with 1 if not, as illustrated in Equation 4.2.

After this step, the ICA is calibrated to perform the artifact removal. Each new sample of EEG data can be corrected by projecting it into the source space via Equation 4.1, rejecting the artifactual sources and projecting it back into the sensor space via multiplication with **B**, the inverse of matrix **A**, as seen in Equation 4.3. The resulting data (here denoted by $\overrightarrow{\mathbf{x_{ec}}}$) are considered free from main sources of artifacts [Rao, 2013].

$$\overrightarrow{\mathbf{x_s}} = \mathbf{A}\overrightarrow{\mathbf{x_e}}, \overrightarrow{\mathbf{x_e}} \in \mathbb{R}^N, \mathbf{A} \in \mathbb{R}^{N\times N} \tag{4.1}$$

$$\mathbf{B} = \mathbf{A^{-1}} \begin{pmatrix} a_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & a_N \end{pmatrix}, \begin{cases} a_i = 0 \text{ if component i is artifactual} \\ a_i = 1 \text{ if component i is valid} \end{cases} \tag{4.2}$$

$$\overrightarrow{\mathbf{x_{ec}}} = \mathbf{B}\overrightarrow{x_s} \tag{4.3}$$

For the preprocessing of the optical eyetracking data, artifacts also need to be accounted for. Most of artifact removal steps are addressed using calibration mentioned in Chapter 3. By means of the calibration phases done prior to the actual eyetracking recordings, pupil deformations in the camera are associated with locations in a given referential. Other locations are then extrapolated from this initial association. Depending on the device, the referential is either the output image of a scene camera or a considered computer screen. The mapping of the gaze to the referential occurs for each pupil separately. Consequently, to adequately work with the user's gaze, the recordings from both eyes are either averaged to one value or only the position of the dominant eye is used [Holmqvist et al., 2011].

Each camera frame returns one pupil ellipsis and thus one location in the reference frame. To collect relevant data, eyetracking has to be performed over time to identify gaze patterns. However during this time, the subject can change position, changing the location of the eyes in the camera frame. This effects pupil deformations seen by the camera and thus the associated gaze location. A series of calibration methods can be performed to compensate for these errors. Specifically, an initial calibration is required to locate the user's gaze within the reference frame. Over time, drifts of eye gaze location are observed, shifting the perceived gaze location away from its effective location [Crossland and Rubin, 2002]. For this purpose, a full recalibration (i.e. performing the initial calibration anew) or drift corrections can be performed in regular intervals, eliminating these artifacts.

From these continuous gaze locations, the four types of eye movements can be extracted. Most notably fixations can be calculated. Practically, a fixation is determined when the gaze location is kept within a restricted visual region for longer than 100 milliseconds (ms). This region is called the foveal region and corresponds to the visual area in which objects are perceived clearly by the user [Holmqvist et al., 2011]. This region covers a 5-degree circular arc radius around the center of the pupil. A fixation ends once the gaze makes a leap outside of the foveal region. If this leap is abrupt, it's counted as a saccade. If performed slowly, the movement is counted as a smooth pursuit. The transformation of the raw input into these metrics also serves as an artifact removal method where quick gaze jitters in and out the same foveal region are rejected as noise from analysis. The extraction of a fixation from gaze locations is further described in the Algorithm 1. After extracting all these types of eye movements, gaze patterns can be analyzed.

Through both methods of preprocessing the EEG and Eyetracking signal can be cleared of most artifacts. Unless the user has been unusually active during the manipulation, these preprocessing steps sufficiently remove artifacts [Rao, 2013, Holmqvist et al., 2011]. With the cleaned data, methods for the extraction of information relevant for interface evaluation can be developed.

Algorithm 1 Extracting fixations from gaze locations

$X \leftarrow []$ ▷ X : Empty Gaze Location array
$t_{thres} \leftarrow \frac{SR_{ET}}{10}$ ▷ SR_{ET} : Eyetracker sampling rate
$d_{thres} \leftarrow D_{ET} \times \arctan 5°$ ▷ D_{ET} : Distance between eye and referential
while Eyetracker is running **do**
  Obtain new gaze location x with timestamp t
  $d = \sqrt{(x - X.mean())^2}$
  if $d \leqslant d_{thres}$ **then**
    X.insert(x)
    if $X.size() > 0$ **then**
      $t_{start} = t$
    end if
  else
    if $X.size() \geqslant t_{thres}$ **then** ▷ If gaze remained long enough
      $x_{loc} \leftarrow X.mean()$
      $t_{end} \leftarrow t$
      Return $x_{loc}, t_{start}, t_{end}$ ▷ Location, start and end times of a fixation
    end if
    $X \leftarrow []$ ▷ Emptying X of all values
  end if
end while

4.3.2 Extracting Information

The proposed BMI setup aims to extract valuable information about a studied interaction. What constitutes valuable information about an interface can take many different forms depending on the interface, the user and the pursued task. When working with our chosen modalities (i.e. ET and EEG), the information which will be extracted from individual modality recordings can be considered first.

We begin by looking at the information contained within the eyetracking recordings. This modality serves, as stated previously, to gauge the orientation of the user's attention in the explored space and thus give contextual information. This is done by utilizing fixations as both saccades and blinks leave the user blind to the environment [Simons and Levin, 1997]. Smooth pursuits do not often occur in interactions, unless smoothly moving elements are present. By focusing on fixations, information such as the duration and number of fixations can be considered. If the scene camera is properly calibrated, fixations are associate with the environment. By looking at fixation locations, regions of interest (ROIs) where more or longer fixations have performed, can be identified within scenes [Dupont, Antrop, and Van Eetvelde, 2014]. Even without knowing the scene layout, these ROIs can create a visual heatmap giving information about interface layout with any system adaptations. While inferences can be made about the interface due to ROIs, these deductions are strongly dependent on the context of the interaction. To enhance the informational value of ROIs, they have to be considered in relation to other gathered information such as the EEG data.

Before doing so however, the informational value of EEG recordings can be explored on its own. EEG data are more complex than ET data. Consequently, they can be analyzed in different ways. When considering the continuous EEG signal recorded over the full duration of the interface manipulation, one of the simplest analyses consists in looking at the frequency spectrum of the signal. Most of the EEG activity recorded from healthy patients takes places within the frequency range of 1 to 50 Hz. This range has been traditionally divided into a series of frequency bands. These bands are commonly referred to as the delta, theta, alpha, mu, beta and gamma bands. Each of these bands respectively contains higher frequency ranges (e.g. the delta band covers 1 to 4 Hz while the gamma band covers 35 to 80 Hz). When aiming to extract information about an interaction, the signal powers of these frequency bands can be compared to each other. A proportional increase of a higher frequency bandpower (such as beta or gamma bands) indicates a higher state of arousal or anxiety. Similarly, more power in lower frequencies (such as alpha or theta) suggests a more relaxed state of the user [Stenberg, 1992]. This property of EEG frequency bands is most notably used in neurofeedback training [Mandryk et al., 2013] where users can use it as a metric to track their state of mind.

Regarding EEG frequencies, another technique to obtain information about the user mental state is possible through connectivity. Connectivity involves the statistical influence recordings from one electrode have onto the recordings gained from other electrodes. While being a promising technique revealing rich information, connectivity requires computationally expensive algorithms which hardly keep up with real time, as well as a greater number of recording electrodes. Connectivity will not be considered within this work.

Beyond frequencies and connectivity, the EEG signal can also be observed as a time series. However, time series are difficult to manipulate or compare between one another as they have different durations and relevant activity may occur at different moments within them. Consequently, a generic way is required to probe the signal appropriately. Here, Event-related Potentials (ERPs) can be used. ERP designate local patterns (containing one or more extrema) occurring in the recorded EEG signals after the onset of a specific event. The occurrence of these patterns depends on specific underlying cognitive processes (e.g. surprise or face recognition) [Luck, 2014]. By identifying the relevant ERP, the associated cognitive processes a user is currently experiencing can be obtained, increasing our understanding of the user's perception of the interface. However, as discussed above, the proposed setup cannot simply rely on generic events relayed from the system-side components. Thus, event-related potentials need to be examined in greater detail to identify which ones can be produced and be relied upon by our BCI to gain insights into an interaction.

4.4 Event-related Potentials

Many different Event-related Potentials have been studied in the context of BCI [Luck, 2014]. Most of these potentials are restricted to very specific and strictly controlled situations. In the context of our application however, ERPs must be identified which can be applied in a broader and more adaptive context [Pfurtscheller and Da Silva, 1999]. First of all, ERPs have to be identified which are applicable in a bi-modal system combining Eyetracking and EEG and which give information about the user's perception of the interaction in one form or another.

4.4.1 Sensory, Attentive and Pre-attentive Potentials

As detailed above Event-related Potentials refers to patterns in the EEG signal occurring at a specific onset in regards to specific event. Practically, these patterns are observed by extracting same-length segments of the EEG signal called "epochs". Commonly, these epochs start either at, or slightly before, a given event onset and last typically for about one second [Rao, 2013]. Epochs are often grouped according to their corresponding event and are averaged thereafter. Averaged epochs, also called grand averages, are used to adequately identify the pertinent potentials contained within them.

The study of ERPs is commonly employed in BCI. Active BCI have notably been employing ERPs as the main interaction paradigm. Indeed, the events used as context information to locate ERPs can in some cases also induce potentials. One of the most well-known examples of this is the P300 potential. As the name indicates, the P300 is a positive potential occurring around 300 ms after the onset of an event. This potential occurs most commonly in the parietal cortex region. The event necessary for its appearance is related to the oddball paradigm [Donchin and Coles, 1988]. In other words, the P300 occurs when an unexpected but relevant stimulus appears in the visual field. This phenomenon is usually utilized in the context of a P300 speller paradigm. In this kind of application, a matrix or array of elements is presented to the user. All elements of this table light up in a semi-random sequence. This randomness ensures that the moment of illumination of a target element remains unexpected to the user, causing a prominent P300 (see mark 2 in Figure 4.3). Other similar potentials, such as the N400 have been employed in non-BCI settings to identify the presence of semantics inconsistencies or other incongruent linguistical properties during reading [Kutas and Federmeier, 2011].

This kind of application of ERPs relies on the conscious processing of the presented stimulation. There are however also applications which rely on lower-level paradigms. One such example are the Steady State Visually-evoked Potentials (SSVePs). These potentials occur when a focused visual element lights up with a constant frequency [Müller-Putz et al., 2005]. They appear with the same frequency as the flashing of the visual element. This type of potential has been used in similar selection situations as the P300 speller, by highlighting different interface elements at different rates. As the potentials arise on a lower level of perception, provoked by sudden sensory information, their use requires less concentration on the part of the user. Codebook Visually-evoked Potentials (CVeP) function in a similar way, relying on flashing sequences of stimuli with shifting onsets rather than changing frequencies [Riechmann, Finke, and Ritter, 2015].

In order to be observable, P300 and N400 require for user's conscious the processing of a very specific type of stimulus. On the other hand, SSVePs and CVePs imply that regular and controlled flashing is present within the user's visual field. Either of these situations is however rarely present in common interface manipulations, as surprise or overloading the user's visual field with flashing stimuli are not desirable. One type of potential whose required paradigm can commonly occur during manipulation are Error-related Potentials (ErrP). Indeed, these potentials rely on the occurrence of a procedural error which is noticed by the user [Chavarriaga and Millán, 2010]. Errors are common during interaction and noticing them helps to evaluate the usability of a system. However, to be identified, the corresponding EEG signal epoch has to be inspected during the occurrence of such an error. Without system context information locating the error event onsets such applications can be difficult.

Between attentive and sensory potentials, there exists another type of ERP. These can be seen at the local extremum in the blue curve marked by 1 in Figure 4.3 : certain ERPs can be present prior to the cognitive processing of the presented stimuli. These potentials are not as "low-level" as the sensory SSVePs or CVePs (i.e. occuring solely due to retinal stimulation with a light flash) but still show a pre-attentive response to the visual stimulation (i.e. before the content of the foveal region could be cognitively processed). Indeed, pre-attentive ERPs typically occur between 20 and 200 milliseconds [Winkler et al., 2005 ; Koelsch, Schroger, and Gunter, 2002] after a stimulus onset and can take a wide number of different roles. The roles of these various attentive and pre-attentive ERPs are described in Table 4.1. The pre-attentive potentials are of interest as they can occur frequently and without the presence of particularly tailored stimulation. When using

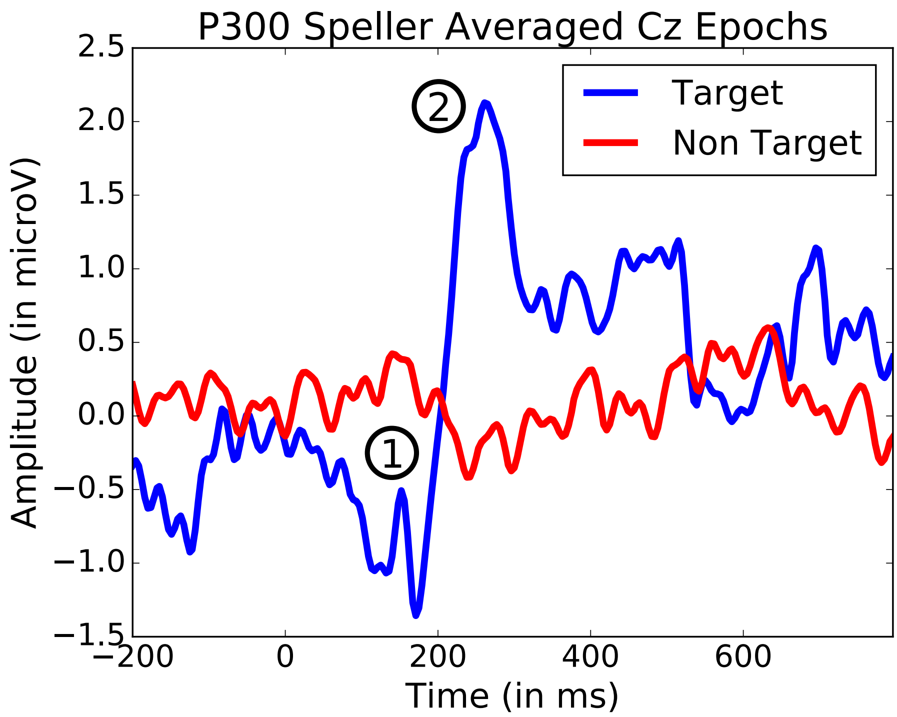


FIGURE 4.3: Average recordings at the Cz channel (according to the 10/20 system) during a P300 Speller BCI application. Participants were asked to focus their attention on a target element. Target corresponds to the target element lighting up. Around mark 1, various early ERPs can be observed. At 2 mark, the main P300 potential can be seen.

Eyetracking data for providing context, the observation and analysis of such early potentials seem, along with the examination of ErrPs, to provide a quite promising approach.

These early pre-attentive potentials can be observed in relation to eyetracking events such as eye blinks, saccades or fixations. As many of the early potentials listed in Table 4.1 occur in response to visual information, analyzing these potentials with our proposed setup should lead to new results. To effectively inspect these potentials, the fixation event is selected among the four types of ET events as the most conducive to this goal. Fixations allow for the observation of Fixation-related Potentials (FRPs) during the visual exploration of an interface. FRPs enable to compare and contrast the information provided by both EEG and ET data.

4.4.2 Fixation-related Potentials

Fixation-related Potentials are ERPs which occur at the onset of an eye fixation. As fixations are well defined temporally and spatially, they offer rich contextual information about the circumstances of the corresponding EEG signal epoch. Since many of the aforementioned potentials are contingent on visual events, their observation through FRPs is possible. The observation of very specific potentials can however be quite difficult when cognitive processes occurring long after fixation onset are concerned, such as for ErrPs. Indeed, while fixations are well located, their duration and the activity performed by the user during the fixation time can drastically vary. The varying duration of fixation [Hooge and Erkelens, 1998] makes it difficult to accurately locate some of the more dynamic potentials (e.g. P300, ErrP, ...). This aspect is exacerbated by the lack of precise

TABLE 4.1: Different types of ERP components which can be observed after onset of certain events of brain activity. Simplified from Luck, 2014

Component Name	Contingent Event Type
P100	Processing stimuli
N100	Unpredictable stimuli
N170	Structural encoding of faces
P200	Selective attention
N200	Stimulus identification, detection of novelty or mismatch
P300	Novelty processing, surprise, unpredictable event
N400	Semantic incongruency
P600	Improbable events

labeling of fixations in regards to the performed task. On the other hand, lower-level and pre-attentive early potentials (e.g. P100, N100, ...) can be located with greater ease and be used effectively with FRP.

Traditionally, FRPs have mostly been employed in the domain of linguistics, providing additional information to evaluate the human processing of language [Holcomb, 1993]. These applications are very controlled in nature but have shown valuable insights into the inner workings of text understanding. Concerning their usage in broader interactions, FRPs have only been applied to a limited scope of scenarios. Notably, interactions scenes were adapted (for example by occluding parts of them) to produce a similar paradigm as studied in other common BCI. When performing visual exploration of such adapted scenes, potentials similar to P300 or N400 have been observed in FRP epochs. For the P300, an equivalent potential occurs when performing a gaze contingent task where stimuli are hidden from the user [Finke et al., 2016] or during experiments with more natural exploration [Brouwer et al., 2013]. Similarly, in a visual environment, which presents elements not congruent with the rest of the scene, potentials comparable to the N400 potential are often detected [Coco, Nuthmann, and Dimigen, 2020]. More common potentials can thus also be observed within FRPs. This is encouraging as it shows the applicability of FRPs in classical BCI interaction. However, regarding the transposability of the proposed BMI system, the use of these specific potentials will not be possible in most scenarios. Rather, pre-attentive potentials can the used to evaluate an interface.

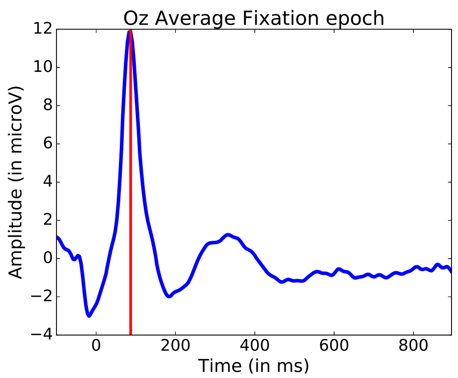


FIGURE 4.4: P100 component observed at the onset fixation in a visual exploration task.

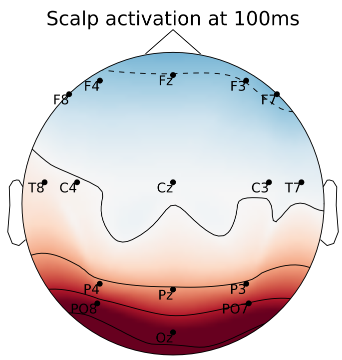


FIGURE 4.5: FRP observation at the scalp of a user, 100ms after fixation onset.

As stated previously, visually contingent potentials can be observed at the onset of fixations and have been studied in the case of controlled FRPs [Jeffreys, 1996]. In the more general context of observing interactions, some of the early pre-attentive potentials need to be considered. Components like the P100 and N240 can give general information such as the attentional cost put into the fixation by the user [Luck, 2014] as well as the perception of faces [Puce, Allison, and McCarthy, 1999] respectively. In particular, the first attention-related property extracted from the P100 component can be quite reliably observed in FRPs in occipital channels. Figure 4.4 and Figure 4.5 show where this prominent P100 is located, both temporally and spatially. For Figure 4.5, a dark red color corresponds to high positive amplitude while blue corresponds to a negative amplitude. The P100 potential is naturally occurring at the onset of a fixation in the occipital cortex. Its attention-related property will help to provide insights into the perception of the interface by the user. The P100 presents comparatively a much lower amplitude than the P300 or N400 potentials, but its observation through FRPs is more stable as it is related to pre-attentive cognitive processes.

The value of all these properties will hereafter be assessed in the context of various applications. However, due to several difficulties arising when inspecting the different potentials, reliable methods are needed to adequately access and assess Fixations-related Potentials first. This is necessary in order to use them in the context of the proposed BMI for interface evaluation.

Chapter 5

Methods for Event-related Potential Analysis

5.1 Required Methods

In the previous chapters, the hardware (i.e. sensors) composing within our BCI system have been discussed. It was also discussed what kind of common biosignal activity could be measured through this BCI. In this analysis done in Chapter 4, FRPs have been identified as a novel but potentially valuable type of signal property which can be collected with our setup. Indeed, FRPs can provide a broad variety of information for the evaluation of interfaces. FRPs also satisfy the previously set transposability criterion and allow access to new understandings about the proposed BCI setup. Now, methods are needed which enable an appropriate handling of FRPs for system evaluation. For an adequate handling of FRPs, we specifically need to conceive methods which respond to the two following questions :

- How can FRPs be reliably observed with two separate modalities ?
- How can information about single moments of interaction be obtained from FRPs ?

Regarding the first question, to be even able to perform this type of FRP analysis on the signal, fixation potentials have to be extracted from recorded biosignals first. As detailed in the previous chapter, FRPs are observed by combining the data from EEG and Eyetracking recordings. Both recordings thus need to be synchronous for the correct identification of FRPs. Different methods have been conceived with the objective to synchronize signals. While most of them are easily applicable and commonly used, most of them require preparations to be taken prior to conducting the manipulation. As there are many everyday situations where these precautions are not necessarily taken, desynchronization of modalities still occurs frequently. For this purpose, an additional method [Wobrock et al., 2019], here called Fixation-based component resynchronization (FCSync), has been developed in this thesis which uses the properties of the P100 FRP potential and its stable location within epochs to synchronize signals. This method is applicable even when crucial information (i.e. the difference between the recording timestamps of both ET and EEG devices) are missing. The FSync is described in detail in Section 5.3.2.

The second question is about how to evaluate FRPs to access their momentary informative properties. So far, ERPs have mostly been evaluated by averaging multiple EEG epochs featuring similar types of events. In everyday interactions, however, an important aspect is the changing of the user's perception as the manipulation progresses. When aiming to evaluate an interface, designers may want to know which part of the interaction caused the user grievances. By observing the ERP grand averages, these changes in perception are often erased and cannot be detected.

When users manipulate an interface, they may repeat a same action many times. This action is however rarely performed in a completely identical context. An example of this would be performing an eye-fixation onto the speedometer of a car. This action is performed multiple times while driving, but can either serve to gauge the speed of other cars or to ensure one is below the speed limit. While mechanically the same action, cognitive processes occurring during these tasks are different. In order to reliably inspect how various properties of an ERP component are changed by such differences, a method must thus be developed which can extract information from single epochs. For this purpose, the method proposed by Hu et al. (2011) has been adapted to be applicable on single trial ERPs. This method [Wobrock et al., 2016], here called Adaptive Multi-linear Regression with dispersion term (aMLRd), allows to reliably extract a potential's amplitude, latency and morphology from an epoch. This permits for a deeper analysis of the signal and a more resource efficient exploitation of the recorded data. The aMLRD is described in detail in Section 5.2.1.

Both methods (i.e. aMLRd and FCSync) for synchronizing recordings and extracting potential properties which address these major issues will be employed in the BCI system to improve the aspects of transposability and information extraction. In this chapter, both methods and their application contexts will be described in detail, starting with extraction of single trial properties. While counter-intuitive, this is necessary as the FC-Sync method is reliant on the aMLRd method to function. For this next section, modality recording of ET and EEG will thus be considered synchronous.

5.2 Accessing Single-Trial ERP Properties

The two previous chapters (3 and 4) detailed the required preprocessing steps for EEG and Eyetracking data to be clean of artifacts and to allow for adequate extraction of information. These methods were employed and the resulting data was segmented into epochs according to fixations. As stated before, the traditional methods for assessing the informational value in FRPs consists either in looking at grand averages, or, in a more algorithmic fashion, using a classification algorithm to validate discriminability between epochs belonging to different groups. An example of such a classifier is the Linear Discriminant Analysis (LDA) as detailed in Bishop (2006). In all these cases, however, the intricacies of ERP changes occurring on single trial level are not explored to much extent beyond simply stating discriminability between considered categories.

As also mentioned in the previous chapter (see Figure 4.4), FRPs present a positive potential in the occipital channels (centered around the Oz location in Figure 3.3) after fixation onset. This P100 component is an always present ERP in fixation-related epochs [Creel, 2019]. It thus constitutes a possible source of information for understanding changes in context between eye fixations. One way to compare this ERP between fixations is to extract its wave signal's properties for each epoch. These include ERP amplitude, latency and morphology. Amplitude refers to the voltage in microvolt (microV) of the component's peak. Latency refers to its temporal position (typically in milliseconds) with respects to the event onset underlying the event (i.e. fixation onset). Lastly, morphology refers to the shape or width of signal occupied by the potential. Mathematically, morphology can be calculated through the duration between the first inflection points of the EEG curve before and after the core ERP extrema. A graphical example of the acquisition of these three ERPs properties is presented in Figure 5.1.

The P100's amplitude has been shown to be dependent on the attentional effort deployed by the user [Desmedt and Tomberg, 1989]. In the context of this work, this correlation will be explored further by taking into account how the different fixation contexts

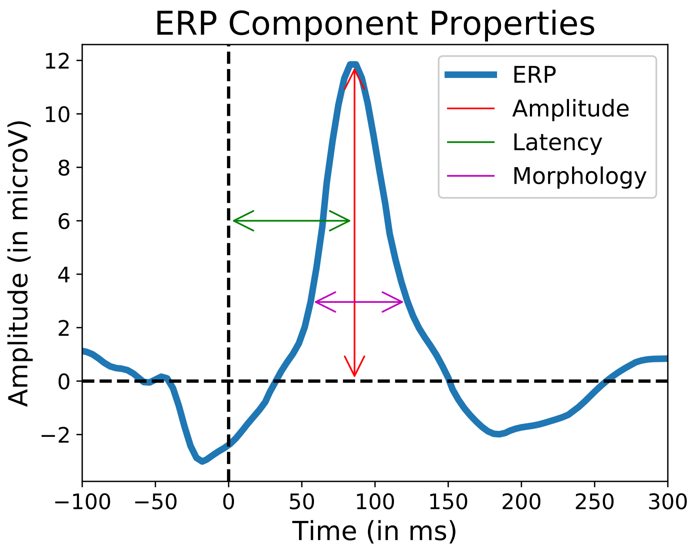


FIGURE 5.1: Three of the properties which can be extracted from an ERP are its amplitude, latency and morphology. These metrics are graphically represented here by the red, green and magenta double arrows respectively. In this example, their values are approximately 12 microV, 90 ms and 60 ms respectively.

impact the P100's amplitude and thus, the user's attentional effort. However, so far, all presented methods used to investigate ERPs relied on summing up multiple epochs. As context of an eye-fixation and attentional effort changes between single fixations, these traditional methods would erase small but meaningful changes in ERP properties.

To get information about ERP properties from single epochs, a method called adaptive Multi-Linear Regression with dispersion term (MLRd) will be applied [Hu et al., 2011]. This method has been already deployed on artificial, preconstructed signals but can be expanded to dynamic EEG signals through segmentation and the use of endemic properties of the averaged ERP signals. This has been explained in Wobrock et al. (2016). For clarity, we will refer to our adapted version of the MLRd as the "adaptive Multi-Linear Regression with dispersion term (aMLRd) " from here on. The MLRd and the aMLRd are detailed side by side below.

5.2.1 Adaptive Multi-linear Regression with dispersion term (aMLRd)

Even when an ERP, such as the P100, remains relatively stable between epochs in regards to amplitude, latency and overall shape, it can still happen that the potential is occluded by other brain activity occurring simultaneously in the epoch. Thus, to be able to contemplate what effect the changes in interaction context have on ERP properties, these ERPs have first to be correctly identified in the considered epoch. This is a non-trivial task and requires to account for the variability and main characteristics of an ERP present across multiple epochs to not misattribute an ERP to the wrong peak in the signal. The aMLRd method offers a solution to this problem and adequately locates and replicates the underlying ERP in warped signals.

The original MLRd algorithm [Hu et al., 2011] was established to identify where a given ERP is located in a warped signal. The MLRd accomplishes this by fitting a set of regressors, representing the ERP and its variability, onto this warped epoch. Practically, each regressor is multiplied by a coefficient and then summed together. The resulting linear combination of regressors forms an approximate isolation of the potential. This isolated potential can then be used to extract the ERP properties with greater ease : by looking at its magnitude for the amplitude, extremum location within the epoch for latency and variance for morphology. In this fitting step, the MLRd method uses a set of regressors as a template to represent an approximate ERP and its variability. In this classical MLRd procedure, the ERP template is created from a dataset composed of artificially modified example signals containing only the isolated ERP (i.e. the latency, amplitude and morphology are artificially shifted). The MLRd uses the principal components of this dataset (calculated via Principal Component Analysis (PCA) [Bishop, 2006]) as a set of regressors. This is a contrast to the Multi-linear regression (MLR) method, which used the averaged isolated ERP and its first derivative (akin to a Taylor series expansion) as regressors. The MLR was found to poorly represent the variability of an ERP. By using principal components, the MLRd provides a better flexibility in property variation allowing for an improved location of the corresponding peaks within warped epoch. According to Hu et al. (2011), the first three principal components are regressors represent the variability of amplitude, morphology and latency respectively. The second (morphology) and third (latency) components thus take the role of the eponymous "dispersion term".

The MLRd algorithm can only focus on one peak changing at a time as it uses one template (i.e. set of regressors) for one peak. In real EEG data, more than one peak can be present. For this purpose, when working with real EEG data, we have created the aMLRd to account for these new complexities. The procedure of aMLRd calculation is detailed in Algorithm 2. The process of our method is also illustrated in Figure 5.2 and described hereafter. Practically, we examine the signal by averaging all epochs which are supposed to contain the considered peak in a specific channel (seen A to B in Figure 5.2). Afterwards this average signal is segmented according to its inflection points (seen in the B image of Figure 5.2). Each segment between two of these inflection points will be considered separately. In the original MLRd algorithm, the average signal segment containing the peak of interest was extracted and duplicated. Each duplicate had the amplitude, latency and morphology of the considered peak shifted around a given scope of tolerance. A PCA was then performed on this set of shifted averages to obtain the regressors. This approach however poorly represents the natural variations of peak properties across all epochs. To improve this part of the method, one could simply perform the PCA on the original recorded set of EEG epochs. However, in these EEG epochs, the peak of interest might not appear clearly or be completely absent. Such occurrences would deviate from the original MLRd, where principal components were extract from a set where all epochs contained the considered peak. To not stray too far from this approach, the recorded EEG epochs can be averaged together in small batches. This ensures the presence of the peak but conserves in part the variability of the peak. This scope of variation in latency, amplitude and morphology of a potential is commonly unknown with regular EEG data. Thus in the aMLRd, recorded epochs from the EEG dataset are randomly shuffled and averaged in small batches (e.g. 3 to 6 epochs) to simulate the variations of the ERP peak while ensuring its presence and prominence (seen in the C, D and E images of Figure 5.2). To ensure that the resulting set is still representative of the considered peak's variability, this process of shuffling and batch averaging is repeated multiple times. The averaging process occurs by using a moving window across the shuffled EEG epochs and which can have a certain amount of overlap. The resulting set of averaged epochs are stored in a new dataset.

On this new dataset, which takes the role of the set of shifted average epochs in the original MLRd algorithm, a PCA is performed. To ensure that the proper principal components are extracted for each considered peak, a gating function is previously applied to the template epoch dataset according to the marked regions (see B and E images in Figure 5.2). The PCA is performed for each gated epoch segment independently (see step F

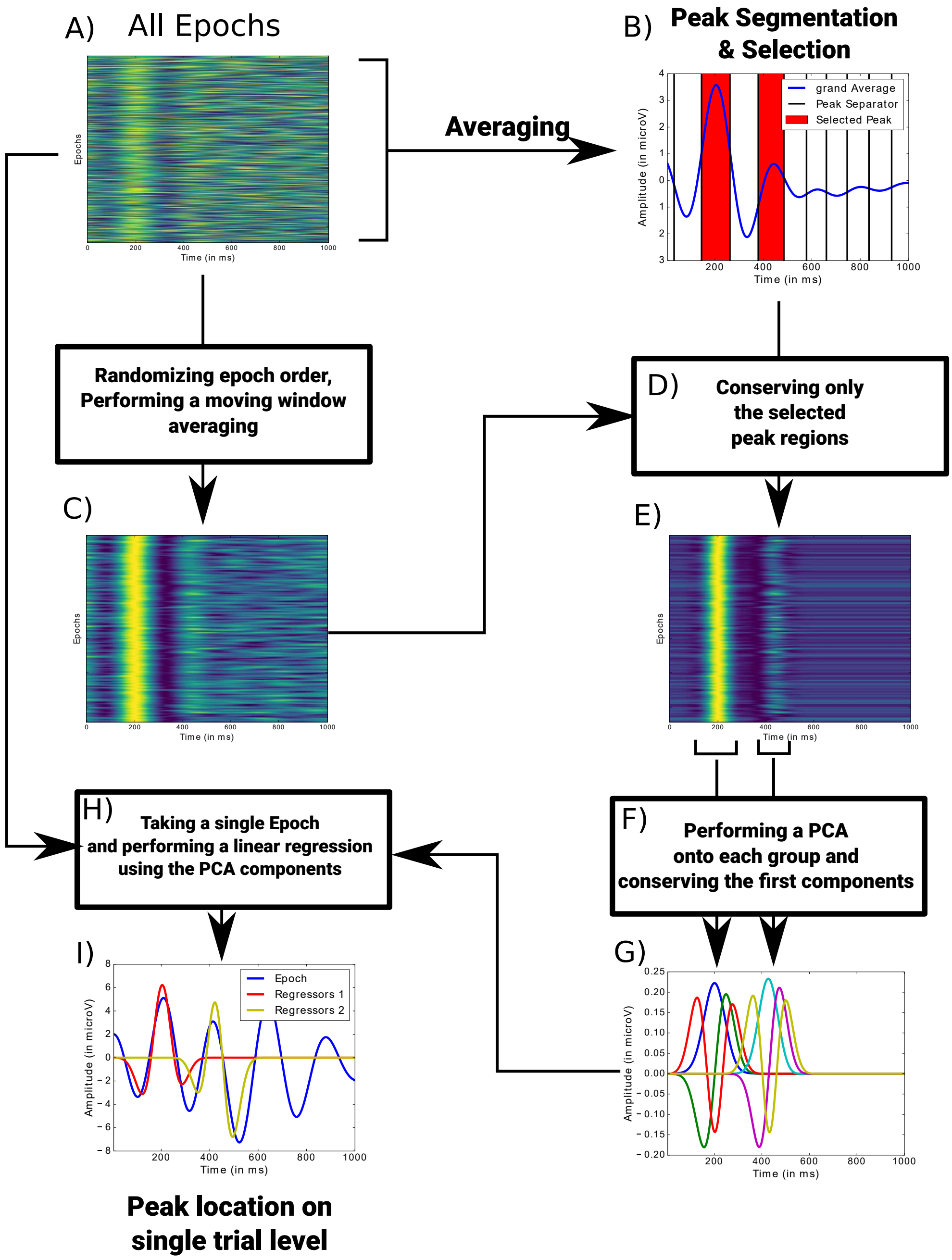


FIGURE 5.2: Application of the adaptive Multi-linear regression with dispersion term on a set of EEG epochs. Within these epochs, two notable potentials can be detected. Regressors for both of these components are created and used to locate components on a single trial level as illustrated in the bottom left image. Presented in Wobrock et al. (2016).

in the same Figure). From the resulting principal components, only the first n ones are considered. n is chosen in such a way that the selected components represent 99 percent of the signal's total variability. This calculation is done using the eigenvalues of the resulting principal components. This percentile can be modified but is usually maintained in order to at least contain the first three principal components of a potential. These components are grouped together (seen in bottom right image G in Figure 5.2) and will then serve as regressors for the analysis of each singular epoch.

For each epoch in which the considered peaks need to be identified, a linear regression is performed (an single epoch from A is used in the step H in the Figure). The set of principal components is regressed onto the considered epoch. By adding up the subset of principal components which were originally used to identify the considered peak, the same peak can be identified within the scope of the warped epoch (seen in bottom left image I of Figure 5.2). By employing this technique, the location of the new peak is obtained, and subsequently, the corresponding region around it can be examined to extract the properties relating to it. If the regression does not result in the creation of the appropriate local extremum, the peak (and consequently the EEG potential) is considered absent from the epoch.

By means of our aMLRd method, components can be identified with greater reliability and examined in more detail. However, certain limitations are still present. For one, the application of this method requires artifact-free signals and an adequate number of regressors. Here, the number n of regressors (i.e. PCA eigenvectors) used was chosen through expertise rather than calculations based on the input signal. Furthermore, the aMLRd method is only able to identify potentials which do not shift too much during manipulation (i.e. they remain at a relatively stable latency during the recording sessions). If other potentials are present in the gated location of a distinct template potentials, they are equated to the template potential by the aMLRd algorithm even when this is not the case. While such cases of overlap are not present in the studies discussed in the rest of this work, this problem of overlap still needs to be discussed. Indeed, especially as eye fixations are often made in rapid succession, different potentials overlapping in the EEG recording are not uncommon within FRPs.

5.2.2 ERP and Epoch Overlap

In naturalistic interactions and free manipulation tasks, events can succeed one another in a rapid fashion [Woldorff, 1993]. This is the case with fixations which typically last around 250 ms, as the brain activity associated with them may sometimes take longer than 250 ms to occur. This is especially important when considering later potentials such as the P300 [Finke et al., 2016], where EEG activity relevant to one epoch can also be present in the subsequent ones if fixations were performed too quickly. In the case of the P100 however, this may be rarely a problem, as the potential is short and occurs 100 ms after fixation onset (as described in Chapter 3, fixations are generally longer). Overlap is, nevertheless a difficult issue to tackle when considering other paradigms (e.g. P300-based object recognition). In these situations, the aMLRd approach is not able to remove a potential without removing other relevant activity. This occlusion between epochs may thus influence the quality of EEG component property identification. This issue also cannot be removed through classical methods of preprocessing.

One way to deal with this problem is through the application of deconvolutional approaches such as proposed in the Unfold MatLab toolbox [Ehinger and Dimigen, 2018]. These methods however rely on knowing beforehand to which category of fixation an epoch belongs (i.e. knowing what kind of activity to expect within the epoch). This information cannot be gathered in the scope of a natural interaction. Nevertheless, this

Algorithm 2 adaptive Multi-Linear Regression with dispersion term

Initialization;
We suppose a set of M epochs
x_i where $\forall i \in [1, ..., M]$, $x_i \in \mathbb{R}^N$ ▷ x_i : Epoch i with N samples
x_{test} where $x_{test} \in \mathbb{R}^N$ ▷ x_{test} : An Epoch to identify peaks in
$X \leftarrow [x_1, x_2, .., x_M]$
Select b_{size} and b_{step}, where $1 < b_{size} < M$ and $1 \leq b_{steps} \leq M - b_{size}$ ▷ Size and Speed of averaging window. We used $b_{size} = 6$ and $b_{step} = 3$
Select n_{comp}, where $1 < n_{comp} < M$ ▷ Nb of Principal Component, often $n_{comp} = 3$
Peak Segmentation and Selection;
$x_{ref} \leftarrow X$.mean()
$L_{inf} = [l_1, l_2, ...] = argmin_x |\frac{d^2 x_{ref}}{dt^2}|$ ▷ Finding inflection points
$x_{seg} = [0, 0, ..., 0]$, $x_{seg} \in \mathbb{R}^N$
for $j \leftarrow 1$ to L_{inf}.size()-1 **do**
    Check if $x_{ref}[l_j : l_{j+1}]$ contains an interesting peak
    if $x_{ref}[l_j : l_{j+1}]$ contains a peak of interest **then**
        $x_{seg}[l_j : l_{j+1}] = [1, 1, ...1]$ ▷ Windowing segments (here square)
    end if
end for
Creating template epoch set;
$X_{temp} \leftarrow []$
X.shuffle() ▷ Initialize empty array
for $j \leftarrow 1$ to M by b_{steps} **do**
    $xt_j = X[j : j + b_{size}]$.mean() $\odot x_{seg}$; $\forall j$, $xt_j \in \mathbb{R}^N$ ▷ Element-wise multiplication
    X_{temp}.insert(xt_j)
end for
Creating Regressors;
$PeakRegs \leftarrow []$
for $j \leftarrow 1$ to L_{inf}.size()-1 **do**
    if $x_{ref}[l_j : l_{j+1}]$ contains a peak of interest **then**
        $regs \leftarrow$ PCA($X_{temp}[:][l_j : l_{j+1}]$)
        $PeakRegs$.insert($regs[1 : n_{comp}]$) ▷ Dispersion term : $regs[2 : n_{comp}]$)
    end if
end for
Application to new epoch;
for k in peaks of interest **do**
    r_{coefs} = regress(x_{test},$PeakRegs[k]$)
    $aMRLdPeak_k \leftarrow$ sum($r_{coefs} \times PeakRegs[k]$) ▷ Clean Representation of the peak
end for

algorithm can still be employed to correct epochs after recording. This method to correct or to account for epoch overlap will not be investigated further within this work. Returning to the extraction of peak properties : in order to be able to perform the aMLRd and extract an EEG potential's properties, epochs first need to be generated. This requires that the recordings from EEG and Eyetracking are adequately synchronized.

5.3 Synchronizing Modalities

FRPs are obtained by segmenting the EEG signal relative to fixation onsets recorded through the eyetracker. As both modalities (i.e. EEG and ET) are rarely captured by the same device, the recordings from two separate origins have to be combined. To perform this step effectively, the recordings have to be synchronous and labeled with the timestamps originating from the same clock. Many different approaches exist to achieve this synchronicity. The classical ones use a timestamp synchronization step between both modalities before performing the main recording session.

5.3.1 Classical Approaches

Recording of eyetracking and EEG data is traditionally performed by means of separate devices. Commonly, these devices are connected to the same computer. This insures that recordings are synchronous, given that the same computer clock is used to write timestamps. However, when considering the large number of different setups and possible interaction scenarios our BCI aims to cover, recordings can also be performed at different computers, especially when working with proprietary software. In this case, other methods of synchronization need to be considered.

One notable example is the utilization of synchronization markers [Thie, 2013]. These consist of additional recordings or punctual markers which are introduced into the EEG and Eyetracking recordings as a separate recording channel. Through these independent markers, reference points for synchronization are obtained [Lee et al., 2011]. Alternatively, special communication protocols can be used. Such protocols rely on their own timestamps (i.e. gather online) ensuring a synchronous communication of the data. One of the prominent communication protocols for physiological data is the Lab Streaming Layer protocol developed by Kothe (2014). However, all these timestamp correction procedures rely on actions taken before or during recording. This means that two unsynchronized ET and EEG recordings, even if gathered simultaneously, cannot be inspected for FRPs aposteriori as their timestamps are incongruent. To solve this issue where reference marker or correct timestamps are missing, other ways of synchronization need to be considered, which we will present here in the form of a novel method.

5.3.2 Fixation-based Component Resynchronization (FCSync)

When considering EEG and Eyetracking, both recordings can be seen as intersecting in the EOG signal. Indeed, eye motions are picked up in the EOG signal allowing its comparison with eye motions recorded through optical eyetracking methods to readjust the timestamps. However, this comparison method to identify shifts in timestamps implies the presence of EOG recordings. Similarly to EOG, electrodes placed in the frontal region of the scalp can harbor eye movement artifacts. Eye motions recorded in the frontal channels are prominent, often overlapping regular brain activity before being corrected for artifacts, as investigated by Croft and Barry, 2000. They can serve as an alternative, but less precise way, to link together EEG and ET data for timestamp correction.

However, is it still possible to synchronize the recordings in cases where neither EOG nor frontal EEG channels are present ? When considering recordings where emphasis is placed on occipital activity (such as the P100 potential), the presence of frontal electrodes may indeed not always be required. In such situations, the stability of the very P100 potential latency can serve as a reference point to synchronize the EEG and Eyetracking recordings. This stable latency is used in a method we developed called the Fixation-based component resynchronization (FCSync). This method is described in detail in the Algorithm 3.

Algorithm 3 Fixation-based Component Resynchronisation

Initialization;
$x_{EEG} \in \mathbb{R}^N$ ▷ EEG Signal
$t_{EEG} \in \mathbb{R}^N$ ▷ EEG Timestamps
$T_{ET} \leftarrow [t_1, ..., t_M], \forall i, t_i \in \mathbb{N}$ ▷ ET Fixation Timestamps
SR_{EEG}, SR_{ET} ▷ EEG and ET Sampling Rates
Static Offset Correction;
$x_{center} = [0, 0, ...0], x_{center} \in \mathbb{R}^{2N}$ ▷ create empty array
for $i \leftarrow 1$ to M **do**
  $d_{time} \leftarrow round((t_i - t_1) \times \frac{SR_{EEG}}{SR_{ET}})$
  $x_{center}[N - d_{time} : 2N - d_{time}] += x_{EEG}$
end for
$t_{offset} \leftarrow (argmax_s x_{center}(s) - N) \times \frac{SR_{ET}}{SR_{EEG}}$
$\forall i, t_i = t_i - t_{offset}$
Linear Drift Correction;
x_{EEG} segmented into $X_{EEG} \leftarrow [x_1, ..., x_M]$ by t_i ▷ At increasing onsets of t_i
Choose b_{size} and b_{step}, where $1 < b_{size} < M$ and $1 \leq b_{steps} \leq M - b_{size}$ ▷ Size and Speed of averaging window. We used $b_{size} = 10$ and $b_{step} = 5$
$Reg_{onsets} = []$
$Reg_{latencies} = []$
for $i \leftarrow 1$ to M by b_{step} **do**
  $x_{mean} \leftarrow X_{EEG}[i : i + b_{size}]$.mean()
  Reg_{onsets}.insert($sum(T_{ET}[i : i + b_{size}])$)
  $Reg_{latencies}$.insert($argmax_s(aMLRd_{P100}(x_{mean}))$) ▷ Latency of aMRLd result, trained using X_{EEG}
end for
$a, b \leftarrow$linear.regression(Reg_{onsets}, $Reg_{latencies}$)
$\forall i, t_i \leftarrow a \times t_i + b$
The new t_i can be used for correct segmentation

A difference in timestamps between two recording clocks can take two possible forms, supposing both clocks have been configured to try to record time accurately. The first form is the presence of a timestamp offset between recordings. This means a static and constant difference between the timestamps is present between recordings at any moment in time. The second type of difference concerns the presence of a linear drift in clock ticking speed. This means that for one recording, timestamps increase faster than for the other. This drift widens the difference in timestamps between recordings over time. In the case of common recordings of both EEG and ET, both forms of timestamp difference should be supposed to be present.

Relying on both the static latency of the P100 peak within a fixation epoch and the previously described aMLRd method of accessing ERP properties, the FCSync method

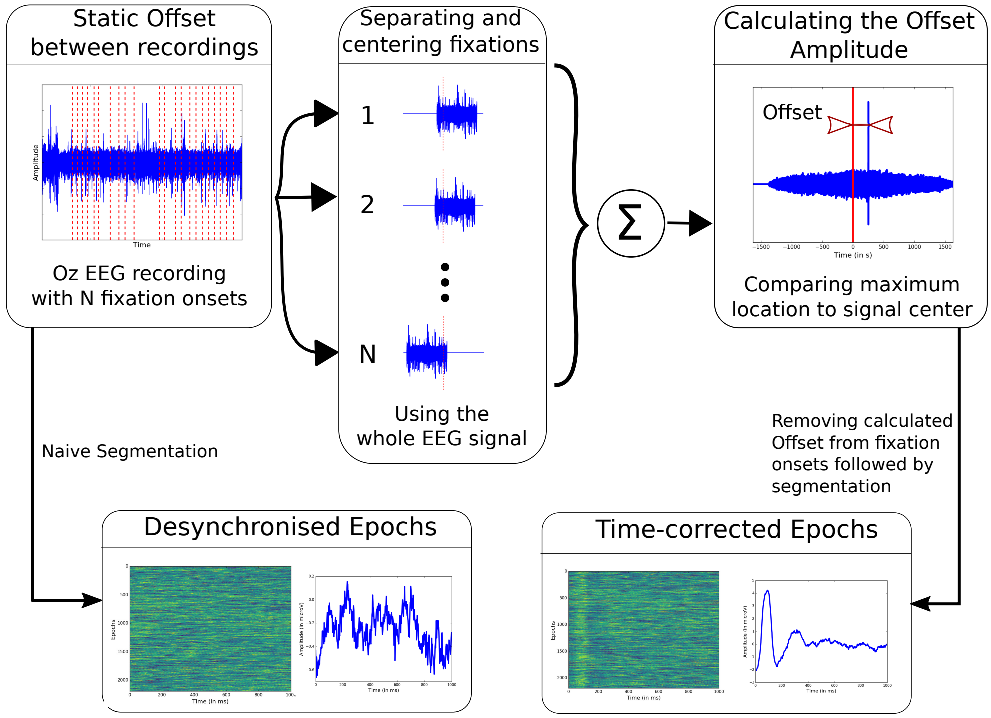


FIGURE 5.3: Fixation-base component resynchronization method employed for correcting static difference in clocks in EEG and Eyetracking recordings.

[Wobrock et al., 2019] can be applied which allows for the resynchronization of both types of clock differences. The procedures of this method are illustrated in Figure 5.3 and Figure 5.4.

When considering a static offset between two clocks, the engendered time difference can often be quite large in terms of signal analysis (e.g. more than 10 seconds). When an unknown offset is present and the EEG signal is naively segmented along the uncorreted, time shifted fixation onsets, potential-less epochs are observed. Figure 5.3 shows the failing of a naive segmentation (e.g. a grand average without potentials) and how the FCSync offset method corrects it. To resynchronize both recording, the P100 location has to be identified. As the static offset is constant between epochs, the P100 is located at an unknown static time duration away from the naive onset of a fixation. Under these circumstances, the totality of the recorded signal needs to be examined. Indeed, for each fixation, the entirety of the recorded signal can be shifted to align each of the fixation onsets (seen in the upper middle image in Figure 5.3). While the time offset between EEG and ET recordings is unknown, time differences remain the same between epochs. Practically, the first fixation is arbitrarily put at the very start of the EEG recording, the following fixations shift the signal temporally in relation to their time difference with the first fixation. The P100 is the only potential within these epochs which is 1) always reliably present [Dimigen et al., 2011], 2) time locked to the fixation onset and 3) occurs independently from other event types. Thus, by averaging these shifted signals, the P100 is the only potential which does no disappear (as seen in the upper right image in Figure 5.3). Thus, by taking the difference between the initial start position (i.e. the start of the EEG signal) and the only maximum present in the averaged signal, the magnitude of the static offset can be calculated. Subsequently, this offset value is removed from all considered fixation onsets to impose synchronization. This results in the time-corrected epochs which are shown in the bottom left image in Figure 5.3.

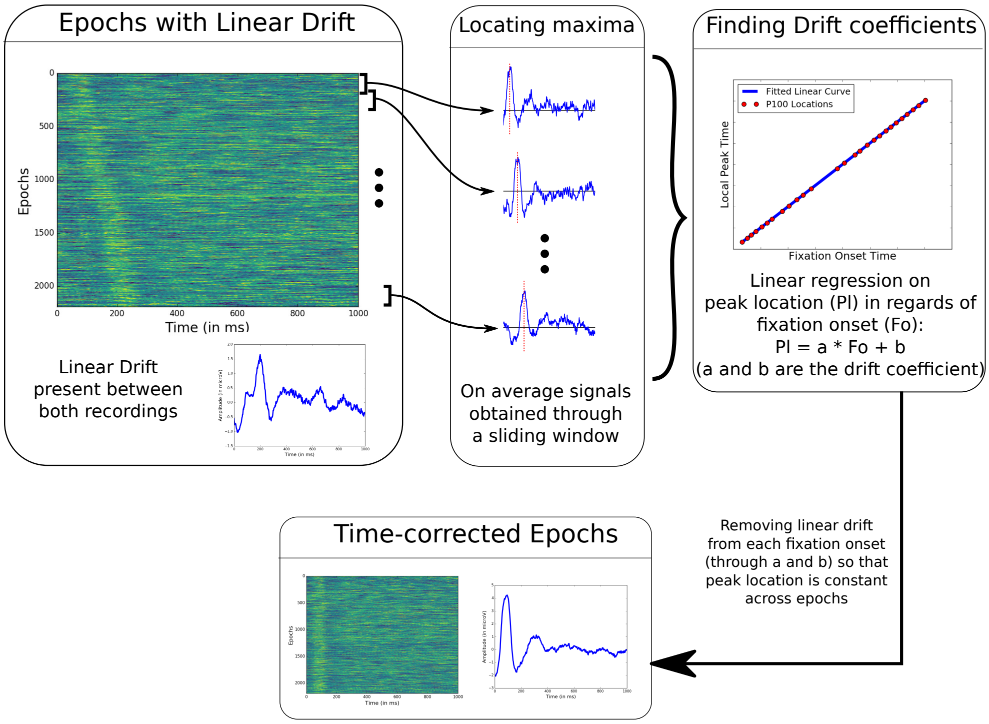


FIGURE 5.4: Fixation-base component resynchronization method employed for correcting linear drifts in clock ticking speed between EEG and Eyetracking recordings.

In the case of a linear drift, shifts in timestamps are often less significant. As both clocks try to keep close to real time, the drift between them is usually smaller than a couple of milliseconds per minute [Walker and Barrack, 2003]. In short manipulations this difference may be barely perceived but can cause larger time gaps over longer sessions. This phenomenon and its correction via FCSync drift correction are illustrated in Figure 5.4. To remove this drift, a naive segmentation of the signal into epochs is performed (see the upper left image in Figure 5.4). This is done in such a way that the P100 component is observable within the obtained epochs. The epochs are arranged in order of chronological appearance, from the earliest to the last. A moving window is then used to average a certain amount of epochs as to make the P100 potential appear more prominent (as illustrated in the upper middle image in Figure 5.4). The size of the moving window should be kept small to ensure that the resulting averages are composed by epochs for which the P100 latency remains almost constant (i.e. this extremum's latency doesn't vary more than a few samples). Within each of the averaged signals, the P100 is located using the aMLRd method and the average onset timestamp of each window of epochs is calculated (i.e. by averaging the naive onset time of used fixations). A linear regression is performed, fitting the relative P100 latencies to the calculated average onsets (seen in the upper right image in Figure 5.4). The coefficients, a and b, obtained from this regression describe the linear drift (see Equation 5.1). By using the naive onset timestamps of each fixations, the equation can be used to identify the expected latency of the P100. As the P100 potential is supposed to present a constant static latency in an epoch, an adequate shift in fixation onset can then be performed to align all P100 latencies between epochs.

$$t_{P100\,latency\,in\,epoch} = a * t_{naive\,fixation\,onset\,timestamp} + b \tag{5.1}$$

Both of these FCSync methods can be performed in conjunction (see Algorithm 3).

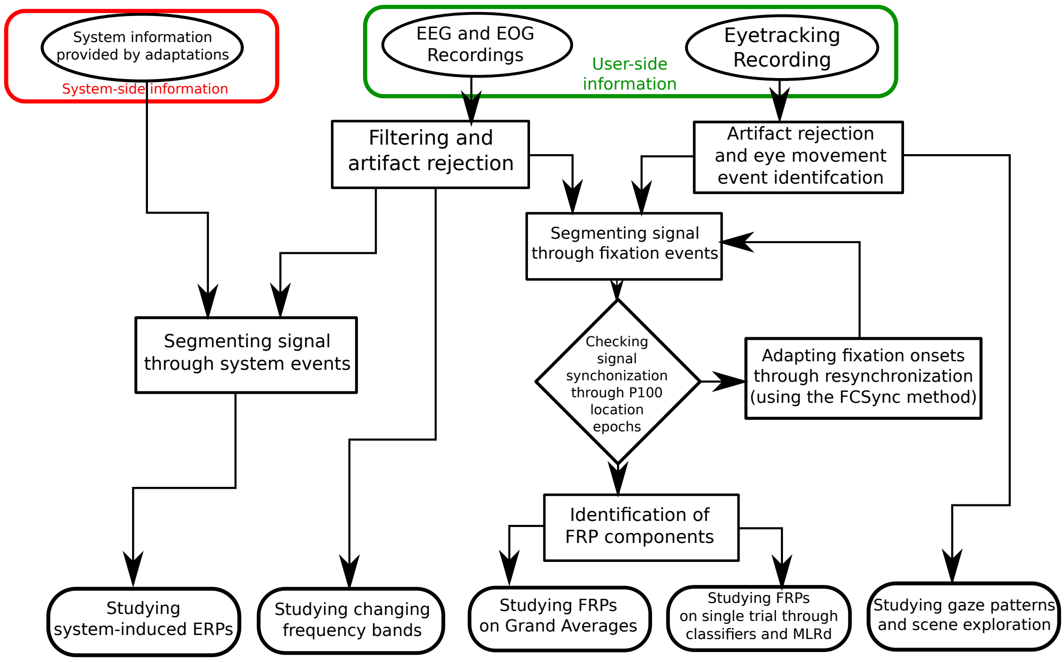


FIGURE 5.5: Process of obtaining biosignal information which can be analyzed to extract new insight about an interaction.

Linear drifts commonly present a quite low magnitude. The averaging step to identify the location of the P100 in the static offset corrections algorithm (upper middle and right images in Figure 5.3) still presents a peak, even when a linear drift is present, which can be used to identify the general location the P100 component. The semi-corrected timestamps, obtained through the FCSync offset correction, are then used as the epochs containing the P100 component in the first step of the FCSync linear drift correction. Both of these correction methods have been used in tandem [Wobrock et al., 2019] and have been shown to present only a 5 ms error margin when correcting properly synchronized data where timestamp differences were manually induced post-recording. This testing was done by utilizing data obtained from two different FRP studies [Wobrock et al., 2017] which will be presented and discussed further in Chapters 7 and 8.

When performing an observation of FRPs, these FCSync methods permit the resynchronization of recordings even when desynchronization occurred. The aMLRd and FCSync methods thus cover all tools needed for the proper extraction and analysis of FRPs. Furthermore, the fact that these methods are applicable in the context of an easily transposable BCI for the evaluation of interfaces, is aligned with the goal of this work.

5.4 System Summary So Far

Before beginning the examination of how our BCI extracts information about a user's perception of an interface, a summary of our proposed system is presented in Figure 5.5.

First, the two recording modalities are inspected separately for artifacts and preprocessed in their respective fashion. Artifacts are removed from the EEG data via filtering and ICA component rejection. Potentially the observation of the EOG signal can be used to identify other eye movement artifacts if not completely removed through these two methods. The eyetracking data are processed from a continuous stream into a discrete set of fixations, saccades and blinks. From this point, both modalities can be contrasted with each other : the timestamps of each modality are compared and the EEG signal is segmented into epochs according to fixation onsets. If these epochs, particularly in the occipital channels, do not present a clean P100 component located at about 100 milliseconds from fixation onset, the recordings can be considered desynchronized. The FCSync methods can then be applied to resynchronize the recordings and extract correct FRPs. This resynchronization step may not be required in most scenarios where correct and maintained synchronization was performed. However, in situations where EEG and Eyetracking recordings were done on separate computers, the FCSync can be used to harmonize the timestamps, thus allowing for the exploration of FRPs and their properties.

As the data are now corrected and artifact free, the processed signals can be analyzed. To extract the maximum information possible, both the cleaned EEG and Eyetracking recordings can initially be examined separately as discussed in Chapter 3. This can notably be done by looking at the different frequency bands via the use of Fast Fourier Transform (FFT) or by examining the gaze patterns within the scene by segmenting the environment into fixated regions of interest. Thereafter, the main analysis will be performed on the FRPs by examining which components are present in the grand averages or by inspecting components in individual epochs by means of the aMLRd method. All these analyses then need to be considered together within the context of the manipulation to identify which properties of the signals are most relevant.

To examine which properties are best inspected to identify user grievances and to evaluate the quality and usability of an interface, a series of three studies has been conducted which aim to mimic real-life interaction with common interfaces. These studies were still done in a controlled environment wherein we are able to gather ground truths about the interaction and to verify the validity and usefulness of our setup for interface evaluation. These studies will be covered in the following three Chapters.

Chapter 6

Applying This BCI To Interaction

6.1 BCI Objectives during Interaction

The bi-modal BCI proposed in Chapter 5 has been designed to allow for an easy transferability between interaction situations on a technical level. It also has been shown to allow access to properties which could contain relevant information about an interaction. However, it's actual applicability needs to be verified. Specifically two aspects need to be investigated. The first aspect aims to verify if the information made accessible through our BCI setup is useful for an in-depth exploration of interaction. As established in the two previous chapters, ET, EEG and ERPs change according to the interaction contexts. The informational value gained from the quantification of these changes needs to be investigated. The second aspect is intertwined with the first : the information gathered from the analyzed bi-modal properties has to remain useful in a variety of situations. Like the technical applicability and transferability of the setup discussed in the previous chapters, methods for analysis are needed which reliably provide information about the interaction and which are transferable between interaction scenarios.

To inform the development of methods which are both useful and transposable, a set of studies have been performed which place participants in common interaction situations. These situations are slightly adapted from their everyday counterparts to log all manipulation information properly, ensuring correct ground truths. To guarantee that the first results obtained in these studies are valuable to the pursuit of both aspects, a situation needs to be investigated which commonly occurs in a wide scope of interactions. By investigating simple visual searches, a set of initial methods for interface analysis can be established which provide useful information and are transposable to other interactions.

Visual exploration of an interaction environment is ordinarily performed at the beginning of any interface manipulation. Users typically require an understanding of the interface layout to orientate themselves within it. Visual exploration serves as an introductory phase to gather this understanding. This phase plays an essential part in the interaction and, while strongly guided by the layout and objective of an interface, relies only on eye movements rather than active manipulation (e.g. button presses, mouse click or more specific hand movements). Accordingly, examining this exploration phase through the proposed BCI, permits an in-depth contemplation of the interaction and allows us to define which methods are most helpful for its evaluation.

To explore the full range of information which can be extracted from the visual exploration phase, stimuli need to be presented which vary in their capacity of being accurately perceived and understood by users. As established in Chapter 2, interaction difficulties can take different forms. They can vary in terms of interface elements, scene composition, pursued tasks and difficulties relating more to the users themselves. In this first study on visual exploration, both element and scene related difficulties were examined in greater detail to help establishing a first set of methods for interface evaluation.

6.2 The Most Common Interaction : Visual Searches

Visual searches are among the most commonly encountered situations in daily life. When working with any interface, exploring it and localizing the relevant elements is one of the first tasks a user performs naturally [Ahlberg and Shneiderman, 2003]. A visual search thus proves to be one of the most appropriate interaction contexts for our BCI to explore concerning the informative value that the interaction holds.

Free visual searches are frequently explored cases in the literature. Similar bi-modal BCI have already been used to explore such situations [Kamienkowski et al., 2012]. However, in most of these applications, the explored scene proposed a limited and very controlled scope of stimuli to their users, to only study a specific research hypothesis. Many of such explorations thus consist in reading tasks [Baccino and Manunta, 2005] or propose the exploration of scenes containing predominantly faces [Kaunitz et al., 2014], such as in stadium tribunes. In some cases, free searches, which are closer to searches encountered in everyday life, are investigated. In these cases, the environments to explore were still often modified to induce specific potentials [Finke et al., 2016; Coco, Nuthmann, and Dimigen, 2020; Wenzel, Golenia, and Blankertz, 2016]. To investigate user grievances in a wider scope of applications, such constraints or adaptions need to be minimized in order to approach real-life like interaction.

To still introduce different kinds of difficulty into visual searches, different steps can be taken. When exploring an interface, one of the factors determining the ease to start the interaction relates to the visual saliency of relevant interface elements [Gong et al., 2015]. While the visual saliency of certain elements within the interface does not fundamentally change the interaction, it does influence gaze patterns [Henderson et al., 2007]. Visual saliency also expresses itself in the differences one interface element has to another. These differences relate to perceptual dissimilarities in element color, shape and orientation [Enns and Rensink, 1990] with the rest of the interface. Indeed, the more similarities are present between interface elements, the more visual attention is required to parse and understand the interaction appropriately. For example, a bright orange is much more visually salient when surrounded by blue coffee cups rather red apples. Furthermore, general scene composition can also have strong effects on the perception of difficulty in a scene [Ionescu et al., 2016]. To implement difficulty in this first study for our BCI, we change elements presented in an interaction scene in two different ways : the number of objects present and the color differences between objects.

Practically, when considering a simple visual search task, a certain type of interface element needs to be found. To increase difficulty on a scene and element, the quantity of interface elements is modified. Similarly, the color similarity between target and non-target elements has been changed.

6.3 Study One : Materials and Methods

Creating a visual search study follows a couple of rules. In the context of this study, a visual search task is presented to the user. The chosen task is to find a set of objects within a larger scene. This objective remained the same during the entire manipulation but the searched objects changed from scene to scene. The only measured aspect which changed during the different trials (i.e. scenes) is the number of present elements in the scene and the color similarity of objects contained within. This first study serves as an initial investigation into the usefulness of our proposed BCI setup. To ensure exact ground truths the scene objects were entirely created from computer generated stimuli, referred to as shapes. Indeed, while working with completely artificial shapes as interaction objects is uncommon in daily-life interaction, it ensures a better evaluation of the obtained results. The observation of controlled stimuli creates an environment in which a reliable set of method can be established which can be built upon in subsequent studies.

6.3.1 Experimental Protocol

The procedure the participants had to follow is illustrated in Figure 6.1. Participants were presented with a scene on a 1920x1080 pixel computer screen, which was split into two sections: a smaller "target scene" on the left side and a larger "search scene" on the right side. The search scene was initially hidden (see the second image in Figure 6.1). Both parts of the whole scene contained a variety of geometric shapes. The target scene contained between three and five non-overlapping shapes, while the search scene contained a much greater quantity of sometimes overlapping shapes (see third image in Figure 6.1). Participants had to locate the shapes presented in the target scene within the search scene. In detail, to select a shape, participants were told to fixate an individual target shape (measured with the Eyetracker) in the search scene and simultaneously execute a keyboard input (i.e. spacebar press) to validate their selection. A short sound, either a soft positive bell or harsh negative buzzer sound, was played depending on the validity of their selection. Concurrently, a small feedback image was also shown in the upper left part of the screen. If the selection was done correctly, the target shape was removed from both scenes (see fourth image in Figure 6.1). To ensure participants remained vigilant during the entire experiment, in rare cases (8% of scenes) some target objects shown in the target scene were missing in the search scene. In these cases, the participants had to perform the same selection procedure (i.e. simultaneously fixating and performing a keyboard press a second time) on the shape in the target scene, which was missing in the search scene. A scene was completed when all target objects were either selected or three minutes passed after the unveiling of the search scene. This three minute threshold was introduced to ensure that spontaneous decalibrations (e.g. due to sudden subject motions or eyetracker imprecision) would not make the scene uncompletable by the participant.

6.3.2 Scene Stimuli

The participants were confronted with a series of 138 scenes. These scenes varied in terms of color similarity between target and non-target objects (in 3 categories: close, medium and far similarity) and in terms of the amount of objects (in 2 categories: few and many objects) present (which includes how strongly the objects overlapped) in the search scene. An example of each of these 6 conditions is represented in Figure 6.2. 23 preconstructed scenes of each conditions were shown to the subject in a random order. To introduce participants to the task, a short tutorial session featuring ten additional scenes was presented prior to the actual recording session. This also allowed participants to familiarize themselves with the operation of the interface.

As mentioned above, the objects presented within these scenes were abstract geometric shapes. To reduce perceptual difference related to the size of the objects, the overall surface area of each object was scaled to maximally fit within a square bounding box. The scaled shapes were randomly selected from a larger list during scene creation. Shapes were also randomly rotated before being placed in a scene to minimize perceptual difference due to object orientation. The positioned shapes could be both convex or concave, ranging from simple (such as squares or circles) to more complex forms (such as half-moons or arrows). A uniform color was applied to each individual shape with no contouring color. The distance in color D between any two of these shapes (here labeled i and j) was determined by Equation 6.1.

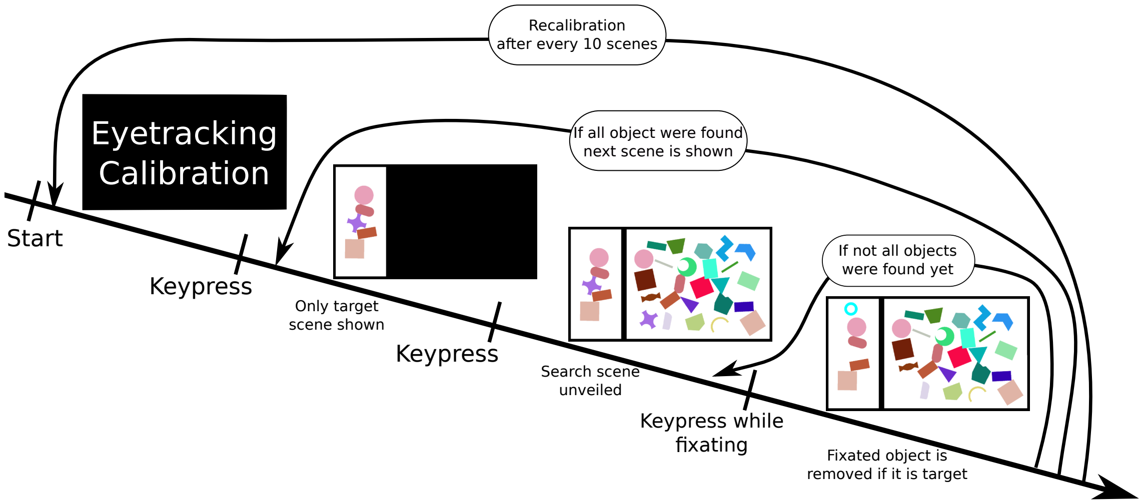


FIGURE 6.1: The experiment Procedure : First, a calibration of the eyetracker is performed. Once completed, the participant performs a spacebar press showing a scene. At first, only the target scene is shown. Upon performing another keyboard input, the search scene is revealed. The fourth image illustrates a correct selection, the participant performed pressing space while fixating on a target object in the search scene. This removes the shape from both scenes (here the purple cross) and displaying a feedback image (cyan circle) in the upper left corner. This process is repeated until all target objects are removed. After this, the next search scene is displayed

.

$$D_{ij} = \sqrt{2(R_i - R_j)^2 + 4(G_i - G_j)^2 + 3(B_i - B_j)^2} \tag{6.1}$$

With R_i,G_i,B_i corresponding to the red, green and blue values respectively of a selected color for a given shape i. This equation was chosen to approximate the human perception of color differences [Robertson, 1990]. In scenes featuring "close" color similarity, the value of distance D was kept between values of 0 and 200. This range changed to 200-475, and 475-765 to consider color similarities "medium" or "far" respectively. In scenes featuring many objects, the centers of each shape were placed at a Euclidean distance of at least 100 pixels apart. In scenes featuring few objects, this minimal distance was increased to 170 pixels, resulting in a presence of less objects in the scene to fit this condition. In practice, during the scene creation step, an object center was programmatically placed at a random location in the scene. In parallel, a scene mask was annotated with a circular area with defined radius (i.e. 100 or 170 pixels) around this center. Each following object center was procedurally placed at the edge of the marked areas and their associated circular area were added to the scene mask. This was repeated until the object areas filled the whole scene mask and no new object center could be placed in it. Occlusion between objects was not quantified, however, as expected occurred less frequently in scenes with a higher minimal distance between object centers.

6.3.3 Apparatus

EEG activity was recorded through 16 electrodes positioned at the locations Fz, Cz, Pz, Oz, F3, C3, P3, F4, C4, P4, PO8, PO7, F7, F8, T7 and T8 respective to the 10-20 system [Acharya et al., 2016] pictured in Figure 3.3. Additionally, horizontal and vertical Electrooculography (EOG) was also performed using 4 additional electrodes placed according to Figure 3.1. Two g.tec g.USBAmp biosignal amplifiers were synchronized via

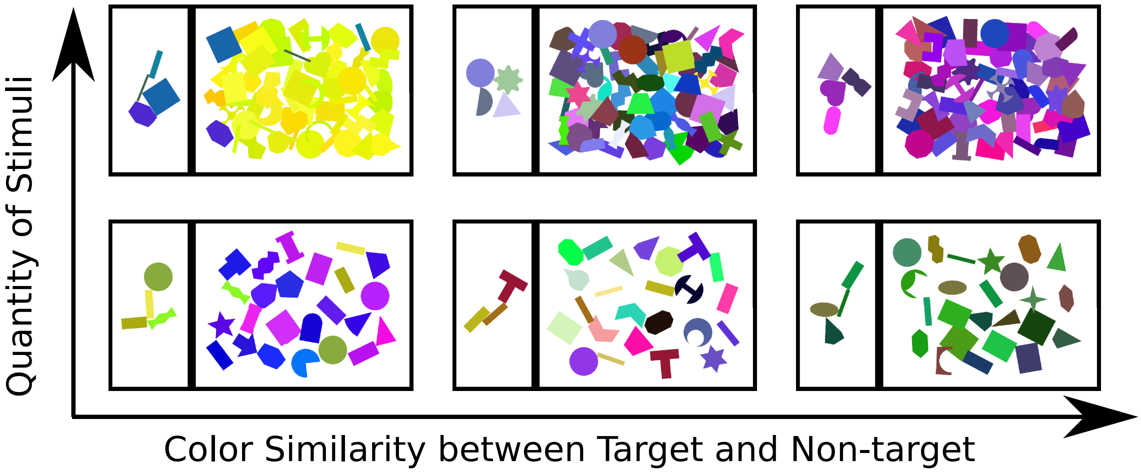


FIGURE 6.2: Examples for each of the six types of stimulus scene presented to the participant : top to bottom show the variation in the quantity of objects, left-right illustrates the color similarity between target and non-target objects.

g.INTERSync cable and used for the EEG recording (at 256 Hertz). Eye movements were recorded using an LC Technologies Eyefollower Desktop Lab Eyetracker. This desktop eyetracker, does not require a chin rest as it permits head motions in a 76 x 51 x 40 cm region between 46 and 97 cm away from the screen. In this region, the eyetracker returns gaze points with an accuracy < 0.4 degrees. The Eyetracker was positioned below the interaction screen and calibrated through a 9-point screen calibration at the start of the experiment. The setup was similar to the setup presented in Figure 6.3. The device was also recalibrated using the same procedure after every tenth scene was completed, or between any two scenes, if the subject deemed it necessary. Eye gaze samples were recorded with a frequency of 60 Hz for each eye in alternation (resulting in a sampling rate of 120 Hz when considering both eyes). Both devices were operated by two different computers whose clocks were previously synchronized. Participants were seated facing the screen at a distance of around 80 cm. As a result, their foveal region during the study encompassed a circular screen region with a radius of about 80 pixels.

6.3.4 Participants

Twenty-one volunteers participated in the study. The participants were aged between 19 and 38 years (mean 26.5 ± 5.0, 14 female). All were recruited from the local student, university visitors, and staff population and were either paid for their expenditure of time or granted course credit. All participants had no known prior or current pathological neurological condition (based on self-report). All participants had normal or corrected-to-normal vision. None of the participants reported color blindness of any kind. The experimental procedure and written consent form for this study were approved by the ethics committee at Bielefeld University, and adhered to the ethical standards of the Declaration of Helsinki [Williams, 2008]. All participants gave their informed written consent to participate in the study.

6.3.5 Artifact Correction and Preprocessing

Once a participant completed the entire 138 scenes (i.e. the end of Figure 6.1 is reached), their data were saved. Every information presented to the participant was also logged

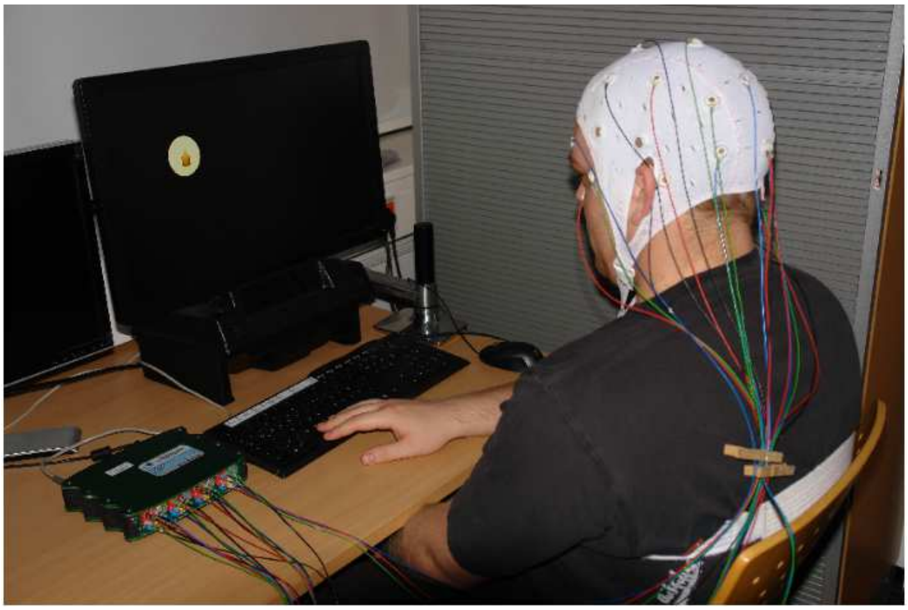


FIGURE 6.3: Example of how the recording sensors were set up. The eyetracker is placed below the interaction screen and the EEG is placed next to the subject. The interaction scene seen on the screen is not related to this presented work.

with the respective timestamps. This included which stimuli were presented to the subject, moments of action (i.e. spacebar presses), feedback shown, sounds played and performed recalibrations of the Eyetracker. The data recorded in this study can be obtained at the link : `https://pub.uni-bielefeld.de/record/2943918`.

Once logged, the artifact corrections discussed in Chapters 4 were performed on the data. Fixation locations and onset times were extracted from raw eye-tracking data of the subject's dominant eye. The fixation detection algorithm registered a fixation when the subject's eye gaze remained in a 5-degree eye-rotation angle area for at least 100 milliseconds. The EEG data were first filtered using a 0.1 Hz highpass and an 8th order Butterworth 45-55 Hz notch filter. This was done to remove low frequency artifacts as well as artifacts originating from device voltages. Subsequently, artifactual components, resulting from eye, head and other muscle movements, were removed from the signal using an Independent Component Analysis (ICA) performed on the EEG channels [Jung et al., 2000]. Artifactual ICA components were identified using the Multiple Artifact Rejection Algorithm (MARA) [Winkler et al., 2015].

Furthermore, due to potential desynchronization between recording modalities, the FCSync methods were employed. As discussed in Chapter 5 , the EEG and Eyetracker recording timestamps were adjusted for a static offset and a linear drift using our resynchronization method [Wobrock et al., 2019]. With artifacts and timestamp errors being removed, the data can now be analyzed.

6.4 Study One : Data Analysis

The data obtained from the bi-modal recording setup are scrutinized in different ways. As mentioned in the previous chapters, data from either modality are firstly examined separately. Both EEG and ET data on their own, contain information relevant to understanding user grievances and interaction difficulties. In this regard, however, ET data often hold less detailed information about underlying cognitive processes than EEG data. Separate analysis will be used to determine the general tendencies of interaction expressed during the manipulation. Once these individual modalities have been investigated, a more detailed interpretation of the data can be performed by examining the combination of both modalities such as through FRPs. The obtained results will subsequently be discussed to identify potential user grievances introduced in the data. There, our first set of BCI methods for the acquisition and exploration of insights into a user's perception of an interface will be developed. All the findings in this section are also published in Wobrock et al. (2020).

6.4.1 Eyetracking Results

As already addressed in the preceding chapters, eye movement data can be decomposed into fixations, saccades, blinks and smooth pursuits. As the participants were presented with static scenes, no meaningful smooth pursuits were recorded during scene exploration. Blinks happen naturally but are rarely influenced by common interaction paradigms [Jakob, 1998]. The remaining fixations and saccades offer a way to assess how and at which speed the participant explored the scene. From this information alone, inferences about the difficulty of the interaction and the layout of the interface can be made [Pretorius, Calitz, and Greunen, 2005].

As a starting point, one superficial metric which can already reflect the effort required to explore a scene is the average amount of fixations performed. The average number of fixation per explored scene type can be seen in Table 6.1. The number of fixations is highest (with 82 fixations on average) when participants are confronted with scenes containing many shapes presenting similar colors. The number of fixations is lowest (at 17 fixations on average) when the scene presents only few shapes and the colors between target and non-target stimuli are very dissimilar. Both of these dimensions of complexity influence the number of fixations required to explore the scene. Interestingly, the changes in fixation number according to changes in color similarity is much larger in cluttered scenes (37 and 18 fixation differences on average) than sparsely populated ones (7 fixation differences on average). This indicates that both dimensions of complexity (i.e. color similarity and number of objects) interact to guide the user's perception of the interaction.

TABLE 6.1: Number of Fixation performed on each type of scene presented to the participants. Data are averaged over all participants, with standard deviations given in parenthesis.

	Many Shapes	Few Shapes	Any Amount
Close Color Similarity	82 (± 50)	31 (± 15)	58 (± 45)
Medium Color Similarity	45 (± 29)	24 (± 11)	35 (± 25)
Far Color Similarity	27 (± 20)	17 (± 7)	22 (± 16)
All Color Similarity	52 (± 42)	24 (± 13)	39 (± 35)

Metrics based on average amount of fixations, however, face the problem that the absolute numbers of fixations already vary quite a lot between participants and scenes. This can be seen in the standard deviations depicted in Table 6.1. For an analysis which is in line with the goals of the proposed BCI (i.e. utility and tranposability of the methods), using metrics which are directly tied to the complete duration of a task presents a set of issues. The strong variability and ease of being influenced by external disturbance, which prolong the task, makes it difficult to utilize these absolute metrics effectively without a prior calibration. As such, using metrics which abstract the time domain could prove more beneficial for a quicker application of our BCI. Those can include metrics describing aspects of fixations, such as average fixation duration or the relative duration of fixations on a given visual area in the interface. This latter metric is notably used to create attentional heatmaps [Špakov and Miniotas, 2007], which allow to gain a clearer view for designers about which part of an interface matters most. Similarly, more information can also be extracted from saccades by looking at the saccade lengths (i.e. how far the gaze jumped between fixations) and how often the participant's gaze jumped back to already visited visual areas. In the context of this study, it is also possible to take into consideration the general layout of the UI. Therefore, the number of fixations onto the target and search scenes can also be considered.

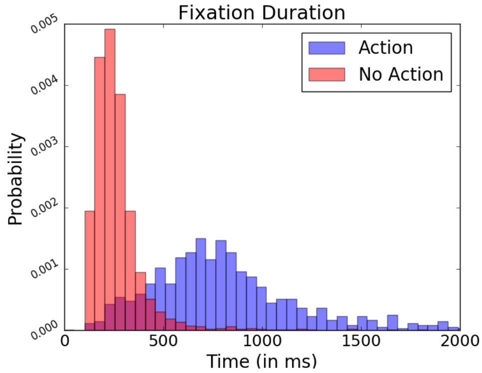


FIGURE 6.4: Fixation durations during the totality of the study, averaged over all considered participants. "Action" refers to the presence of keyboard presses during fixation.

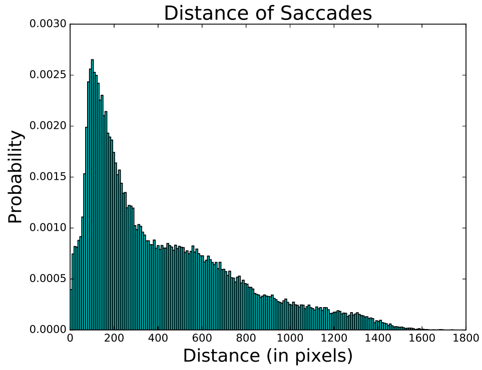


FIGURE 6.5: Saccades distances during the entire study, averaged over all considered participants. The maximum probability is located at 95 pixels.

When looking at the actual fixation duration, little difference can be observed depending on the category of a fixated scene. Indeed, while the amount of fixation seems to vary (see Table 6.1), the duration of fixations mostly stays the same. However, when considering the presence of keyboard presses during a fixation, one notices that fixation which occurred during keyboard presses are distinctly longer than action-less fixations. This is illustrated in Figure 6.4, where the histograms present a percentage distribution of fixation durations depending on the presence of simultaneous actions (i.e. spacebar presses). The different types of spacebar presses (i.e. for selecting a target object and selecting an absent object), were not differentiated at this point due to the rare occurrence of absent objects. In the case of action-less fixations, one notable aspect is that their duration follows a relatively narrow distribution centered around its maximum at 250 ms. This is a common duration for fixations [Baccino and Manunta, 2005]. It however also shows that most fixation epochs may be so short that common potentials (like P300 or N400) occur within the time window of the following fixation. As discussed in Chapter 4, this creates potential problems of overlap between epochs. On the other hand, fixation associated with keyboard press events show an average duration of around 700 ms, making them less prone to overlap, even if they present a broader distribution. Short fixations (i.e. fixation lasting for a duration of 250 ms or less) are also much more common than keyboard press associated fixations: short fixations represent 46.1 ($\pm$ 9.8) percent of all fixations. This value increases to 62.5 ($\pm$ 8.2) percent of all fixations if the threshold for a short fixation is increased from a maximum duration of 250 ms to a maximum duration of 300 ms.

Considering metrics relating to saccades, the distribution of saccadic distance also remains unchanged across scene categories. A histogram detailing their distribution across the entire study can be seen in Figure 6.5. The exploration of the scene is done using saccades of similar length disregarding how the underlying scene is organized. Here also, a maximum in the distribution of saccade distance occurs. It's value is located around 95 pixels, which corresponds to an eye globe turn of around 8 degree (see Figure 6.5). In contrast to the fixations, where fixation duration distribution (see Figure 6.4) generally decreases below its most common value, saccade distance distribution plateaus between 300 and 600 pixels. This shows that most saccades aim to explore the vicinity of the previously fixated region in this study rather than jump further away. Additionally, the distances obtained from looking at saccades give an indication about the size of the visual interface the user is currently operating on.

The participants frequently switched their attention between the target and the search parts of the entire scene. These eye movements are of particular interest as they inform about the moments when the participant sought out the instructions about which shapes they had to find. These moments of returning to the target scene are more commonly present when looking at scenes presenting higher degrees of complexity (i.e. a scene with more objects and less color difference between target and non-target shapes). However, while more switches between scene parts occur, the total time spend on the search scene is still much higher in scenes presenting more complexity. Indeed, while 70 % of the time is spent on the search scene part of simple scenes, it grows to 80 % in complex scenes. These results are presented in greater detail in Table 6.2. The standard deviation of this relative time spent on the search scene is kept around 10 % for all scenes categories. This indicates, that while variations still exist between participants and scenes, these are less prominent than variations in the amount of fixations as shown in Table 6.1 above.

TABLE 6.2: Average percentage of task time spent fixating on the search scene depending on general scene category. The standard deviation is placed in parenthesis

	Many Shapes	Few Shapes	Any Amount
Close Color Similarity	79.8 (± 6.8)%	72.1 (± 8.5)%	76.1 (± 8.6)%
Medium Color Similarity	75.5 (± 9.5)%	70.4 (± 10.4)%	73.1 (± 10.3)%
Far Color Similarity	72.5 (± 12.1)%	70.5 (± 12.7)%	71.5 (± 12.4)%
All Color Similarity	76.1 (± 10.1)%	71 (± 10.7)%	73.7 (± 10.7)%

More complex metrics could be considered to characterize the variation of gaze patterns across scenes. However, to keep this BCI for interface evaluation simple and transposable, additional more complex Eyetracking-based metrics are not considered. Indeed, the metrics considered so far can already serve to gain a general overview about the interaction. Further information about the interaction can be obtained when analyzing the data provided by the other BCI modality : the EEG recordings.

6.4.2 EEG Results

As mentioned in Chapter 4, data from the EEG signal can either be analyzed in terms of relative frequency bands or in terms of temporal ERPs. Using relative (and not absolute) frequency bands is necessary in these circumstances as the duration of the recorded signal (and consequently the magnitudes of values gained from its Fourier Transformation) changes between the scene categories. This difference in duration can be inferred by contrasting Table 6.1 and Figure 6.4. Figure 6.6 shows the average relative bandpower (through the colored bars) and its standard deviation. The overall frequencies recorded over an entire task change little from one condition to the other. The frequency profile does remain consistent with low frequencies being more present than higher one. This may indicate that participant attentiveness is not necessarily influenced by task difficulty.

It is to be noted, however, that bandpower standard variation (represented by the error bars in Figure 6.6) is high within scene categories. Thus, other aspects of interaction may strongly influence relative band frequency.

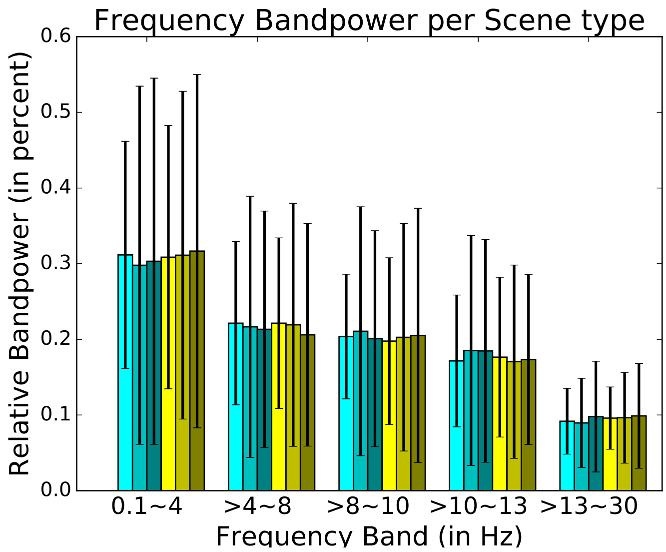


FIGURE 6.6: Relative Frequency bandpower from 0.1 to 30 Hz by type of scene. Blue corresponds to many objects, yellow to few objects. The lighter the bar color the more similar the target and non-target objects are in color. The bars represent the standard deviation of relative band power. Here data from all participants were used.

For a more rigorous analysis of the EEG signal, the observation of ERPs in the time domain can be considered. To observe ERPs, a type of interaction event is required. To this end, spacebar presses performed to select target objects were taken into account. Relying on such events, which would mean system adaptations have to be adequately logged, is technically against the spirit of transposability of the BCI stated in the motivation of this work. As these ERPs occur during an important moment of the studied interaction, investigating their properties is still desirable in this preliminary step to create adequate analysis methods.

Observing the EEG activity at the onset of keyboard presses shows the presence of certain prominent ERPs around the Pz electrode location. These potentials are represented in a visualization of the Pz EEG Channel in Figure 6.7. Regarding these potentials, the question arises : how is ERP latency and amplitude affected by the type of feedback presented to the user? As stated previously, users were made to hear a positive bell sound or a negative buzzer sound, depending on the validity of their input. To answer this question, these potentials were analyzed and showed changes in latency and amplitude depending on the type of feedback which was presented to the user after their keyboard press (as also seen in Figure 6.7). Interestingly, the correct selection feedback presents an earlier increase in EEG activity. An incorrect selection presents a much higher EEG activity but also a delayed one. These ERP are modulated by the feedback, being possibly connected to Error-related Potentials (ErrPs), as they give insight when the user perceives a mistake with the interaction.

From the results gathered through these analyses, the EEG signal has been shown to contain properties which allow to discriminate between different moments of interaction. So far, information about entire scenes as well as about specific moments of active interaction (i.e. keyboard presses) can be gained. However, as stated in the previous

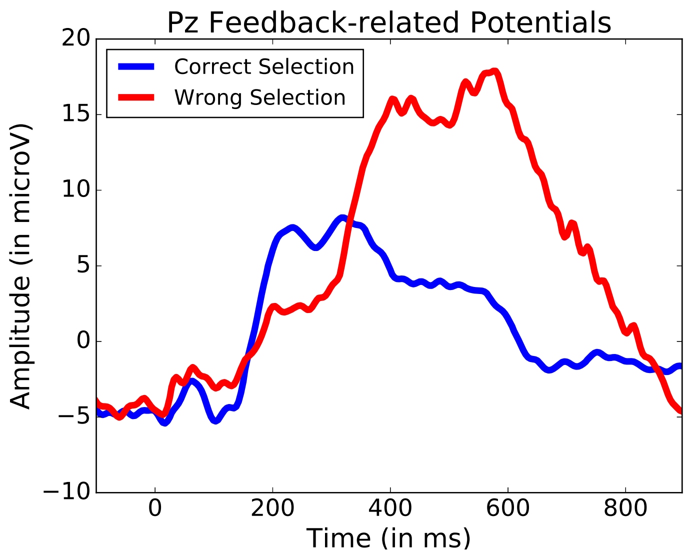


FIGURE 6.7: Event-related potentials occurring in the Pz EEG channel at the onset of the feedback related to a correct or incorrect selection of stimuli via spacebar presses. Data from only one participant was used.

chapters, our objective is to identify momentary changes across an entire manipulation, not only moments of targeted interest (e.g. element selection). Both of the discussed EEG evaluations are not sufficient to satisfy this goal. Other techniques are required to locate interesting EEG segments and to analyze them within the interaction context. As detailed in Chapter 4, the providing of context data, and thus the segmentation can be done by combining ET data with the EEG signal in the form of FRPs. We thus investigate FRPs and their properties to determine if differences between other, less prominent moments of interaction can be identified.

6.4.3 FRP Results

Within this study, when considering FRPs, the EEG signal was segmented into 1 second long epochs starting 200 ms before fixation onset. As discussed in the previous chapters, when observing FRPs, two main approaches can be taken to inspect the FRPs' features. The first approach consists in identifying the peaks present within the considered epochs' grand averages. Here, one of their most prominent features is a noticeable positive peak at 100 ms after fixation onset in the occipital Oz channel. This peak is not always noticeable on single epoch level, but becomes visible when averaging multiple epochs together. The P100 potential remains clearly present and visible when averaging any subset of epochs, even those associated with short fixations times (i.e. with a duration inferior to 250 ms). In all these averaging of subsets of epochs, the P100 potential is the most prominent potential but it isn't the only one which is present. Another potential is present in the frontal channels of the EEG. This N100 potential presents a negative peak occurring in the occipital P100 potential at around 100 ms after fixation onset. Both of these potentials and their location on the scalp are illustrated in Figure 6.8.

The second approach when looking at ERPs consist in looking at the variations in amplitude, latency and morphology of a peak during the study. These properties are obtained through the aMLRd method detailed in Chapter 5. When looking at the variation of the P100 peak across the various categories of scenes, a distinct difference in the amplitude of the P100 peak can be noticed in the grand averages. The more complex and

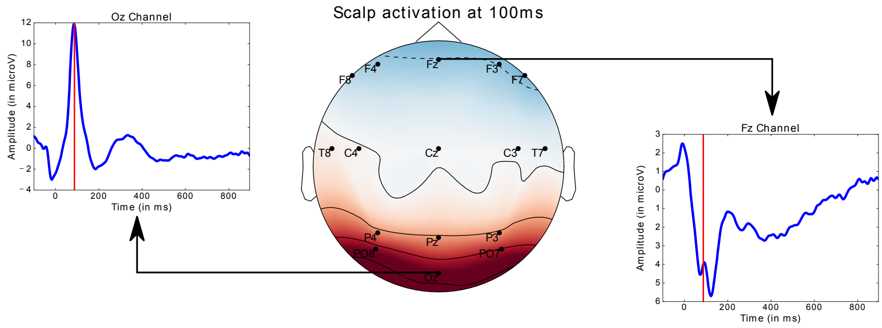


FIGURE 6.8: Averaged Fixation-related potentials recorded during the entire study at the Oz and Fz channels

difficult a scene is organized, the higher the amplitude of the P100 potential gets. This changing in the amplitude of the P100 potential is also strongly present when observing the contents of the foveal region (i.e. the circular area of 80 pixel radius around the center of the ET-returned fixation location) during a fixation. When this foveal region, where they perceived elements clearly, contained an increased number of different objects, the P100 amplitude increased in accordance with the number of objects. This phenomenon is shown in Figure 6.9, where an increase of objects in the foveal region results in a higher P100 peak amplitude. This same phenomenon was also observed when the foveal region contained more colors which were clearly distinct from one another. Similarly, this is illustrated in Figure 6.10. In this context, a clearly distinct color translates to a color distance D (from Formula 6.1) larger than 200 to its neighboring objects.

As opposed to the P100 potential, the frontal N100 potential does not increase or decrease in amplitude consistently with the number of objects or colors present within the currently fixated foveal region. Indeed, while the N100 is the only other FRPs consistently present in epochs, it does not seem noticeably influenced by the various complexity dimensions introduced into the interaction at first glance. This lack of change to the P100 according to these dimensions can be seen in Figures 6.11 and 6.12. For this reason, the N100 will not be explored further in this work.

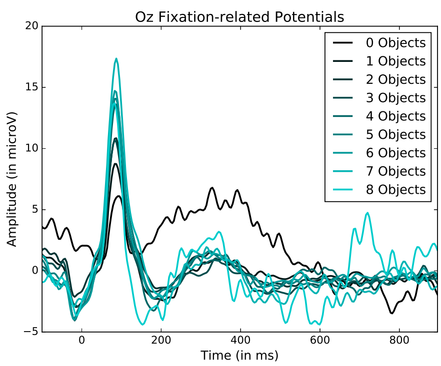


FIGURE 6.9: P100 Fixation-related potentials variation along the quantity of objects present in a fixation

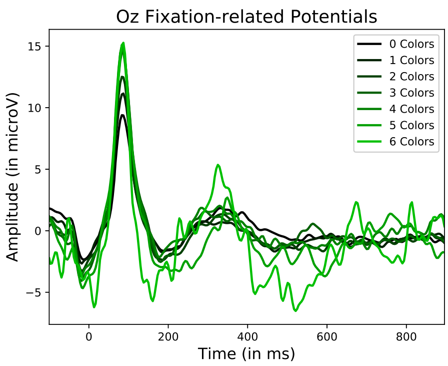


FIGURE 6.10: P100 Fixation-related potentials variation along the quantity of colors present in a fixation

Unfortunately, the P100 amplitude occurs at a very early moments of the fixation, meaning only a limited amount of information is available, and can be expected to be related to preattentive perception about the interaction rather than deeper insights about

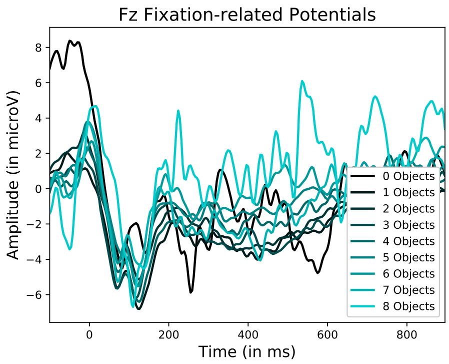


FIGURE 6.11: N100 Fixation-related potentials variation along the quantity of objects present in a fixation

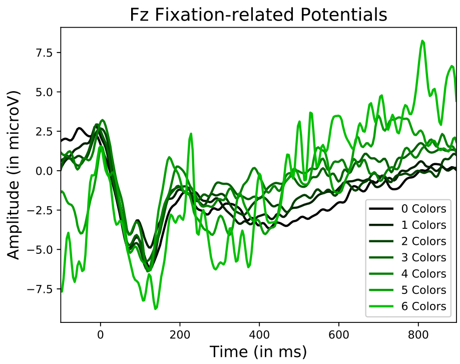


FIGURE 6.12: N100 Fixation-related potentials variation along the quantity of colors present in a fixation

the user's decision process. Fixations can strongly vary in duration as seen in Figure 6.4, making it difficult to identify and compare later ERPs contained within the rest of the FRP epochs. The high number of shorter fixations indicate the exploratory nature of the interaction. Fixation associated with manipulation tasks are longer and subsequently may allow for the analysis of these later potentials. However, even in these cases, due to their duration and the associated strong standard deviations, it remains difficult to observe potentials occurring after 500 ms (see Figure 6.4). With the ErrPs detected previously (see Figure 6.7) occurring at varying moments during a fixation, they are difficult to accurately locate with these FRP epochs. Nevertheless, with these first observations from the EEG, ET and FRP data, a first set of methods can already be created to evaluate the user's perception of the interface.

6.5 Methods for Interface Evaluation

From all the gathered data and the above early analysis, different types of information about the interaction have been accessed. All three ways of probing the data (i.e. ET, EEG and FRP) demonstrate that insights about the user's perception of the interaction can be gathered. Accordingly, the obtained information has to be analyzed in different ways to gain information about the underlying interaction. More specifically, to adequately retrieve theses relevant insight, the information needs to be categorized according to the type of difficulty or grievances the user faces. Among the three sources of information, general scene related difficulties can be examined using the ET-extracted information. Similarly, FRP properties can be compared to the contents of the fixated foveal region to obtain information about local interface complexities. Finally, the information contained within the EEG data frequency bands can help categorize moments of interaction which permits to contextualize moments of difficulty that the user faces. All three types of user grievances can be used differently to advance evaluation. Below three methods are proposed which translate the data into statements about these three user grievances. These three methods are : the General Difficulty via Eyetracking (GDET) method, the Stable Peak Property Quantification (SPPQ) method and the Segment Frequency Bands Analysis (SFBA) method.

6.5.1 General Difficulty via Eyetracking (GDET)

As discussed previously in Table 6.1, the number of fixations performed on a scene varies with the different dimensions of the scene organization (i.e. the number of objects and the color similarity between objects). The more objects are present and the closer they are to each other, the more complex the scene gets. The more the complexity of the scene grows, more the amount of fixations during a scene also increases. This correlation indicates a connection between the gaze patterns and the scene complexity. If the overall objective is to optimize interaction efficiency, the amount of fixations needed to complete an interaction scene should to be minimized. To accomplish this, the obtained results indicate that the dimensions of color and element number should not be cumulated. However, such observations are obtained by looking at metrics which are difficult to transpose between interactions. Notably the time and number of fixations change significantly between participants, scenes and tasks [Castelhano, Mack, and Henderson, 2009] and thus the interaction as a whole. To minimize the difficulties that these metrics pose for quick application and generalization (e.g. requiring calibration to know which completion time is normal for a participant), time independent ET metrics can be applied to estimate the perceptual complexity of a scene to the user. This is the objective of our first evaluation method. Time independent metrics, could be compared among participants without requirement to calibrate the BCI to a specific individual's average performance. To identify which measures are most effective for this, different fixation metrics have been compared.

When comparing different fixation metrics, three of them have been identified which can provide information about the general layout complexity of the observed scene in a time independent fashion. The first of these metrics (M1) is the relative amount of longer fixations. Longer fixations, in this case, refer to fixation periods lasting longer than 250 ms, due to the distribution observed in Figure 6.4.

The second metric (M2) concerns the amount of new visits of visual regions. While the first metric informs abstractly about the temporal cadence of the exploration, this second aims to inform about the spatial cadence of exploration. Similar to the object regions used to create the stimulus scenes, a visit of visual region means that the user fixated a location of the interaction scene. When fixating, we annotate a circular region of 80 pixel diameter centered around this location in the scene. This annotated region is termed a visual region. If the participants fixate outside of any visual region, a new visit of a visual region is assumed and a new visual region is labeled. When the participant's gaze jumps outside of one visual region and later back into it, the fixation is labeled as a revisit of a visual region. If two successive fixations are performed in the same visual region, the second one is neither labeled as a visit nor as a revisit. Looking at the ratio of revisits to visual regions indicates how many times the user had to reinspect already explored regions of the scene.

The last chosen metric (M3) concerns the relative time the participant spends fixating the target scene (see Table 6.2). Indeed, this part of the whole scene informs about how often the subject went back to inspect the instructions and how long participants spend exploring the search scene. More abstractly, this metric is chosen to represent a specificity of the studied interaction and should be replaced by other similar metrics in other interaction scenarios. The three chosen metrics are thus :

- M1 : the relative amount of long fixations (i.e. fixations longer than 250 milliseconds [Hooge and Erkelens, 1998]).
- M2 : the relative number of times the subject revisited previously fixated regions (i.e. fixation within the foveal region of any prior fixation)

- M3 : the relative number of fixations associated with a task specific element (e.g. here, the percentage of fixations performed on the search scene)

These three metrics encode different aspects of the process of exploration of a presented scene. To validate if these metrics help to identify the different levels of scene layout complexity, a visualization of their distribution in the feature space was performed. This was done by using the t-stochastic neighborhood embedding (t-SNE) [Maaten and Hinton, 2008] algorithm utilizing the three metrics as input features. Practically, the t-SNE algorithm was given a two dimensional matrix having the same number of rows as scenes and 3 columns as input. The second dimension (the columns) consisted of the three previously discussed metrics. This algorithm aims to visualize high dimensional data in a way to make natural clusters more visually apparent. The results can be seen in Figures 6.13 and 6.14 showing the same representation but different labeling of the output data. In the first representation (Figure 6.13), coloration indicates the amount of objects in the scene. In this representation, the clusters associated with many and few shapes show barely any overlap. Indeed, in terms of number of objects presented within a scene, many and few objects are very much distinct in these metrics. On the other hand, in the second representation (Figure 6.14), coloration is performed according to color similarity. Differences observed between color similarity are difficult to distinguish between close and medium color similarities. Far color similarities between target and non-target objects are however graphically more distinct from the other clusters. This indicates that the far color similarities are perceived differently from close or medium color similarities by the user. When improving the interface, this insightful representation can be used to adjust color differences between interface elements to make user progress faster or slower.

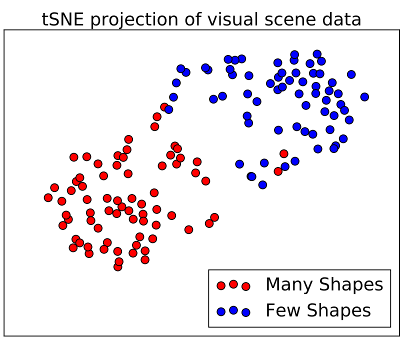


FIGURE 6.13: t-SNE projection of Fixation Data, colors distinguish object quantity

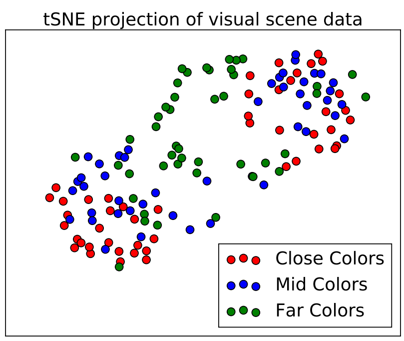


FIGURE 6.14: t-SNE projection of Fixation Data, colors distinguish color similarity

To formalize this type of scene complexity discrimination and to make it more easily usable for other situations, a classifier was trained and tested using these three metrics as inputs. Practically, this was done by performing a multi-class LDA classification. The results were verified using a 5-fold crossvalidation. The resulting accuracies presented in Table 6.3, confirm the conclusions gathered from the t-SNE projection. The classification rates show over chance level accuracy for all differentiations. While proving that differentiations between scene categories can be done thanks to these metrics, they are not accurate enough to reliably and precisely identify which type of scene was observed. Nevertheless, evaluating their classification score helps to identify how complex a scene appears to the user. This overall allows a probing of the general difficulty of an interaction.

TABLE 6.3: Multiple linear classification applying the combination of 3 visual time independent features.

	Classification Accuracy	Chance Level
Discriminating Number of Objects	69.18% (± 3.86)	50% (2 Classes)
Discriminating Color Similarity	53.63% (± 7.75)	33.33% (3 Classes)
Discriminating All Scene Types	26.28% (± 3.48)	16.67% (6 Classes)

Overall this analysis of ET data presents a promising and useful way to gauge the complexity between different scenes (and possibly even between users) without requiring prior calibration of any user averages of these metrics. The utility of this information and how to integrate it into various interfaces has to be investigated on a case-by-case basis. This three metric approach averages data over an entire scene. As such, it only provides a general information about the interface layout out of a specific scene and subsequently about general scene related difficulties. In the rest of this work, we will refer to this method of interface analysis as the General Difficulty via Eyetracking (GDET) method. Below we formalize it in Algorithm 4. To explore other types of difficulties, quicker changes should to be examined and integrated in the evaluation framework. As mentioned, properties in FRPs can provide more information about individual scene elements.

6.5.2 Stable Peak Property Quantification (SPPQ)

The amplitude of the P100 peak has been shown to vary with the objects contained within the fixated foveal region (see Figures 6.10 and 6.9). The amplitude of the P100 peak is influenced by the attentional effort put into processing the contents of the fixation by the participants [Luck, 2014]. Attentional effort is comparatively less intuitive to incorporate into the interface evaluation than the previously discussed general scene complexity. To analyze the utility of the P100 potentials, their amplitudes can be extracted from all fixation epochs using the aMLRd algorithm detailed in the Chapter 5. The P100 was also analyzed in terms of latency and morphology but no significant differences were examined between the categories, due to their overall distribution being very narrow. The case of absence of latency variation is notably due to the FCSync synchronization method used between ET and EEG data. However, looking at the variation in P100 amplitude for different conditions, significant differences can be observed. These differences are represented in the boxplots featured in Figure 6.15. For information, the red central line of each boxplot shows the mean value of the distribution. The upper and lower ends of the box mark the second and third quartile of the distribution respectively. The upper and lower horizontal strokes outside the central box mark the general range of the distribution. The other markers (i.e. circles, or crosses in later boxplot figures) show the outliers of the distribution.

Different types of scenes can present very different amplitude distributions of the P100 peak. In Figure 6.15, four pairs of epoch subsets were created and the P100 amplitudes compared within each pair (represented by the brackets linking boxplots together). As indicated by the previous Figures 6.10 and 6.9, higher P100 amplitudes are recorded in FRPs related to fixations performed on regions featuring many objects and many different colors. When applying a Student t-test onto both the set of epoch pairs dividing scenes by number and color of objects, p-values below 0.001 were obtained (as marked by the two ** brackets in Figure 6.15). These two pairs of distributions can be considered as containing many of the same fixations since the presence of many different colors

Algorithm 4 General Difficulty via EyeTracking

Creating GDET Feature Set;
An interaction has N scenes
F, the set all of fixations occurring during interaction
Set T_{dur} ▷ Long fixation duration threshold, often 300 ms
Set T_{dis} ▷ Fixation distance threshold, often 80 pixels
Set $A_{cond}(x_{location}, x_{duration})$ ▷ Condition function specific to the interaction
$S_{GDET} \leftarrow []$ ▷ Initialize GDET Feature set
for $i \leftarrow 1$ to N **do**
  Scene i contains M fixations
  $L_{visited} \leftarrow []$ ▷ Set of visited locations
  $n_{long} \leftarrow 0$
  $n_{revisit} \leftarrow 0$
  $n_{condition} \leftarrow 0$
  for $j \leftarrow 1$ to M **do**
    $f_{loc}, f_{dur} \leftarrow F[i, j], f_{loc} \in \mathbb{R}^2, f_{dur} \in \mathbb{R}^+$ ▷ Location and duration
    $n_{condition} += A_{cond}(f_{loc}, f_{dur})$
    if $f_{dur} \geqslant T_{dur}$ **then**
      $n_{long} += 1$
    end if
    for $k \leftarrow 1$ to $L_{visited}$.size() **do**
      if $T_{dis} < \sqrt{(f_{loc} - L_{visited}[k])^2}$ **then**
        $n_{revisit} += 1$
        $L_{visited}$.insert(f_{loc})
        break; ▷ ends For loop iteration
      end if
    end for
  end for
  S_{GDET}.insert($[\frac{n_{revisit}}{M}, \frac{n_{long}}{M}, \frac{n_{condition}}{M}]$)
end for
Use GDET Set
Plot tSNE(S_{GDET}) or classify S_{GDET} according to other scene labels

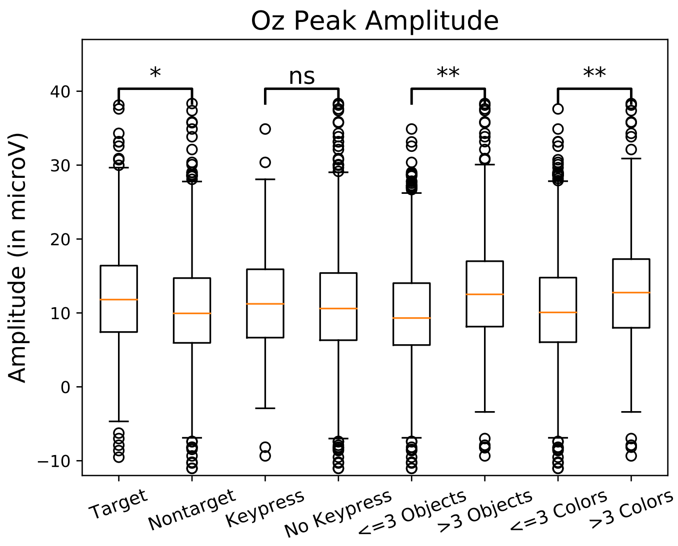


FIGURE 6.15: Variation of P100 amplitude across different conditions of fixations. Each bracket above the boxplots compares two distributions via the Student t-test.

implies the presence of many shapes. Nevertheless, information about the visual complexity of scenes is still reflected within the P100 amplitude. Such information could be used to identify which visual parts of the interface are particularly hard to parse by the user. These visual regions could then be simplified accordingly.

Additionally, a difference between fixations on target and non-target objects can be observed from the P100 amplitude distribution. While less significant than color or object quantity variation, this difference still scores a t-test p-value inferior to 0.01 (as marked by * in Figure 6.15). As such, the attentional cost of a fixation could not only be attributed to the quantity of visual information within the foveal region but also to the relevance of this information to the task at hand. While the meaning of information relevance changes drastically between interaction scenarios (and making it more difficult to transpose), the pre-attentive nature of the P100 potential may indicate that the ease with which a target stimuli is identified depends on a priming of the user to this target stimuli. This aspect needs additional research to be verified however. Finally, the P100 amplitude did not show a significant amplitude difference when comparing fixations during which the participant performed a spacebar press or not. This difference scored a t-test value superior to 0.1 (as marked by ns in Figure 6.15). This confirms that the P100 amplitude mostly varies due to pre-attentive observations but does not contain information about how the participant acts in accordance with the perceived stimuli.

The P100 potential occurs early after fixation onset. It's ability to reflect attentional cost as well as pre-attentive perception of the environment gives information about the affordance of the interface. When applied to the different zones of the interface, its properties may thus help to identify how difficult it is for the user to adequately and quickly parse the contents of each region of the environment. In practical terms, the P100 amplitude can be used as a reference point to tune the complexity of scene regions, making visual elements easier or harder to parse depending on the attention which is demanded from the user at this time [Kahneman, 1973]. If the scene is too easy, users may get bored and reduce their attention, potentially resulting in more errors later on. For the interaction scenario considered in this experiment, this can be done by increasing or reducing the amount of interface elements in a given region or by making the color of certain elements more salient when compared to the surrounding interface elements.

This aspect of the P100 amplitude should be explored in greater detail in the following studies as the relationship between interface affordance and attentional cost can take many different appearances depending on the considered interaction scenario. In the rest of this work, this method of analysis of FRPs will be designated as the Stable Peak Property Quantification (SPPQ) method. Below, the SPPQ is detailed in Algorithm 5. Both the GDET and SPPQ approaches used so far were related to eye movements and exploration of the environment. However, more general task changes may also occur during interaction which can be logged and analyzed to the benefit of the evaluation.

Algorithm 5 Stable Peak Property Quantification

Creating SPPQ Feature Set;
An interaction has N fixations
$S_{SPPQ} \leftarrow []$ ▷ Initialize SPPQ Feature set
$X \leftarrow [x_1, x_2, ..x_N], \forall i, x_i \in \mathbb{R}^M$ ▷ FRP epochs with M EEG samples (from one channel)
aMLRd.selectPeak(X, $P100_{bin}$)
aMLRd.createTemplateAndRegressors(X)
for $i \leftarrow 1$ to N **do**
    $x_{aMLRd} \leftarrow$ aMRLd.getPeak(x_i)
    if Maximum present around 100 ms **then**
        $amp_x \leftarrow \max(x_{aMLRd})$ ▷ P100 Amplitude
        $lat_x \leftarrow argmax_t(x_{aMLRd})$ ▷ P100 Latency
        $l_1, l_2 \leftarrow argmin_x|\frac{d^2 x_{aMLRd}}{dt^2}|$
        $mor_x \leftarrow l_2 - l_1$ ▷ P100 Morphology
        S_{SPPQ}.insert($[amp_x, lat_x, mor_x]$)
    end if
end for
Using SPPQ Feature Set;
1 : Statistical tests between available fixation categories
2 : Classifications between available fixation categories

6.5.3 Segment Frequency Bands Analysis (SFBA)

As seen in Figure 6.7, aside from FRPs, other non-fixation-related potentials are also present in the EEG signal. These potentials are related to keyboard presses. By using epochs extracted from the onset of keyboard presses, a binary classification by means of the LDA algorithm can be used to determine if the feedback received from the press was positive or negative. The classifier was tested using a 5-fold crossvalidation which resulted in an accuracy of 82.2% ($\pm$ 4.8). This result is however hardly transposable between interfaces, as interactions may be different in terms of how object selection is handled (e.g. they may not give auditory feedback). Active inputs or movements nonetheless occur frequently during interaction and change significantly the meaning behind fixations [Land and Hayhoe, 2001]. Indeed, by identifying the presence of actions during a visual exploration the understanding of the goal of the associated fixation changes. The nature of the fixation changes from a fixation aimed only for exploring the interface to a fixation for monitoring a choice. This implies that different cognitive processes were present.

By looking at the duration of fixations, a noticeable difference can already be observed depending if a button press occurred during the fixation or not (see Figure 6.4).

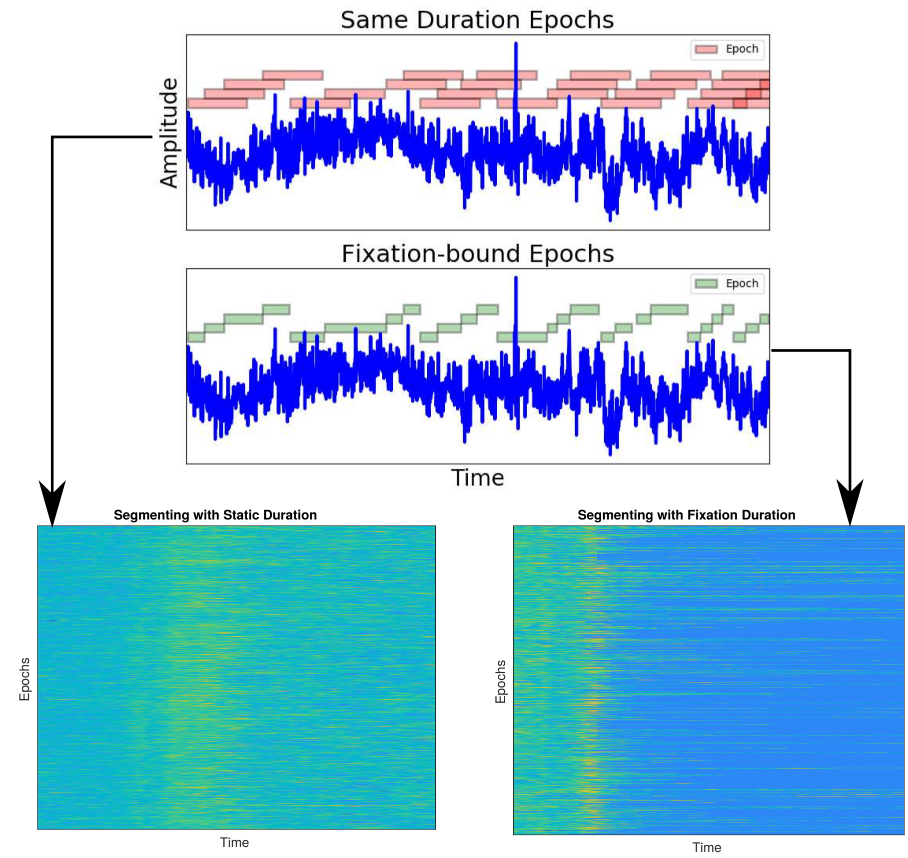


FIGURE 6.16: Different ways of segmenting the EEG. The lower left figure shows a regular segmentation at fixation onset with constant epoch length (here one second). The lower right figure shows a segmentation bound to individual fixations. The time axis of the both bottom figure represent one second of signal.

This difference in fixation duration distribution is quite significant but an overlap persists between action and non-action fixations, meaning that a clear identification between key press and exploration fixations is not always possible. A different method is thus required to identify which purpose the fixation served. The problem could be avoided by logging the keyboard presses, which is mostly easy to implement. However, what actions the user performs is different depending on the considered interaction. Accounting for all possible actions across all possible interactions is costly and necessarily introduces system-side adaptation workload which we aim to avoid.

One possible approach to still discriminate between fixations relies on the fact that, here, ErrPs present in the signal induce a strong change in frequency composition of the EEG signal. However, these ErrPs are often not contained in the regular duration of FRP epochs. As such an alternative way to segment the signal can be used. Rather than constructing epoch featuring the same length, the signal can be segmented according to fixation duration. This is done in Figure 6.16, where the upper plot shows a segmentation into same length epochs at fixation onset, while the lower plot segments the signal into fixation-bound epochs. In this Figure, both plots show one second of signal for reference. In practice, the fixation bound epochs are truncated at the end of the fixation. This can occur before or after the end of the fixed length epoch (see beginning of the deep blue color in Figure 6.16). When observing the EEG in a fixation-bound epoch fashion (see Figure

6.16), the motor action is guaranteed to be present within each segment (as the keyboard press needed to occur while fixating) and exclusive to that epoch. Looking at the relative frequency bands presented by each type of fixation-bound epoch (i.e. containing action or not), a significant difference can be observed between both categories. This is represented in Figure 6.17 where the band powers of fixations featuring keypresses are compared to those without keypresses. The epoch containing an action (i.e. spacebar press), presents a higher amount of high frequencies with respect to non-action epochs. In both of these cases, the frequency standard deviation, marked by the error bars, remained minimal.

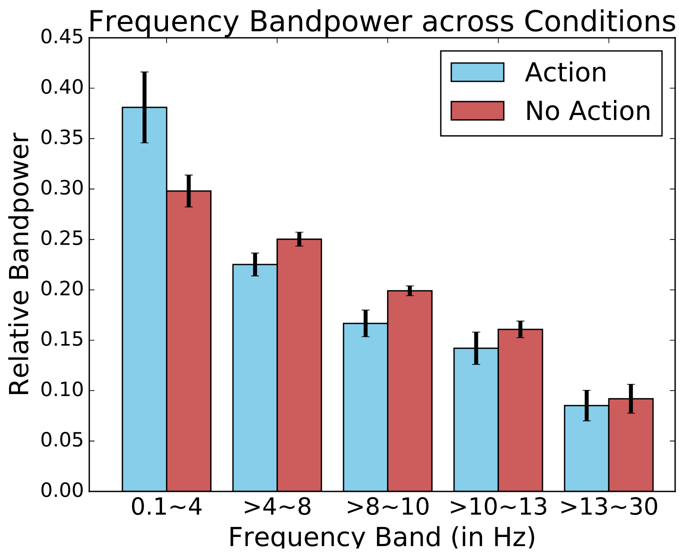


FIGURE 6.17: Variation in relative bandpower depending if an action took place during the fixation interval or not.

By using these frequency bands relative magnitudes as features in the classifier, the identification of epochs which contain actions become possible. Indeed, by performing a 5-fold balanced LDA classification using frequency bands as features an average accuracy of 83.11% (± 4.71) across all subjects is observed. One reason for the strong differences between both categories may also be due to the presence of potentials relating to motor activity. Indeed, these potentials often display a prominent low frequency resulting in a singular potential peak which stands out among the other EEG activity recorded during manipulation (see Figure 6.7). This activity could however also stem from event-related desynchronization (ERD) being caused by motor activity [Pfurtscheller et al., 1997].

Whatever the case, a difference in frequency bands (here mostly in 0.1 to 4 Hz) allows to locate quite effectively moments of interaction during which more noticeable activity occurred than is possible when simply looking at regular same length signal epoching. The fixations identified by this method can thus be given a more precise context. Indeed, as see Table 6.4, identifying these fixations can help to understand other effects the scene complexity has on the user. Here, both dimensions of complexity, influence the number of actions (and subsequently wrong selections) the user performed. At the highest complexity (i.e. many objects with close color similarity), the participants performed the most actions. Additionally, the fixation-bound epochs can be further analyzed to possibly identify later types of Event-related Potentials. This could help to detect the location of Error Potentials which occurred during them. However, the discrimination between ErrP types (as done in Figure 6.7) would still be difficult when the exact action onset cannot be determined accurately through this approach.

Overall this technique helps to locate the fixations during which motor actions occur. For brevity, in the rest of this work, this method will be designed as the Segment Frequency Bands Analysis (SFBA). This method is further detailed by Algorithm 6. To improve the evaluation, knowing the presence of these fixations may help to gain a more complete image of the interaction and, more specifically, it indicates which elements of the interface are monitored for active interaction and which are only fixated for explorations. By adding this approach to the two previous evaluation methods regarding scene and element difficulty, task complexity can also be evaluated to a certain extent.

TABLE 6.4: Number of actions performed on each type of scene presented to participant. The shown data are averaged over all participants.

	Many Shapes	Few Shapes	Any Amount
Close Color Similarity	6 (± 0.74)	4.6 (± 0.48)	5.3 (± 0.58)
Medium Color Similarity	5.2 (± 0.73)	4.7 (± 0.55)	5 (± 0.55)
Far Color Similarity	5 (± 0.74)	4.4 (± 0.28)	4.7 (± 0.5)
All Color Similarity	5.4 (± 0.66)	4.6 (± 0.4)	5 (± 0.52)

Algorithm 6 Segment Frequency Band Analysis

Creating SFBA Feature Set;
$X \in \mathbb{R}^M$ ▷ Full EEG Signal
$S_{SFBA} \leftarrow []$
$F \leftarrow [0.1, 4, 8, 10, 13, 30]$ ▷ Initialize SPPQ Feature set
Interaction contains a series of N same type events
$\forall i \in [1, N], \{ts_i, te_i\}$ ▷ Start and End times of events
for $i \leftarrow 1$ to N **do**
  $x_{seg} \leftarrow X[ts_i : te_i]$
  $f_{rels} \leftarrow [0, 0, 0, 0, 0]$
  for $j \leftarrow 1$ to 5 **do**
    $f_{rels}[j] \leftarrow \sum_{freqs}(\|FFT(x_{seg})[F[j] : F[j+1]]\|) / \sum_{freqs}(\|FFT(x_{seg})\|)$ ▷ Calculate Relative Bands
  end for
  $S_{SFBA}.insert(f_{rels})$
end for
Using SFBA Feature Set;
1 : Statistical tests between event categories
2 : Graphical Comparisons (t-SNE) between event categories
3 : Classification between event categories

6.6 Examining the Methods

The goal of this study was to establish methods, which can be employed to analyze the data of an EEG/ET BCI setup, that are both effective for evaluating interaction situations as well as being transposable between these situations. We have proposed three methods (i.e. GDET, SPPQ and SFBA), all of which are not reliant on system-side information. These three methods provide insights into various difficulties surrounding the interaction. Notably, user grievances on scene element level, on whole scene level and on task level can be identified.

6.6.1 Application for User Grievances

The first method which was established helped with the examination of difficulties associated to scene variation during manipulation. The GDET method relies on different metrics gained from eyetracking. These metrics are meant to be time independent. Indeed, as regular interaction time varies strongly between tasks, users and environments [Szalma et al., 2004], relying on time independent metrics makes information easily transposable. Here, transposability is only hindered by the use of the search/target scene labeling. Broadly labeling the scene is however a quite inexpensive adaptation of the BCI system, only requiring to mark static regions before manipulation. Overall metrics inform about the general gaze patterns employed by the user to explore the scene and to gather information. Similar to common eyetracking analysis, the method offers a generic approach to inspect interaction and estimate the general difficulties encountered during exploration of an interface in a scene by scene fashion. As such, it allows to identify and to compare the perceived complexity of scenes within a larger manipulation in a quantifiable way. This gives designers reference points when considering interface improvements.

The second method which was established considers the peak properties (amplitude, latency and morphology) of the P100 potential. This method is called the SPPQ method. Only the P100 amplitude was shown to provide information about the interaction so far, among other FRPs and their properties. Indeed, the P100 amplitude correlates to attentional cost associated to a fixation [Desmedt and Tomberg, 1989]. Thus, it allows to examine how much effort it took the user to explore the scene. Accordingly, this amplitude can be employed to locate the regions of varying saliency. While such information individually may not be significantly different from eyetracking analysis, it offers an added value: while eyetracking analysis requires multiple revisits, through the use of the aMLRd method on single trial FRP, information about visual processing difficulties can be obtained from a single visual visit of the scene region. This helps to identify visual regions which are difficult to parse for the user, more quickly and clearly. This method could thus allow to tune scene composition more effectively to the desired interaction. The P100 amplitude offers a metric to quantify and potentially reduce visual effort when the scene is too complex or increase it when the scene is too simple as to speed up or slow down scene assessment [Cooper et al., 2014]. Furthermore, differences between target and non-target objects were identified using this method which need to be investigated further. However, in this study, as the number of fixations grew with the complexity of the scenes, examining target/non-target object P100 amplitude differences may be skewed by this scene complexity. This aspect of the P100 amplitude will be investigated in the following studies.

The last SFBA method looked at information obtained from fixation-based segmentation of the signal. By using this method, the presence of other types of events can be determined without relying on the precise onset of system triggered events. The current setup does not allow to explore the scope of identifiable actions in depth. However, as this identification relies on frequencies, prominent ERPs present during fixations can be distinguished. As such, this approach could be used to identify different types of tasks which were performed during the interaction. One such example would be the identification of the presence of Error Potentials. These potentials could then help to detect the type of error which was done [Spüler and Niethammer, 2015]. Indeed, the profile of an ErrP is different between errors the user notices themselves and errors noticed through explicit system feedback. For this study, the SFBA allows to inspect how scene complexity modulates the tasks performed within it. This can help designers to locate specific scenes which demand more activity from the user and adapt scene elements to lower this demand.

Our three methods thus offer new ways for designers to quantifiably assess the interaction on different levels. While being rather superficial, these methods serve as a transposable entry point for the analysis of interfaces. They aim to make interfaces more efficient and satisfactory in each design cycle iterations. From these starting points, more in depth analysis can be made relying on the specificities of each considered new interaction. The three method serve as an applicable and valuable initial framework to expand on.

6.6.2 Expanding and Improving Methods

Our three methods were elaborated using a study on controlled stimuli and scenes. Indeed, the study was designed to allow for the properties, that these methods require, to be easily gathered and inspected. However, many manipulations offer more freedom to the users in the frame of a naturalistic set of stimuli and scenes. While our framework was shown to be transposable on a theoretical level, this property needs to be verified. One other aspect which also requires verification is the discrimination of target and non-target stimuli using P100 amplitude discovered above. To inspect both of these issues as well as expand and improve our set of methods further, a second study was performed proposing a similar visual search task in more naturalistic environments. All of this is presented in the following chapter.

Chapter 7

Expanding the Methods to Natural Interactions

7.1 Objectives for Natural Interaction

In the previous chapter, a set of methods was developed with which data obtained from the bi-modal BCI setup could be reliably analyzed and user grievances identified. We have shown that information about the stimuli scenes and their perceived complexity can be extracted by this set of methods using the properties contained within the data. This information can then be applied to evaluate and potentially improve the interface and the underlying interaction. The set of methods was however established and explored only for very specific artificial and controlled scenes. Furthermore, participants interacted with these scenes through an eyetracking based object selection paradigm.

Both the scenes and the interaction paradigm used in the previous study are not present in common daily interaction. So far, the scenes contained a set of random shapes, with flat colors, semi-randomly inserted in a two dimensional scene. The established methods and their utility thus need to be tested in a more natural interaction context to validate them for common interface evaluation. A natural interaction context means that the interaction occurs within environments more frequently encountered in a user's daily life. Natural scenes (i.e. environment associated with natural interaction) often contain much more irrelevant information and have objects organized in a nuanced way. For examples, natural scenes present depth, shadows, a much less salient color palette, 3D shapes, motion and even distractor objects. In the previous chapter, the proposed interaction was tailored to explore key facets of interface evaluation. Common natural interactions have a more targeted scope and thus do not present the same range of varying complexity as can be explored in artificial interactions. Therefore, we have conducted two additional studies to investigate the effect of changes in scene elements and in interaction task. In this chapter, the proposed BCI setup and the methods developed for the extraction of relevant information will be applied to a natural interaction setting and will focus on the effects of object variety on the user's perception.

Specifically, the SPPQ method using the P100 amplitude will be inspected in this real-life like setting to ensure that its informational value (i.e. monitoring of attentional effort), which was discussed in the chapter 6, remains usable. The real-life like interaction setting also allows for the inspection of properties which were only briefly discussed before. This includes the analysis of amplitude differences in FRP between fixations on target and non-target objects. Another major aspect concerns the exploration of difficulties as they change between users instead of depending on the interaction context itself (i.e. scene objects, whole scenes or task related changes). This includes the identification of users which have difficulties with the interface and what sets them apart from other participants. For this purpose, we will establish an additional method to complete to our set of three.

Gaining insights into these complementary aspects is beneficial when evaluating an interface. Particularly in common real-life interaction, users rely on expertise or preconceptions with the environment when beginning a manipulation. Their biases serve as their starting point for their visual assessment of a scene [Downing, Moore, and Brown, 2005]. The exploration of fluctuations in the data between users objects can thus help to identify these biases for individual users. To explore these biases in parallel to the validation of the transposability of the previous three methods, a natural interaction will be studied which offers variability in objects and user expertise without changing the layout of the entire scenes or nature of the interaction task significantly. Practically, a visual exploration of kitchen scenes has been selected.

7.2 Emulating Natural Interaction in a Laboratory Setting

Natural interactions are defined here as a subset of HMIs which occur in less controlled environments. These types of interactions are commonly encountered in daily life and are often explored in the industrial and ergonomics field of HMI. Due to their high number of degrees of freedom, these types of interactions are very rarely studied with biosignal monitoring in a laboratory setting. Indeed, the laboratory setting generally changes how natural interactions are commonly experienced. Furthermore, natural interactions need to be simplified to allow for an easier monitoring of ground truths for a subsequent analysis. This is necessary as, otherwise, the freedom present in natural interaction introduces noise and confounding factors which are mathematically expensive to remove from the data [Meißner et al., 2017]. This simplification and reduction of interaction is thus necessary to fit within the technical constraints of laboratory studies.

In order to still approach the investigation of natural interaction with biosignals, approximations need to be made. For example, pictures of real-life scenes [Coco, Nuthmann, and Dimigen, 2020; Kaunitz et al., 2014] or pictures of approximation to real-life objects have been used [Finke et al., 2016]. To align the stated objectives and the need for control over the scene, an interaction paradigm has to be chosen where real-life interaction is only marginally simplified. When working with visual exploration of scenes, examining pictures of scenes is similar to examining real-life scenes in most conditions. While it is possible for differences to occur between pictures and real scenes when certain interaction paradigms are used [Pönkänen et al., 2010], the major difference is the 3D nature of real scenes. If movement is restricted and the scene is clearly presented, this difference however plays an insignificant role. For our purposes, using pictures of real-life scenes is therefore sufficient. To ensure that a wide variety of objects and user expertise can still be explored, scenes are selected where the manipulation context remains similar. For this purpose, we explore object selection tasks within the aforementioned home kitchen scenes.

A home kitchen setting proposes a wide array of different tools, utensils and consumables aimed to solve a wide array of tasks ranging from cleaning to cooking. By using the kitchen environment, visual exploration can thus be performed and evaluated as it relates to different contexts and, more relevant to our study, to different objects proposing different functions and rarities. As kitchen interaction are common in every-day life and used by a significant part of the population, these different scenes will allow for difference between users to manifest themselves. First and foremost, however, kitchen scenes allow for a replication of many of the interaction aspects considered in the previous study, but in a real-life like interaction context. This especially includes the search and selection of specific objects of varying shapes and sizes.

7.3 Study Two : Materials and Methods

This study used a home kitchen environment. To respect the constraints detailed above (i.e. limiting muscle artifacts and adequately labeling stimuli for ground truths), the study was performed on a set of previously prepared images picturing kitchen scenes. In order to reproduce aspects of the last study, tasks requiring a wide variety of objects serving as adequate tools were presented to the participants. The data from this study were also employed to validate the FCSync resynchronization method between eyetracking and EEG data [Wobrock et al., 2019] and furthermore to test the aMLRd algorithm, as discussed in Chapter 5 [Wobrock et al., 2016].

7.3.1 Experimental Protocol

Figure 7.1 presents the study protocol the participants had to follow. They were presented with a series of 40 pictures of scenes. At the beginning of the study, the participants were shown a written task composed of a single sentence. The statement was written in a white Arial font on a fully black background. The statements were written in German and spelled out a task commonly performed in the kitchen. The statement was restricted to tasks which would traditionally be performed within a home kitchen environment. "Peel an orange" is an example for such a task. The participant had as much time as they wanted to read and memorize the task statement presented to them. Once memorized, the subject performed a single keyboard input (i.e. a spacebar press) and the scene associated with the statement was display. The scenes always showed a kitchen counter top featuring a collection of objects congruent with the kitchen environment.

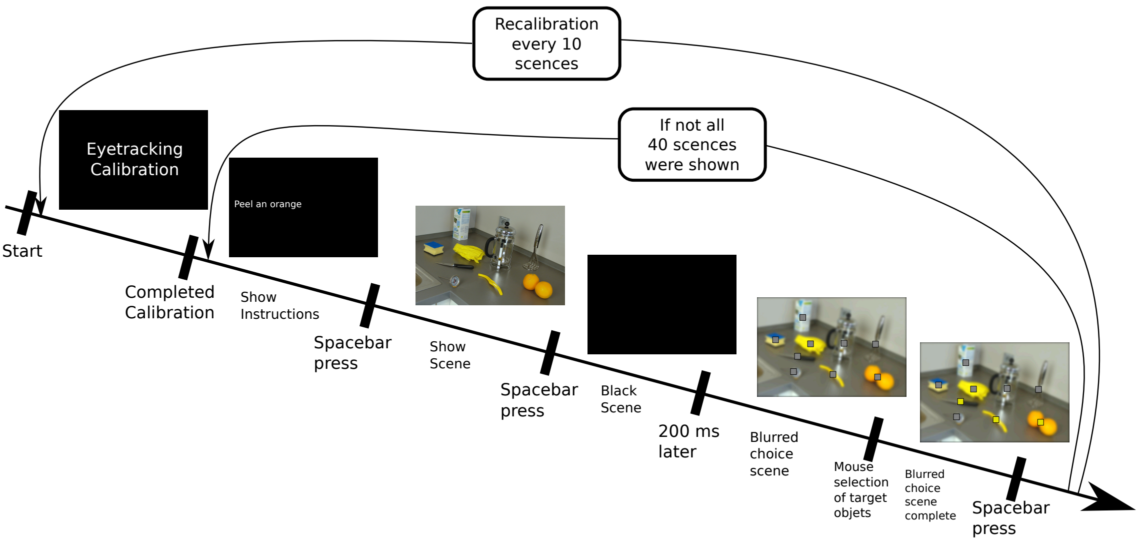


FIGURE 7.1: Task procedure which participants had to complete for each scene in this Study.

Once the scene was unveiled, the participant had to explore it and identify the objects they need to accomplish the task they were given. Once again, participants had as much as they wanted to familiarize themselves with the content of each image. However, participants were also asked to delay their selection as little as possible. EEG and eyetracking data were only recorded during this phase of scene exploration of the study to document this natural exploration part of the interaction. After another keyboard input, the screen was blacked out for 200 ms and a modified version of the same scene was shown. This second time, the scene was heavily blurred using a spatial gaussian blur filter. A gray square was superimposed onto each object present within the scene. The participant, using the computer mouse, could click on these squares to select or deselect the associated object. The participant had to select (by clicking the mouse cursor) the square associated with the objects they would choose to use to accomplish their written task. Once they were satisfied with their selection, participants were asked to perform a final spacebar press to validate the scene. Subsequently, the screen was blacked out for two seconds and the participant was confronted with the next written task. This was repeated until the participant viewed and selected objects for all 40 scenes. The order in which these scenes were presented was randomized, and thus appearing in different sequences for each participant.

Participants were introduced to the general study protocol through an example task and scene which was presented to them prior to the actual study. This tutorial presented a scene and a written task not present in any other scene. The participants had the opportunity to repeat this tutorial scene if they desired. Once familiarized with the protocol, the participant started the actual study. Prior to the experiment, the eyetracker was calibrated. It was recalibrated three times during the experiment, each time after 10 scenes were completed.

7.3.2 Stimuli

The written tasks presented to the participants were chosen to represent generic kitchen tasks, commonly performed on a daily or weekly basis. This included tasks such as "cut vegetables", "clean the dishes" or "cook pasta". Generally, these tasks were chosen to be performed with objects exclusively found within a kitchen. The chosen tasks covered different areas concerning the preparation of food, handling of kitchen utensils and cleaning tasks.

Each one of the 40 pictures featured between seven to nine objects. These objects were spaced out as not to overlap and be clearly distinguishable from one another. All objects were also elements which can be found in any kitchen. They could range from kitchen tools, to foodstuff such as vegetables, fruit, or other processed comestibles. Cleaning products were also present in certain scenes. The kitchen tools ranged from knives and fork to pots as well as specialized tools (e.g. steak thermometer or orange peeler). All scenes can be seen in Appendix A.

Within each scene, there were between two to six objects which could be considered target for the given written task. Other than being target or non-target, objects could also vary in their "commonness" to the participant. Indeed, in certain scenes, objects were present which are not commonly found in most western households. Their number varied from 0 to 3 depending on the scene. These objects included elements such as uncommon fruit or specific tools such as an egg-yolk separator, corncob holders or orange-peeler tools. In some scenes, certain uncommon objects were considered target objects while in others they were considered non targets. Figure 7.2 presents an example of a scene which was shown to the participants containing objects of all types of commonness (numerical labeling added here for clarity). Which categories an object belonged to was previously labeled. Practically, this labeling was performed by hand creating a mask image where the corresponding pixels were changed to non-zero values, corresponding to a numerical index arbitrarily assigned to the object. This way, during analysis of the data the fixations could be projected onto this mask to determine which objects were looked at.

Objects were categorized into six distinct classes, along three dimensions, for each scene. The first dimension indicates if the object was selected by the participant during

FIGURE 7.2: One of the scene images that was presented to participants, with the goal to peel an orange. The scene contains four types of stimuli : 1- Simple target stimuli, 2- Simple non-target stimuli, 3- Complex target stimuli and 4- Complex non-target stimuli.

the study. The second dimension indicates the commonness of the object. These objects will be referred to as common and uncommon in the rest of this chapter. The last dimension recorded if the object was marked as target or non-target during the prior labeling. The first dimension is the only one which changes between users. Ultimately, this third dimension was removed to avoid confusing situations where annotators and users perceived the validity of an object differently. Furthermore, the labeling of "commonness" of the objects is based on the subjective perception of the annotators, meaning that this perception may vary.

7.3.3 Apparatus

The setup of this study was identical to the one for the first study presented in chapter 6. EEG activity was recorded through 16 electrodes positioned at the locations Fz, Cz, Pz, Oz, F3, C3, P3, F4, C4, P4, PO8, PO7, F7, F8, T7 and T8 respective to the 10-20 system [Acharya et al., 2016] as represented in Figure 3.3. Additionally, horizontal and vertical Electrooculography (EOG) was also measured using 4 further electrodes. Two g.tec g.USBAmp biosignal amplifiers were synchronized via g.INTERSync cable and used for the EEG recording (at 256 Hertz). Eye movements were recorded using an LC Technologies Eyefollower Desktop Eyetracker. The Eyetracker was positioned below the interaction screen and calibrated through a 9-point screen calibration at the start of the experiment. The device was also recalibrated using the same procedure after the completion of every ten scenes. This was done using the same 9-point calibration provided by the eyetracking software. Eye gaze samples were recorded with a frequency of 60 Hz for each eye in alternation (120 Hz total). Both devices were operated by the same computer. Participants were seated with their face at a distance of around 80 cm from the screen. As a result, their foveal region during the study encompassed a circular screen region with a radius of about 80 pixels. EEG and Eyetracking data were recorded on the same computer in this setup.

7.3.4 Participants

Eleven volunteers (5 female) participated in the study. The participants were all recruited from the local student, university visitors, and staff population and were either paid for their expenditure of time or granted course credit. All participants had no known prior or current pathological neurological condition (based on self-report). Participants had normal or corrected-to-normal vision. The experimental procedure and written consent form for this study were approved by the ethics committee at Bielefeld University, and adhered to the ethical standards of the sixth revision of the Declaration of Helsinki [Association, 2001]. All participants gave their informed written consent to participate in the study.

7.3.5 Artifact Correction and Preprocessing

Once a participant completed the entire 40 scenes, their data were saved. Every step presented to the participant was also logged with the respective timestamps. This included the order in which the scenes were presented to the subject, when each new phase began and when moments of action occurred (i.e. keyboard inputs). Fixation locations and onset times were extracted from raw eye-tracking data of the subject's dominant eye. The fixation detection algorithm registered a fixation when the subject's eye gaze remained in a 5-degree eye-rotation angle area for at least 100 milliseconds. The EEG data were firstly filtered using a 0.1 Hz highpass and an 8th order Butterworth 45-55 Hz notch filter. This was done to remove low frequency artifacts as well as artifacts originating from device voltages.

Subsequently, artifactual components, resulting from eye, head and other muscle movements, were removed from the signal using an Independent Component Analysis (ICA) performed on the EEG channels [Jung et al., 2000]. Artifactual ICA components were identified using the MARA, an automated artifact detection algorithm developed by Winkler et al., 2015 and eliminated afterwards. Here both the EEG and Eyetracking recording were performed on the same computer ensure the usage of the same timestamps. As such, no additional FCSync resynchronization or timestamp correction was required.

7.4 Study Two : Data Analysis

This study was conducted with the aim of addressing three points of the proposed BCI setup for interface evaluation. In the objective of developping a transposable BCI framework, the first point to test is if the methods we established in the previous chapter (i.e. the GDET, SPPQ and SFBA methods) remain applicable and valuable to examine natural interactions. Natural interactions also introduce additional complexities. The two remaining points addressed in this chapter aim to investigate what additional information can be gained through our BCI setup when difficulties arose during interaction. Specifically, the two issues relate to : firstly, difficulties about the type of objects contained in the scene (i.e. difficulties in identifying task-relevant objects due to uncommonness or lack of familiarity) and, secondly, difficulties relating to differences between users (i.e. certain users performing better than others). To study these three points adequately, the collected data are analyzed using the same protocol as in the previous study (see Chapter 4). Accordingly, data gathered from each modality are first inspected individually and, in a second phase, combined as FRPs.

7.4.1 Eyetracking Results

Eye movements serve as a starting point for the observation of an interaction. In this second study, the natural interaction environment proposes new characteristics which could modify interaction. The first one is the presence of a non-blank background. Localizing the fixations performed within the scene shows that 96.92% ($\pm$ 1) of the fixations (across all participants) were performed on a scene region which contained at least one defined object of interest for the task. Thus, almost all fixations can be considered as task relevant fixations, not exploring non-relevant regions. Similarly to the study of Chapter 6, these fixations lasted on average 250 ms interspersed with saccades covering a distance of around 90 screen pixels. As these characteristics are congruent with other visual search tasks (as reported by Baccino and Manunta (2005)), we can conclude that exploration in our natural visual searches does not noticeably modify eye movements with respect to the exploration of artificial scenes.

From here, a more detailed analysis of the different categories of objects present in the scene was performed. As mentioned before, objects present within the scenes were characterized according to the three different dimensions : target/non-target, common/uncommon and selected/not selected by the participant. Participants selected objects labeled as target in 93% ($\pm$ 2) of cases, presenting a great overlap between the two categories of target labeling and selection. Regarding the 7% discrepancy, the major part was due to the difference in opinion in what defines a target objects, rather than lack of knowledge of the utility of the object in question (i.e. uncommonness of the stimuli). Due to this large overlap and the quite limited number of fixations available (on average 533 fixations over the whole study, $\pm$ 200 for a participant), the target labeling dimension will not be discussed further from here on. It is mostly redundant with the selection dimension and applies to less than 5 fixations for most participants. Inspecting the two remaining dimensions (i.e. selection and commonness), major differences can be noted in fixation duration on these object types. Table 7.1 shows average fixation duration across all participants. These results are furthermore subdivided according to object commonness and selection. We can see that selected objects are fixated about 200 ms longer on average than non-selected objects. This happens regardless of commonness of the objects. Comparing commonness, there is no noticeable difference in average fixation time between common and uncommon objects regardless of selection. The participant thus appears to spend more time fixating on regions containing objects they selected. These regions are named regions of interest (ROIs) in the rest of this work.

TABLE 7.1: Duration of average Fixation in milliseconds, considering the selection and commonness of the fixates object.

	Common Objects	Uncommon Objects	Any Commonness
Selected Objects	576 ms($\pm$ 59)	605 ms($\pm$ 92)	578 ms($\pm$ 60)
Non-selected Objects	372 ms($\pm$ 62)	371 ms($\pm$ 41)	372 ms($\pm$ 41)
Average	568 ms($\pm$ 56)	402 ms($\pm$ 46)	511 ms($\pm$ 51)

Interpreting eye tracking patterns on an entire scene level becomes more difficult due to the lack of ways to quantify the difficulty of a scene. Indeed, the complexity of each scene is more closely linked to the subjective perception of task and object complexity rather than on the quantity or color similarity in the scene composition. Nonetheless, the same approach as in the previous study can be employed to gain insight into the perceived difficulties of the interaction. Therefore, we first consider the number of fixations performed on each scene. The number of fixations on a scene mostly stays around 15 fixations on average but this value can vary with the presented scene. This information is illustrated in Figure 7.4 where each boxplot represents a distribution of the number of fixations performed to complete all 40 scenes. The index of these scenes increases from left to right and can be seen in Appendix B. For explanation of the structure of these boxplots, please refer to the description of Figure 6.15. Indeed, certain scenes (such as scenes 11 or 12) present much broader variations in the number of fixations employed between participants. This indicates, that while scenes and tasks were created to present a uniform level of difficulty to the user, certain tasks and scenes were perceived as more difficult. Differences in the number of fixations become even more apparent when looking at the fixation number profile by participant. Figure 7.3 presents differences in the number of fixations performed onto all scenes for each participant. The differences observed between the participant boxplots (and their standard deviations across scenes) can be quite large (such as between participants 2 and 7, presenting an average 8 and 20 fixations respectively). The differences in the number of fixations indicate that participants explored the interaction scenes differently. Given these differences, it may be possible to find a correlation between eye motions and certain user task performance metrics, such as completion duration. This would allow to quantify actual user performance from biosignals recorded during the interaction.

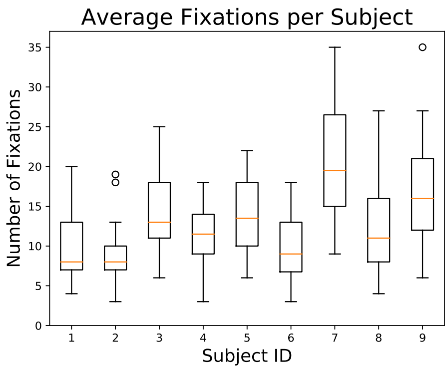


FIGURE 7.3: Number of fixations across all scenes for each of the participants.

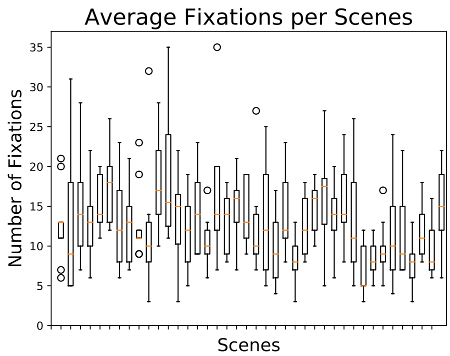


FIGURE 7.4: Number of fixations across participants for each of the 40 scenes.

In summary, eye movements occur similarly in the exploration of natural scenes as they did during artificial scenes exploration. Due to the less constrained nature of the interaction, it becomes however more difficult to accurately categorize the different gaze patterns. Nevertheless, the differences that can be observed through ET show prominent variations between both the scenes and the participants. This indicates that strategies of interaction change depending on the scenes and the participants. We complement these observations with an analysis of the EEG signals of the same scenes.

7.4.2 EEG Results

With respect to the EEG activity, analysis also becomes more challenging in natural interaction. Here the signal cannot be analyzed as it relates to input or feedback events, as none were present during the exploration. Alternatively, the signal of a participant can be analyzed as it relates to an entire scene. As stated previously, the same 40 scenes were presented to all participants. However, there are no specific categories of scenes which can be compared, only objects within them. As no markers for objects are present, they cannot be located through the EEG signal alone. Categorization of scenes can, nevertheless, still be considered by using different criteria to group scenes together. Scenes vary in the number of selected objects as well as with the amount of complex objects within them. As scene duration varied strongly (as shown in Figure 7.4), the EEG signal of these scenes can best be analyzed by comparing the relative frequency bands. The five frequency bands are illustrated in Figure 7.5 where they are contrasted to the quantity of selected objects present in the scene. In each frequency band, variations related to the number of selected objects are not significant (see low average and standard deviation represented in Figure 7.5), as all yield p-values > 0.1 derived from a Student t-test. Variations within each category (represented by the error bars) are also low, which was not the case in our previous study. This lack of variability within frequency bands may be due to the absence of other interaction-related events (and subsequently associated ERPs). This can be attributed to the fact that stimulus selection was only done after exploration (as shown by the protocol in Figure 7.1) and thus no motor action (e.g. spacebar presses or mouse clicks) were performed during exploration.

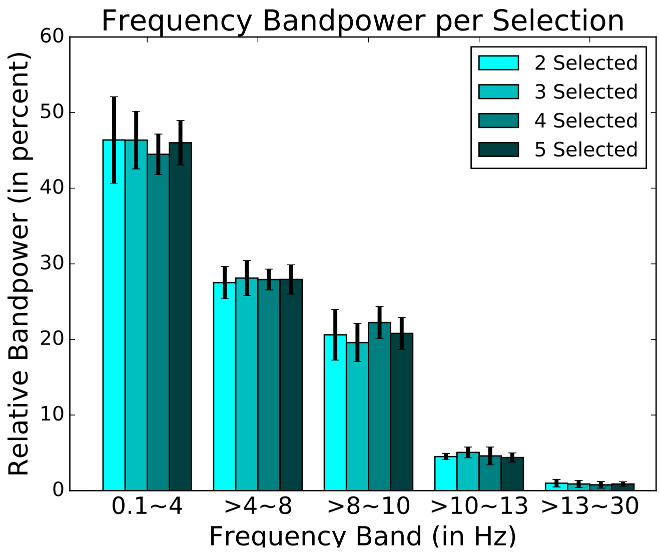


FIGURE 7.5: Average relative frequency bands relating to the different scene types. The vertical bars provide the standard deviation. Scenes were categorized depending on the quantity of objects selected by the participant within them. This Figure was generated using the data from one participant.

Although these results indicate an absence of strong difference in frequency bands between any scene in this study a further analysis is performed to ensure the validity of this statement. The relative frequency data from all scenes and all participants are projected into a two dimensional space using the t-SNE algorithm [Maaten and Hinton, 2008]. Data were labeled according to the explored scenes. While this procedure promotes a data driven clustering of the data, highlighting scene similarity, here, no noticeable clusters appear in the representation of Figure 7.6. By labeling the data according to the various participants (as is shown in Figure 7.7, which is a recoloring of Figure 7.6), the presented data seem to create different clusters, grouping data from the same participant. To generate both of these figures, relative frequency bands from all recorded EEG channels have been used as input features to the t-SNE algorithm. While differences in relative frequencies between participants are expected [Klimesch et al., 1998], these differences can help elucidate the manipulation differences observed through ET. It may be possible that participants presenting relative frequency bands closer to one another had similar performance during study. The overlapping of the different cluster may thus allow to group participants together in terms of mental state [Hanslmayr et al., 2007]. This however needs to be tested when both modalities are combined through our methods.

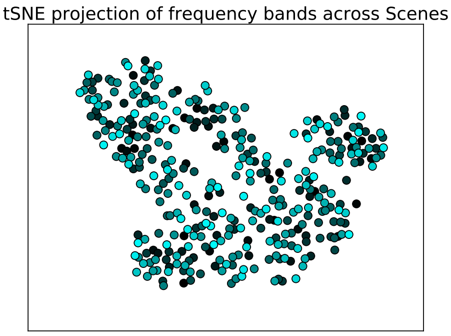


FIGURE 7.6: t-SNE projection of the relative frequency band. Colors represent the scenes.

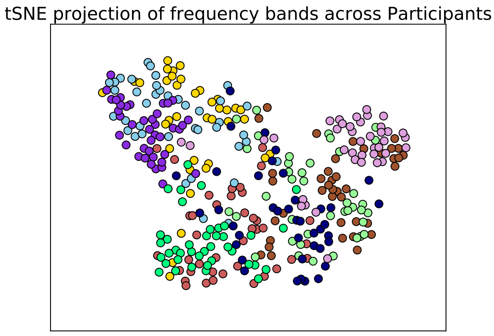


FIGURE 7.7: t-SNE projection of the relative frequency band. Colors represent the subjects.

Differences between participants in terms of relative frequency bands are also quite small. Indeed, when performing a similar frequency representation as in Figure 7.5 between participants, differences are generally non significant when performing a Student t-test. While differences in the power in a same frequency band are present, they do not function as properties for comparing the performances of participants. Such differences in performance are also generally difficult to observe when considering data from a whole scene. Indeed, these recordings are long and present varying attention during exploration. These variations are lost when calculating the relative frequency bands. Consequently, a more focused analysis of the EEG signal through FRPs should be performed.

7.4.3 FRP Results

Similar to the previous experiment, the data are segmented into epochs of one second duration at the onset of eye fixations. Each fixation is categorized by the contents of the foveal region. The EEG data epochs are then analyzed with respects to how they relate to the contents of their corresponding fixation within a scene. Here also, only two major potentials can be observed within the average epochs. These major potentials consist of an occipital P100 and a frontal N100 (identical to those shown in Figure 6.8). This study took inspiration from the work of Finke et al., 2016, where participants were tasked to find an object within a gaze contingent scene and produced a potential similar to the P300 when finding the relevant object. However, despite a similar interaction paradigm, our recorded data did not present any potential similar to the P300 potential. The P300 potential is commonly associated with surprise, relevance and possibly novelty [Goldstein, Spencer, and Donchin, 2002]. As task instructions were however given in advance in a written form (as opposed to the paradigm proposed in Finke et al. (2016)), these notions of surprise and relevance were absent.

The P100 potential shows variations in amplitude across the epochs of the study when observing the grand averages of epoch subsets. Indeed, repeating the approach from the previous experiment (i.e. Chapter 5), FRP amplitudes were found to be correlated with the number of objects present within a fixated area. In contrast to the previous study however, while differences can be observed between categories in certain participants,

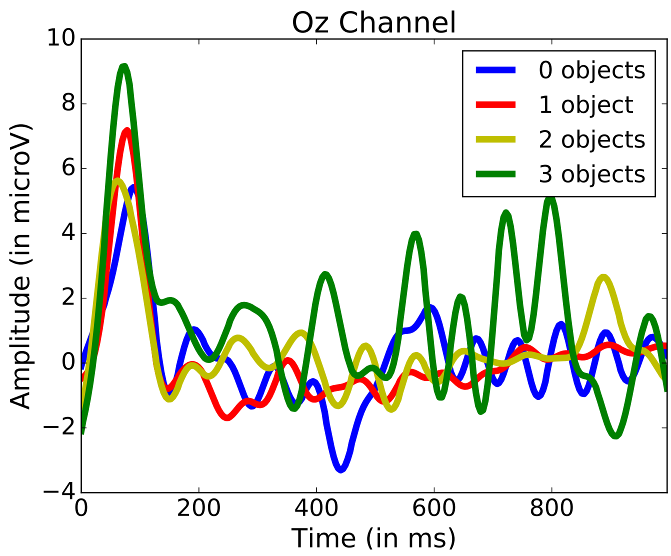


FIGURE 7.8: Average Fixation-related Potential depending on the number of objects present in the foveal region. The graph only represents the averaged data from a single participant.

these differences in the P100 amplitude between number of objects in the fixation region are not as clearly pronounced as in the last study. This is illustrated in Figure 7.8, where the 3-object curve does present the highest amplitude, but the 2-object curve does not present the second highest. In contrast, to the study discussed in Chapter 5, an increase in the amount of objects present in the foveal region of the fixation did not necessarily correlate as well with an increase in Oz channel P100 amplitude in the average epoch. This holds true for the majority of the participants of this study.

In the previous study, difference in P100 amplitudes was also found to correlate with whether the fixated object was target or not. This could not be investigated further in the first study due to the overlap between salient colors and target objects : more fixations on target objects occurred on scenes where many objects were present, conflating these labels. Here, comparisons between selected and non-selected objects as well as common and uncommon objects can be made much more clearly. In Figures 7.9 and 7.10 grand averages of the EEG recordings for each participant are displayed. Comparing the grand averages between these categories, differences in P100 amplitude in the occipital P100 peaks can be identified. When comparing FRPs between fixations on common and uncommon objects, the amplitude difference is around 0.2 microV, in favor of uncommon objects, and is only barely visible in Figure 7.10. However, when comparing FRPs between fixations on selected and non-selected objects (see Figure 7.9), this difference is larger (around 0.9 microV) in the favor of selected objects. Both of these differences in peak amplitude show strong variations between participants. Effectively utilizing this amplitude differences is thus inconclusive. However, these first results indicate that the P100 amplitude may be a better indicator for object selection over object commonness.

Both of these grand average observations can be further investigated by comparing the distribution of the P100 amplitude obtained from the application of the aMLRd algorithm at the beginning of each epoch. These distributions are visualized in Figure 7.11. Three pairs of distributions are illustrated : target/non-target, common/uncommon and $<=1$ object/ >1 object in the foveal region. As expected, no significant differences between the distributions can be observed between each of the three pairs of distributions.

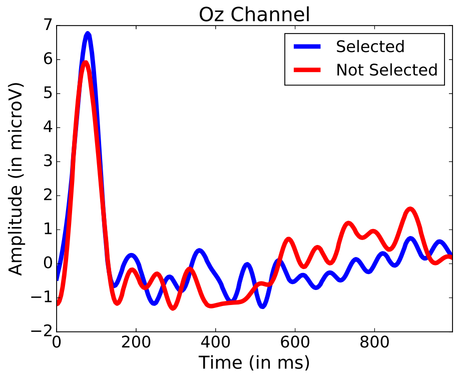


FIGURE 7.9: Average fixation-related epochs containing selected and non-selected objects. Used data from one participant.

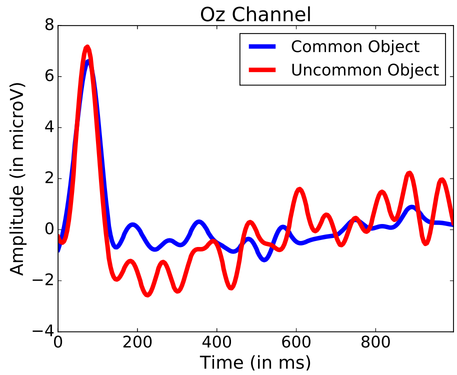


FIGURE 7.10: Average fixation-related epochs containing common and uncommon objects. Used data from one participant.

By performing a set of Student t-tests between these pairs of distributions, all returned p-values are greater than 0.1. This is the case between quantities of objects, between target and non-target objects as well as between common and uncommon objects. However, as observed in the previous ERP plot analyses, variations (illustrated by the bars representing standard deviation) within each category are significant. This indicates that although one cannot discern between the proposed categories using the P100 amplitude, there are still other factors at play which drive the changes in this potential, and thus attentional fixation costs. When observing P100 amplitude variations across participants, more noticeable differences in distribution can be observed. These are displayed in Figure 7.12.

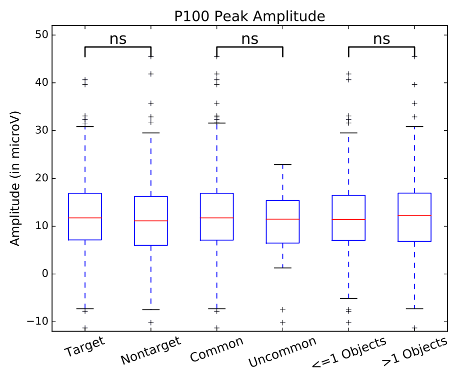


FIGURE 7.11: P100 amplitude distribution for different fixation categories.

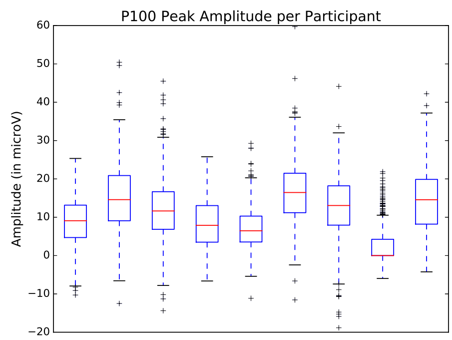


FIGURE 7.12: P100 amplitude distribution across all nine considered participants.

In this second series of boxplots (i.e. Figure 7.12) differences in P100 distribution are much more apparent. The absence of such difference on a task level may be strongly related to the clear visual distinction between objects within a scene. As visual saliency is less accentuated in scenes, when compared to the last study, the differences in attentional effort are minimized. The lack of these scene-related complexity also makes differences between subjects more apparent which may help to better determine which expertise or preconceived notions each user brings to the interaction. For this purpose, the results obtained from each modality inspection can now be compiled and discussed in greater detail using our previously established set of methods.

7.5 Adapting our Methods

As stated above, the goal of this study was threefold. The first objective was to inspect how well the three methods elaborated in the previous chapter (i.e. GDET, SPPQ and SFBA) could be used in a natural interaction environment. The second objective was to investigate if differences between objects are detectable from biosignal recordings. Indeed, differences between scene elements are commonly present in everyday interactions. Our intention was to elaborate a method which would allow to reliably gather information about these common object-related difficulties during natural visual exploration. The final objective was the creation of a method enabling the identification of user-level difficulties. All previous methods did not allow to distinguish between participant performance. Participant differences are however important when considering interface evaluation, as each participant may have his own way of approaching an interaction. From the early analysis reported in the previous sections, observations and conclusions about all of these three objectives can be made.

7.5.1 Transposing the GDET, SPPQ and SFBA Methods

In Chapter 5, three methods (GDET, SPPQ and SFBA) were established. Each of them focused on the identification of difficulty and grievances at different levels of interaction. The first of these methods, the GDET, aimed for the exploration of eye movements in order to identify difficulties related to scene layouts. This was done using three different ET-related metrics : ratio of long-lasting fixations (i.e. longer than 300 ms), ratio of visual region revisits and ratio of fixations on the instruction scene. As there is no instruction scene in the present study, an alternative needs to be proposed. Previously, we mentioned the creation of a mask wherein the image pixels are associated with objects in a scene. The information about the search scene fixations from the previous study (i.e. the third ET-related metric) is thus replaced by information about the ratio of fixations performed on visual regions containing objects pixels versus fixations performed outside of these regions (i.e. on the image background). This type of information is related to the task which is performed by the participants here and requires system adaptation. Similarly to its counterpart metric of Chapter 5, this diverges from our initial goal of ensuring the transposability of our setup, but, due to the static nature of the scenes, this adaptation is feasible prior to interface evaluation. Indeed, preparing these masks prior to interaction and applying them to the BCI setup limits the impact of this adaptation. Similarly to the previous study (see Figures 6.13 and 6.14), the data obtained through these metrics can be plotted for an initial analysis. Figure 7.13 shows the scene-wise t-SNE projection when using these three ET metrics as features. Lighter coloring of points corresponds to longer scene completion times.

As can be seen from Figure 7.13, multiple clusters of data are present, indicating an underlying structure in scenes perception across participant. However, when applying a labeling reflecting the overall duration of each scene (as is done in Figure 7.13), none of the clusters seems distinct from the other. Other categorizations of scenes can be superimposed onto this distribution to examine which aspects of the interaction drive the differences in scene perception. While the exact ground truth labels are missing in this study, this view of clusters of participants and scenes still allows for the designers to evaluate the interaction qualitatively. This could, for example be done by locating a participant who showed an overall good performance during interaction and comparing how other participants are located in this distribution with respect to his scenes. In the previous study, clear differences between established categories (i.e. color similarity and object quantity) could be observed. These are absent here. The scenes themselves cannot

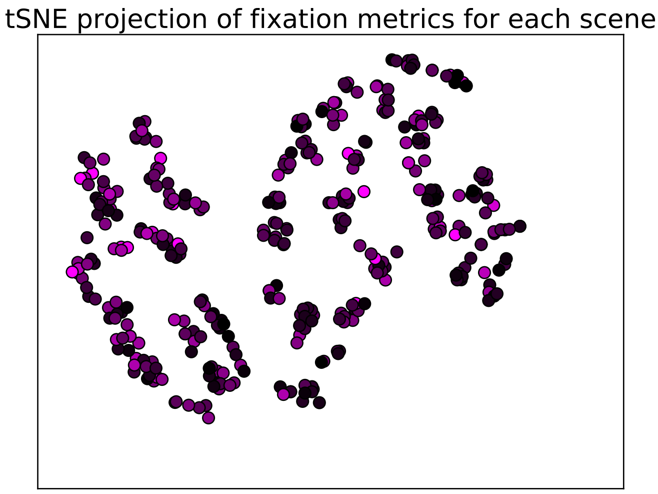


FIGURE 7.13: t-SNE projection created from three different metrics : ratio of long fixations, ratio of revisits of areas, ratio of fixations on scene objects. Each point represents a different scene. Scenes from all participants are present in this Figure. The lighter the color, the longer a participant needed to complete the associated scene.

be labeled or be grouped together, implying that a meaningful categorization of the tasks can unfortunately not be performed.

The second proposed method, the SFBA, aimed for the identification of changes in tasks performed by the user through EEG relative frequency band analysis. For this purpose, fixation-bound epochs of varying length are created and transformed to obtain comparable frequency bands. While their segmentation and their evaluation are feasible in this study, the considered visual exploration phase did not contain any motor action outside of the concluding button press. Investigating this method is thus not adapted to this study as interaction was entirely separate from the visual exploration. Indeed, the differences in frequency band variation were minimal during the experiment for a given participant (see Figure 7.5), indicating the absence of other interaction relevant potentials in the signal. Still, applying this method allows the designers to ensure that all scenes were performed with the same guiding task and absence of motion.

The third method, the SPPQ, considered the P100 amplitude as a marker for the amount of visual information contained within a region of the scene. Here, the same aMLRd algorithm can be applied for obtaining these values. However, as the presented scenes were designed to display objects clearly and distinctly from one another, no changes in the attentional cost are perceived from P100 changes (as can be seen in Figures 7.8 and 7.10). This indicates that, at least from a pre-attentive perspective, all visual regions in all considered scenes are initially perceived in a similar way. Like the SFBA, the SPPQ can still be used to ensure that no particular visual region cause more attentional cost than others.

All three methods can thus be technically applied in natural interaction. However, their application may not provide the same informational value as in the previous study. Indeed, the informative value proposed varies strongly depending on the observed scene and related interaction tasks. As object density and manipulation tasks do not noticeably vary during the interaction, the methods used for their adequate analysis (i.e. GDET, SPPQ and SFBA) do not yield remarkable information. This still offers an interesting insight to designers regarding the workings of the interaction and allows to exclude the presence of certain difficulties. Nonetheless, the already acquired EEG potential properties (e.g. P100 amplitude) can still be employed to explore other aspects to interaction. Particularly, these properties could be used to assess difference in perception between individual objects within the scene rather than clusters of objects.

7.5.2 Comparing Perceptions of Individual Objects

This study was conceived to investigate how the difference in perception between objects in a scene could be evaluated with the proposed bi-modal BCI setup. For this purpose, each scene presented objects which can be categorized in two different ways : target or non-target and common or uncommon. Being able to identify these differences would allow for a better evaluation of the interface by determining which specific objects within the scene causes difficulties for the user.

Using a classical approach to FRP inspection (i.e. the comparison of grand averages), the lack of differences in potentials could be observed between both of these pairs of categories. Indeed, looking at the signals at fixation onset, no major difference can be observed between target and non-target as well as common and uncommon object fixations (as already discussed in Figures 7.9 and 7.10). However, this aspect can be further analyzed by performing a classification between the proposed categories using a PCA to extract meaningful features and an LDA as classification algorithm. The results of this classification are displayed in Table 7.2. For all presented classifications, classes of objects were balanced and segmented into 5 equally sized sets used for crossvalidation. The resulting values represent the area under curve (AUC) of the Receiver Operating Characteristic (ROC) curve, which reflects the performance of the linear classification more accurately than regular accuracy measurements [Bishop, 2006]. Practically, the ROC curve plots the ratio of correct classification versus the ratio of incorrect ones. By shifting the calculated classification hyperplane along its normal vector across the examined data set, multiple points for the ROC curve are created. The more the ROC arches upwards, the easier the classification allows for the correct discrimination of the data.

TABLE 7.2: Classification AUC percentage of the ROC curve using the first second after fixation onset from all 16 EEG channels. Changes depend on the category considered and whether the signal or its FFT decomposition were used as input features.

	Time Signal	FFT Amplitudes
Selected v Non-selected	51.8 %($\pm$ 3.24)	59 %($\pm$ 4.63)
Common v Uncommon	51.7 %($\pm$ 6.02)	54.3 %($\pm$ 3.75)

The resulting AUC are all close to the chance level (i.e. 50 % as 2 classes of objects are present). Using the time signal information as input shows a chance level classification. This indicates that there isn't any presence of major time-wise changes in the recorded signal, like the presence of another potential, when comparing signal profiles between categories. However, when using the FFT amplitude (i.e. the absolute value of the FFT coefficients) of these same signal segments (equal in length) as inputs to classification, an above chance difference between categories can be observed in both cases. While the AUC percentages (see Table 7.2) are too low for effective use in BCI, they indicate minute differences being present in potentials between target and non-target categories. This observation is in agreement with changes of the P100 amplitude, which show small differences when displaying grand averages (Figures 7.9 and 7.10). However, it is also likely that other ERPs are implicated in this differentiation as differences in P100 amplitude distributions are not statistically significant between both considered categories (see Figure 7.11).

Although this classification is not really a reliable method for our setup, this observation nevertheless indicates that differences between amplitudes of potentials are somewhat present. Particularly, the difference in selected and non-selected object fixations show, while true object selection ERPs (like the P300) may be absent from the recordings, small changes in existing potentials occur. As the main potentials which are present appear pre-attentively, a certain priming of the participants may have occurred. A similar effect has already observed been in face recognition studies [Schweinberger, Pfütze, and Sommer, 1995]. The observation of the P100 peak amplitude can still be of value to identify objects that the participant has primed to identify. However, in the future other possible relations to these differences between individual objects also should be analyzed.

The detection of which objects are most relevant for the participants can however still be made using ET data only. Representing the exploration of the scene using a heatmap, the visual areas containing the objects selected by the participant were fixated more often than the rest of the scene. An example of such a heatmap representation can be seen in Figure 7.14. This method is used in classical eyetracking studies. It also allows to gain information about the various regions of interest to the user. While this method is valuable and denotes the complementary power of the modalities used in our BCI, the method is generic and thus not further discussed in this work. While it can still be employed, more interest is put on novel explorations of user grievances.

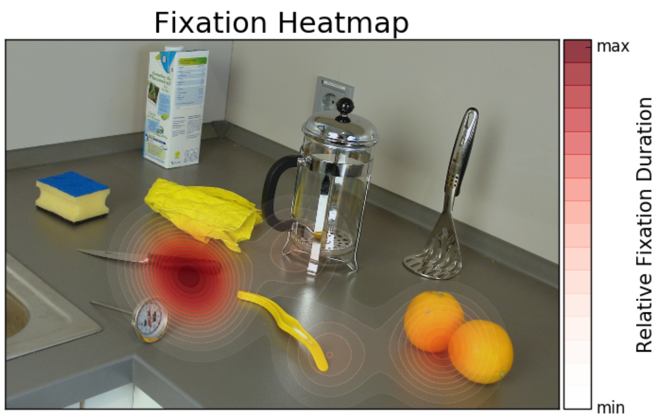


FIGURE 7.14: Heatmap representing the fixations performed on a given scene by a participant. The darker the color, the longer the corresponding region was fixated: here mainly the knife and the oranges.

From other observations made in the data analysis, results showed strong variations between the participants of this study. Attempting to develop a novel method quantifying these results may thus help to gain more information about perception differences on user-level, improving our understanding about the interaction as a whole.

7.5.3 User-dependent Potential Variance (UdPV)

The recordings obtained during this study have shown to present significant differences between participants in both the ET and EEG data. In ET, this concerns the amount of fixations performed during the study for individual scenes (see Figure 7.3). In EEG, it concerns variations in P100 amplitude across the study (see Figure 7.12). Both of these observations indicate that different participants can have different ways of approaching the same task. As the instructions, scenes and objects were however the same between participants, it is interesting to identify which aspect of their approach to the interaction changed their performance so significantly.

To identify the origin of these differences, P100 amplitudes of each scene were plotted with respect to the time necessary to complete the considered scene. All P100 amplitudes of a considered scene, colored according to the participant they belong to, are represented in Figure 7.15 and plotted against the time it took for that participant to complete the scene. The amplitudes of the P100 component between participants are also affected by general voltage offsets and differences in the preparation of the EEG cap. In order to account for the differences, the average amplitude offset, represented by the average P100 amplitude of a participant across all scenes, was subtracted from all P100 amplitudes in order to center the amplitudes. This metric, here termed noramlized P100 amplitude, is used for the Y axis of Figure 7.15.

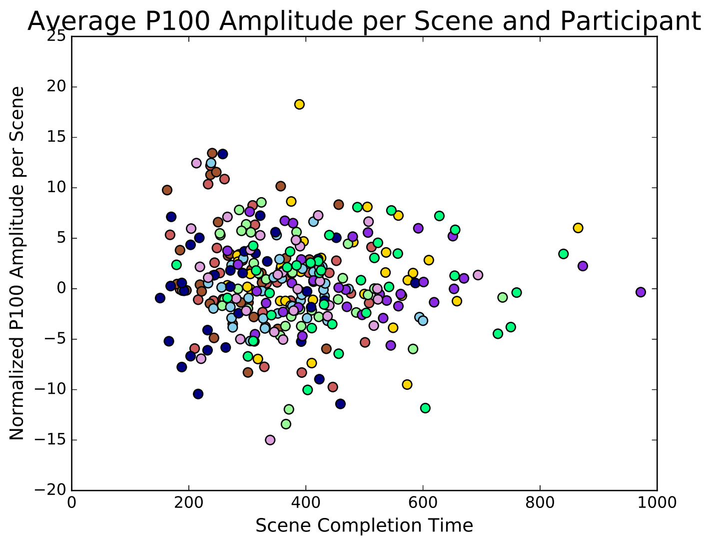


FIGURE 7.15: The average P100 amplitude recorded during a scene, plotted against the total duration of completion of each scene. Each color refers to a different participant.

The resulting plot of Figure 7.15 shows a scattering of data. A noticeable fact is that participants which were able to complete individual scenes faster presented a higher variation in average P100 amplitude over all scenes than participants which needed more time to complete each scene. This observation is investigated further by examining the standard deviation of all the P100 amplitudes and comparing it to the total duration required by a participant to complete the study. This new comparison is displayed in Figure 7.16.

From Figure 7.16, a strong linear correlation can be observed between P100 amplitude standard deviation and the overall duration of the study. By performing a linear regression of the data, a Pearson correlation coefficient of -0.676 can be obtained. When

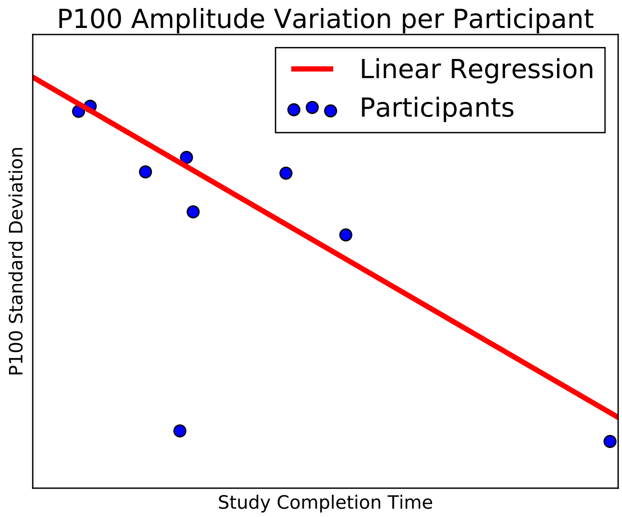


FIGURE 7.16: Standard deviation of the P100 amplitude recorded during the while study, plotted against the total duration of the study. Each point corresponds to a different participant. A linear regression was performed indicated in red.

rejecting the one "outlier" participant, this coefficient increases to -0.93. Thus, a linear correlation between P100 amplitude variation and study duration exists. This improvement of completion time with higher standard deviation of the P100 amplitude could be explained by a better use of attentional effort made by the user. Indeed, participants who prioritized certain objects in the scene over others may be more performant to determine which objects to choose. In opposition, participants who gave each object the same amount of attention took more time to finish a scene. Thus, analyzing the variation of P100 amplitude gives an indication about how the attentional effort of the participant is spent and subsequently how proficient they are in the task.

This result obtained through our bi-modal BCI setup thus allows to gain an insight into expertise and proficiency between participants during a visual exploration task. We term this new method of analyzing P100 standard deviation to completion time, the User-dependent Potential Variance Method (UdPV). This method is detailed in the Algorithm 7. While this method requires more research to confirm and generalize this finding, it can also provide an additional way to gauge interaction for designers and engineers aiming to improve interaction.

7.6 State of the BCI

The presented bi-modal BCI setup has been tested in a natural manipulation scenario focused on the variation of individual objects in the interface. Analyzing the data from this study demonstrated that the previously elaborated set of methods are technically appropriate. The utility of these methods was however reduced due to the nature of the proposed interaction. Additionally, observations relative to complexity of interface objects and differences between participants have been made. Both of these aspects show that the utility of the setup can be expanded beyond the scope initially targeted. The different complexities of interface objects can be explored using common heatmap methods while user-related difficulties can be explored using a new method : the UdPV.

Algorithm 7 User-dependent Potential Variance

Initializing UdPV;
A study has N participants
$P_{var} \leftarrow 0$ ▷ Initialize potential variation array
$M_{perf} \leftarrow 0$ ▷ Initialize performance metric
for $i \leftarrow 1$ to N **do**
    Participant performed M fixations
    $X \leftarrow [x_1, x_2, ..., x_M], \forall j, x_j \in \mathbb{R}^K$ ▷ Epochs
    t_{total}, total study duration for the participant
    M_{perf}.insert(t_{total})
    aMLRd.selectPeak(X, $P100_{bin}$)
    aMLRd.createTemplateAndRegressors(X)
    $P_{amp} \leftarrow []$ ▷ Initialize P100 amplitude array
    for $j \leftarrow 1$ to M **do**
        $x_{aMLRd} \leftarrow$ aMRLd.getPeak(x_i)
        if Maximum present around 100 ms **then**
            $amp_x \leftarrow$max(x_{aMLRd}) ▷ P100 Amplitude
            P_{amp}.insert(amp_x)
        end if
    end for
    P_{var}.insert(var(P_{amp})
end for
Create Regression;
$a, b \leftarrow$linear.regression(P_{var}, M_{perf})
Regression coefficients indicate the relation between these metrics

7.6.1 Preliminary Conclusions Regarding Our Set of Methods

The proposed BCI setup for interface evaluation is structured around the identification of difficulties occurring during a given manipulation. The initial three methods (GDET, SPPQ and SFBA), elaborated in the chapter 6, were applied to help with the identification of scene-related difficulties, task-related difficulties and scene element-related difficulties a user might experience respectively. The methods have been shown to be still applicable in this new scenario, requiring little to no system-side adaptation. Their utility was however reduced. The nature of the interaction limited variation which was center-stage in the previous study, kept the stimuli and task simple. This lead to a decrease in the likelihood of difficulties in these regards. While the information these methods yield may thus vary according to the interaction paradigm, they can still provide reference points for designers to help in their improvement of interaction. Be it only to verify that all the difficulties of the interaction remain constant during manipulation.

The bi-modal BCI setup has also shown that additional information can be collected from the considered HMI. Notably, the setup for eyetracking, allows to identify which interface elements the user considers as important via analysis of heatmaps. Furthermore, while still requiring more research, small changes in FRP amplitude have been observed between target and non-target objects. This could indicate a priming of the users toward target objects and focusing their attention more intensively toward the desired object. The identification of these property changes could be particularly useful for designers when considering improvements of efficiency in interaction by gauging how well the user has been primed to the visual exploration of the interface. Overall, this also suggests that other potentials could be present in the epochs which contain additional information about the interaction. All these statements about FRPs do however require further investigation to be validated. For now, no method about their usage can be formulated.

Furthermore, differences in performance between users can be observed with our setup. Indeed, the BCI setup contrast user performance with the variation in their attentional effort. This enables a possible differentiation between users by identifying those users who may feel more lost in the interaction. Through the observation of the P100 amplitude variation, an additional method called UdPV has been formulated. It allows to observe that users, who require more time to complete a scene, are less selective with their attention. This enables the quantification of user difficulty and thus also enables a new way to gauge the interfaces. This additional dimension to interface analysis may offer new insights into efficiency and satisfaction on the side of the user by informing how easy of a time they had allocating cognitive resources.

In summary, the transposability of this setup is maintained, with all methods of chapter 6 being applicable in this experiment. All the proposed methods do however not hold the same potential for a deep analysis into each facet of interaction. An additional method (the UdPV) has been added to our set. By providing information about user-related difficulties, it can give clues to improve the interface in a next design cycle more effectively.

7.6.2 Expanding and Improving Methods Further

This first approach to real life interaction has shown that the developed setup is promising. However, it also covers only a small part of possible visual search interactions in real-life situations and an even smaller part of all interactions. To allow for more generalization, an additional study of natural interaction is proposed. This next interaction focuses on the changes in the manipulation task. In this new interaction, a comparable approach to the one performed in this chapter will be considered. The applicability and utility of all methods considered in this and the previous study will be tested in this new interaction environment.

Additionally, more attention will be put on the differences between participants (i.e. by applying the UdPV method). As the finding of P100 modulation with respect to priming and expertise of subjects is a novel approach, while not being exploitable yet, it needs to be investigated in further detail to ensure its adequate integration into the proposed interface evaluation set of methods. This set of methods reveal itself as promising as the aspect of differences between participants is normally omitted in the iterations of the design cycle [Faulkner, 2003].

Chapter 8

Exploring the Methods in Changing Interactions

8.1 Objectives in Changing Interaction Tasks

The BCI setup and the evaluation methods established in the chapters 6 and 7 have so far been tested in either one or two interaction scenarios. In the first, more controlled scenario, three different methods (i.e. GDET, SPPQ and SFBA) have been established and their utility has been verified. In the second study, the transferability and utility of this set of methods were verified in a natural interaction setting. Additionally a new method (i.e. the UdPV) has been added to the set of methods. Natural interaction covers an even wider variety of scenarios which occur in very different forms in daily life. In this chapter, we conduct a third study to explore other kinds of natural interaction. Specifically, this is done to investigate the validity of transferring the previously proposed set of four methods further and potentially identifying other useful methods for interface analysis.

As noted in the study of Chapter 6, when analyzing the presence of motor actions, the frequency spectrum of the observed EEG signal changes in relation to the performed task. These changes in task also strongly influence how scenes are visually explored (such as the sandwich making studied by Hayhoe and Ballard (2005)). In common natural interactions, the user's tasks often quickly change over the course of a single manipulation. For example, this can happen when browsing an online shopping interface : first, there is a precise task to select a specific type of product (e.g. a specific piece of clothing such as a shirt) and secondly a free exploration of the collection of proposed products (e.g. looking through the list of available shirts). In the previous studies, we focused on the first part of this example as the user has a precise goal. In the second phase of the example, the user explores the interface more freely, not having a precisely-defined goal. It is beneficial for our proposed BCI to be applicable in situations where such different tasks arise and to provide information about which type of tasks is currently pursued. Task identification would allow to gain a better understanding of the users intentions during interaction. If this identification is reliable, it may even allow in the future to adapt the interface actively during manipulation to match the user's current needs.

With the objective of identifying pursued tasks, a further study has been performed confronting participants with a set of cluttered real-life scenes. The study was performed by Shirley Mey and Prof. Dirk Koester [Wobrock et al., 2017]. This study also used a bi-modal BCI setup comparable to the previous two chapters, recording both EEG and ET activity. In the two previous chapters, only changes between the visual composition of scenes have been explored. In this study, we now focus on presenting similar scenes which are explored according to different tasks. The changes induced by these varying tasks will be investigated to potentially widen the scope of applicability of the proposed BCI.

For different tasks to emerge from an interaction situation, the scene which is explored by a user needs to present a certain minimum of relevant elements. Indeed, if the system interface/scene is very sparsely populated, a user's exploration of the scene will be completed too quickly, meaning very little information about the user's task can gained from it. Similarly, few fixations will be produced. Fixations are however an essential source of information for our setup. This problem notably occurred in the previous study where the amount of fixations per scene was rather limited (i.e. 15 on average per scene), while the study in Chapter 6 had more objects and more fixations (i.e. 39 on average per scene). As we conceive our system as a useful addition to interface analysis, we prioritize investigating interactions where the explored scene is more densely populated with information. Practically, information-dense scenes are commonly encountered when exploring natural interactions featuring a great amount of different object. These scenes are referred to as cluttered scenes. Due to the abundance of such scenes in daily life, they offer an interesting starting point to investigate the detection of changes in task. These scenes also allow a closer replication of the paradigm of the first study (see Chapter 6). By presenting a large quantity of objects, these scenes can be composed in such a way as to allow different tasks to emerge. Before creating the scenes, we take a look at what is known about cluttered natural scenes.

8.2 Clutter in Natural Interactions

Exploring cluttered scenes with the help of eyetracking is a paradigm that has been studied extensively in the past [Gao and Vasconcelos, 2005]. Indeed, cluttered scenes are often used in visual exploration in general [Mann et al., 2007]. When exploring clutter, one of the general findings is that participants usually need a much longer time to complete the proposed task [Moacdieh and Sarter, 2012]. While this finding varies based on the expertise of the participant [Ferrari, Didierjean, and Marmeche, 2008], the salient visual decomposition of such scenes also plays a role in their exploration [Itti, Gold, and Koch, 2001]. As cluttered scenes present an overwhelming amount of information to a user, usually the most visually salient regions of the scene grab the attention of the user first, unless the user has prior expertise relating the task (e.g. they know where to look to find a specific object).

Additionally, other factors also influence the exploration of the visual scene. Such factors can be the presence of abstract stimuli [Albert et al., 2005] (e.g. random shapes like in Chapter 6 rather than the objects presented in Chapter 7) or changes in anxiety and emotional perception [Lester, 1968]. The presence of muscle and general motor activity occurring during visual exploration also influences the way in which an interaction is performed [Mey, Koester, and Schack, 2018b]. One major effect on exploration is in general the presence of task differences between explorations [Earley, Wojnaroski, and Prest, 1987] (i.e. what is the user supposed to find in the scene and how). To avoid introducing all these different factors, the effects of which would likely be enhanced by the clutter, an exploration of familiar cluttered scenes was chosen. Such scenes are commonly encountered by participants in their daily life and feature a collection of objects which are also common and congruent with the environment. To limit expertise changes, the task was restrained to directed search and counting tasks of objects within these common scenes. By adequately tailoring the experimental scenes, variations outside of the targeted task changes are reduced. Other dimensions of difficulty such as scene-level (i.e. some scenes being more difficult to explore), object-level (i.e. some objects being more difficult to identify) and user expertise-level difficulties, are minimized this way but can still be residually present. The exploration of these three levels of difficulty will however be performed when analyzing the results of the study through our set of methods.

8.3 Study Three : Materials and Methods

As detailed above, visual searches, especially those performed within cluttered scenes, are difficult to construct due to the interplay of many different dimensions of complexity. Here, we aim to prepare scenes which allow to explore different scenes with different objectives while limiting possible changes due to such factors. While factors such as stress, abstracted stimuli or motor activity can be adequately controlled for, other biases introduced through different user expertise (depending on scene or objects) are more difficult to account for. These biases can not completely be removed, only minimized in relation to task difference, as our system also takes interest in identifying differences in user perception. The removal and minimization of all these factors is done by creating scenes which present common western household locations. Within these locations, common household objects are scattered in an uncommon but realistic fashion. This creates both a familiarity and randomness which ensure that objects and scenes are known while user expertise is minimized. As the number of objects is high, different types of tasks can be created to explore these scenes. Here, we task participants with two classical scenarios : a taskless exploration of the scene and the searching and counting of a specific type of object present in the scene. These two tasks reflect the precise and non-precise goals mentioned above.

8.3.1 Experimental Protocol

The protocol of this study is presented in Figure 8.1. Participants were presented with a series of 16 scenes. Once an initial eyetracking calibration was performed, each scene was preceded by a black screen on which an instruction was written in white text. These instructions could take two different forms. The first type of instruction asked the participant to explore the scene without providing them with any goal ("free viewing task" represented by scenario 2 in Figure 8.1). The second type of task asked participants to count how many instances of a given types of object were present within the scene ("counting task" represented by scenario 1 in Figure 8.1).The initial task of a participant was randomly chosen at the start of the study. Once chosen however, the following 7 scenes proposed the same type of task and the following 8 scenes proposed the other task. All of the 16 scenes were designed to be explorable with either task. The scenes were thus newly shuffled and randomly attributed to a task for each participant. Additionally, participants were presented with a single button as their input device for the interface. The participant had as much time as she desired to read and understand the instruction given. Once ready, the participant pressed the single button placed in front of her to unveil the scene. Once the scene was unveiled (see third step in Figure 8.1), the participant had as much time as she wanted to explore this new environment. The scenes presented indoor household scenes featuring numerous objects (e.g. pots, screwdrivers, coffee mugs, brooms, etc...). Their composition is detailed in the next section. Prior to the start of the experiment, the study supervisor instructed the participants to be as accurate as possible with their response to a counting task (i.e. try to find all objects of that type in the scene). If ambiguity arose during counting, the supervisor also instructed that it was up to the discretion of the participant to choose whether to count the object or not. Once satisfied with their exploration of the scene, participants could press the single button to close the scene. They were subsequently confronted with a black screen (see last image in

Figure 8.1). During this phase, the instructor came into the room were the participant was seated and asked him a question. In the case of a counting task, this question always was "How many of the objects did you count ?". In the case of a free viewing task, the question varied but was still associated with the scene (i.e. "What room did this scene show ?", "Were towels present in this scene ?", "What would you expect a person to do in this scene ?"). These free viewing questions served to ensure that the participant explored the scene and that the experimental protocol was identical to the counting task condition. The participant answered these questions orally. Once the supervisor had written down the participant's response in their response sheet, the participant was instructed to press the button once again. This unveiled the next instruction and upon a second button press, the next scene. This was repeated 16 times with an eyetracker recalibration occurring after the 8th scene.

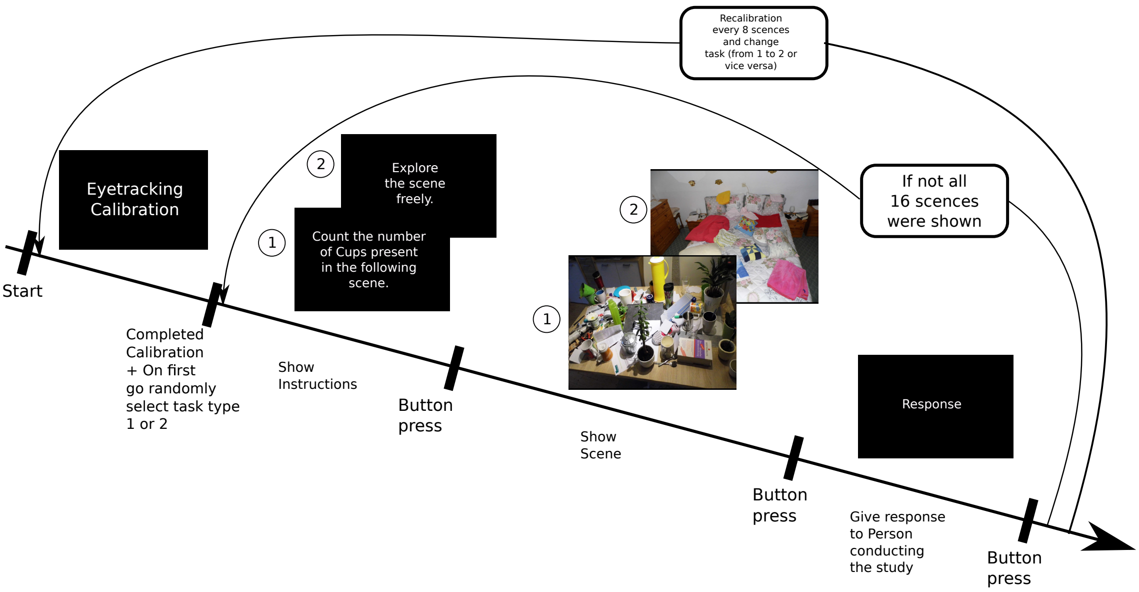


FIGURE 8.1: Task procedure which participants had to complete for each scene. The instruction here shows a counting task (labeled 1) and a free viewing (labeled 2). During the "response" screen (which was completely black), the supervisor of the study asked the participants how many objects of the searched type (here cups) were present in the scene in the counting task. Under free viewing conditions, the question was "Who do you think sleeps in this bed ?"

8.3.2 Scene Stimuli

The 16 scenes had a size of 1280x1024 pixels and were displayed on a monitor featuring the same resolution. Figure 8.2 presents one of the 16 scenes. These pictures showed regular indoor household scenes, featuring a variety of commonly encountered objects. These objects were divided into three categories : counting-relevant 1, counting-relevant 2 and other objects. Other objects refer to a set of objects of various types which were present and congruent with the scene (i.e. paper on the office table), but never inquired about in the counting task. There were also two types of counting relevant object types present in great quantity in each scene (i.e. between 5 and 25 instances of each type per scene). In Figure 8.2, these counting relevant categories are coffee mugs and the flower pots. During the scene order randomization process discussed above, it was also decided which of these two categories of objects, the participant was tasked to count for each counting task. All scenes and their associated counting tasks can be seen in Appendix B.

The consideration of two types of counting objects was done for two reasons. Firstly, our aim was to create cluttered scenes. By composing the scenes with two types of objects in great quantity, this clutter could be better achieved. Secondly, this study was done in the context of student projects, were students would both conduct and participate in the study. To minimize students being primed to identify certain objects, which would induce unwanted expertise bias, a second category of object was offered for counting.

Within the scenes themselves, the clutter was ensured by positioning most of the objects (from all categories) in the center of the image. The objects could overlap but were kept sufficiently apart to be distinct. Statistically, the average searched object of either category fits inside a square of 65 ($\pm$ 50) pixels squared and had an average distance of 105 pixels to the closest object of the same type. The scenes feature various environments around the house including kitchen, bathroom, living room and garage/attic scenes. The visual saliency of these scenes could vary significantly between scenes, with some scenes presenting bright colors while others featured paler ones.

FIGURE 8.2: One of the scene images that was presented to participants. The participants were either be asked to search for flower pots or coffee cups in this scene. During the entire experiment a participant was only asked to look for one of these two types of objects, even when they were shown the scene again.

8.3.3 Apparatus

EEG activity was recorded through 64 electrodes positioned at the locations FP1, FPz, FP2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, M1, T7, C3, Cz, C4, T8, M2, CP5, CP1, CP2, P7, P3, Pz, P4, P8, POz, O1, Oz, O2, AF7, AF3, AF4, AF8, F5, F1, F2, F6, FC3, FCz, FC4, C5, C1, C2, CP3, P5, P1, P2, P6, PO5, PO3, PO4, PO6, FT7, FT8, TP7, TP8, PO7 and PO8 respective to the 10-20 system [Acharya et al., 2016] as represented in Figure 3.3. Additionally, horizontal and vertical Electrooculography (EOG) was also measured using 4 further electrodes. One 64 channel Ag/AgCl electrode Waveguard EEG cap was used to obtain the EEG recordings (sampled at 512 Hertz). Eye movements were recorded using an LC

Technologies Eyefollower Desktop Eyetracker. The Eyetracker was positioned below the interaction screen and calibrated through a 9-point screen calibration at the start of the experiment. Eye gaze samples were recorded with a frequency of 60 Hz for each eye in alternation (resulting in a sampling rate of 120 Hz total). Both devices were operated by synchronized computers. Participants were seated with their face at a distance of around 80 cm from the screen. As a result, their foveal region during the study encompassed a circular screen region with a radius of about 80 pixels. A full eyetracking calibration using LC Technologies software was performed after every 8th scene was explored.

8.3.4 Participants

23 volunteers (9 female, aged 23.3 $\pm$ 2.5) participated in the study. The participants were all recruited from the local student population and were compensated via granted course credit. All participants had no known prior or current pathological neurological condition (based on self-report) and normal or corrected-to-normal vision. The experimental procedure and written consent form for this study were approved by the ethics committee at Bielefeld University, and adhered to the ethical standards of the sixth revision of the Declaration of Helsinki [Association, 2001]. All participants gave their informed consent to participate in the study.

8.3.5 Artifact Correction and Preprocessing

Once a participant completed the entire 16 scenes, their data were saved. Every step presented to the participant was also logged with the respective timestamps. This included the order in which the scenes were presented to the subject, when each new phase began and moments of action (i.e. button inputs). Fixation locations and onset times were extracted from raw eye-tracking data of the subject's dominant eye and LC Technology's fixation detection software. The fixation detection algorithm registered a fixation when the subject's eye gaze remained in a 5-degree eye-rotation angle area for at least 100 milliseconds. The EEG data were firstly filtered using a 0.1 Hz highpass and an 8th order Butterworth 45-55 Hz notch filter. This was done to remove low frequency artifacts as well as artifacts originating from connection between the EEG amplifier and the mains voltage.

Subsequently, artifactual components, resulting from eye, head and other muscle movements, were removed from the signal using an Independent Component Analysis (ICA) performed on the EEG channels [Jung et al., 2000]. Artifactual ICA components were identified using the Multiple Artifact Rejection Algorithm [Winkler et al., 2015]. Eye artifact identification was also applied using the EyeLab toolbox [Dimigen et al., 2011] of EEGLAB Matlab toolbox [Delorme and Makeig, 2004]. Here both the EEG and Eyetracking recording were performed on different computers which were connected together to ensure the usage of the same timestamps.Thus, no additional FCSync resynchronization or timestamp correction was required.

8.4 Study Three : Data Analysis

As we did in the previous two chapters, the data collected from this study are first explored by analyzing each modality (EEG / Eyetracking) separately. Once individually considered, the data from both modalities are combined together for the analysis of FRPs. The combination of both of these approaches permits a more thorough analysis of the interaction. The data are analyzed with respect to changes in scenes, participants (i.e. user expertise) and image contents. In this study, particular interest is put on the differences in recordings between the tasks, the participants had to perform. In the sections below, the free-viewing task will also be referred to as the "exploration task". The dimension of task difficulty will be mainly explored when analyzing the collected data. In addition to this examination, our previously set of established methods (i.e. GDET, SPPQ, SFBA and UdPV) will be exploited again.

8.4.1 Eyetracking Results

We start by examining the data obtained through ET. A set of common metrics about this study is given in Table 8.1. As mentioned above, we made the choice of introducing clutter into the scene because it increases the number of fixations performed on a scene. This is confirmed here as the average amount of fixations for the completion of a scene is 87 fixations. These can be subdivided into 80 fixations for counting tasks and 93 fixations for free viewing tasks. In either case, the results exceed the number of fixations presented in both Chapter 6 and Chapter 7. Thus, the addition of clutter to the scenes increases the number of fixations significantly, allowing our BCI to access more information through a higher number of fixations.

Additionally, another difference in ET metrics is evident between this study and the two previous experiments. The average fixation duration is longer in this experiment, with around 350 ms in contrast to the more common 250-300 ms of the two previous studies. This difference is present in all proposed tasks and for all participants, which already indicates that the fixation duration is dependent on the cluttered nature of the presented scenes. Indeed, objects were more densely placed together on a smaller screen, requiring more time to be processed.

TABLE 8.1: Average eyetracking metrics obtained from the observation of all participants over the two different tasks.

	Counting Task	Free Viewing Task	Either Task
Fixation Duration	350 ms($\pm$ 95)	334 ms($\pm$ 87)	341 ms($\pm$ 91)
Number of Fixations per Scene	80 ($\pm$ 45)	93 ($\pm$ 66)	87 ($\pm$ 57)
Scene Completion Time	27.3 s($\pm$ 15)	30 s($\pm$ 19)	28 s($\pm$ 17)
Time fixating Target Objects	31 %($\pm$ 17.2)	27 %($\pm$ 15)	29 %($\pm$ 16)

With respect to other metrics presented in Table 8.1, we note a faster scene completion time of counting tasks compared to free viewing tasks. Counting tasks were more demanding for the participant as they required higher workload, asking participants to actively find and count objects rather than just freely explore the scene. This difference in completion time is also reflected in the variation of the number of fixations needed to complete the scene. These differences could be due to the fact that objective-less exploration of an environment leaves the user with no concrete endpoint to their search. Consequently, the user takes more time to explore the environment then they would in a guided task with a specific objective [Beck, Lohrenz, and Trafton, 2010].

Finally, when looking at the time spent fixating objects, participants performing a counting tasks fixated target objects longer than during free viewing tasks. The values for time on target objects listed (in Table 8.1) for the free viewing task are incidental as the participant was not told to consider this type of object as target. In the case of free viewing tasks, both categories of counting objects were considered target here. Comparing this metric between conditions provides a baseline for reflecting the natural visual saliency of the object type. Consequently, these differences are explored further in the boxplots of Figure 8.3. For a detailed explanation of the structure of these boxplots, see the description of Figure 6.15. Inspecting this fixation duration difference statistically, participants performed significantly longer fixations onto target objects in counting tasks (where p-value < 0.01 in a Student t-test), in comparison to free viewing tasks. This difference is even more prominent when looking at fixations between counting and free viewing scenes. As our BCI aims to rely on system information as little as possible, these differences may help, similar to the previous study, to locate ROIs relevant to the participant and to identify regions containing target objects. As the time spent on these ROIs is however below 50 % of total scene exploration time, a precise identification of target locations through ET may not be possible.

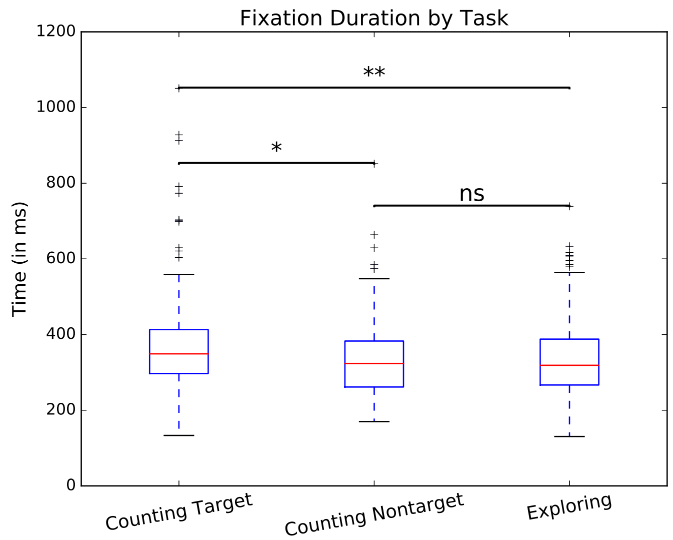


FIGURE 8.3: Fixation duration in milliseconds as a function of the type of task asked of the participant. Within the counting task a distinction between fixation onto target object and non-target objects was made. 'ns' refers to no significant difference, '*' refers to a significant difference with p-value < 0.01. '**' refers to a very significant difference in distribution with p-value < 0.001. Generated by using the data from all participants.

The ET metrics can also be used for a comparison of scene completion time between all the considered scenes and all participants. These comparisons are useful as they allow to detect differences between scenes and participants (and thus potential scene-level and user-level difficulties). All these durations are averaged and presented in Figures 8.4 and 8.5. While some rare scenes take more time to complete than other, overall there are no significant differences in completion time between most of the scenes, with times clustering around 30 seconds (scene number 3 in Figure 8.4 being a rare exception). On the other hand, performing the sample comparison according to the different participants, distinct differences are observed more commonly (see Figure 8.5). This indicates that different participants may have employed different strategies when exploring the scenes. This statement is furthermore strengthened by the high standard deviations listed in Table 8.1.

As such, the analysis of the ET data returns differences between tasks and between participants which need to be investigated in more depth using the results obtained from the EEG and FRP analyses later on when testing our set of algorithms. For now, to understand how these differences present themselves in other modality recordings, the resulting EEG data will be analyzed below.

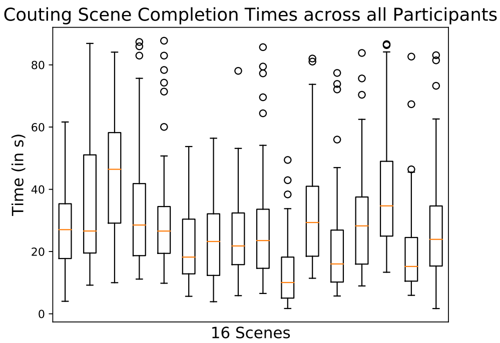


FIGURE 8.4: Exploration Time per 16 Explored Scene.

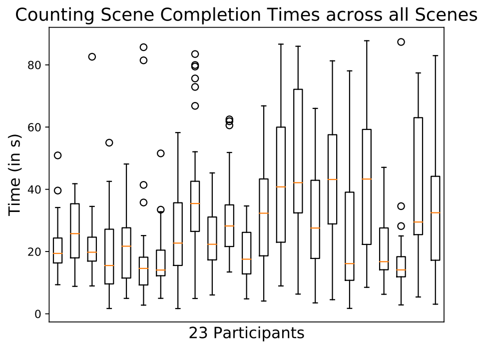


FIGURE 8.5: Exploration Time per 23 participants.

8.4.2 EEG Results

The established set of methods, specifically the SFBA, shows that tasks can be investigated via their relative frequency band powers. Relative frequency bands can help to understand the level of cognitive effort put into performing a given task. Figure 8.6 shows the average relative frequency band decomposition of the EEG signal (at the Pz electrode, see Figure 3.3) over entire scenes depending on the tasks. The analysis of these frequency bands only shows minimal differences between the considered task categories. This observation holds true for all other EEG channels.

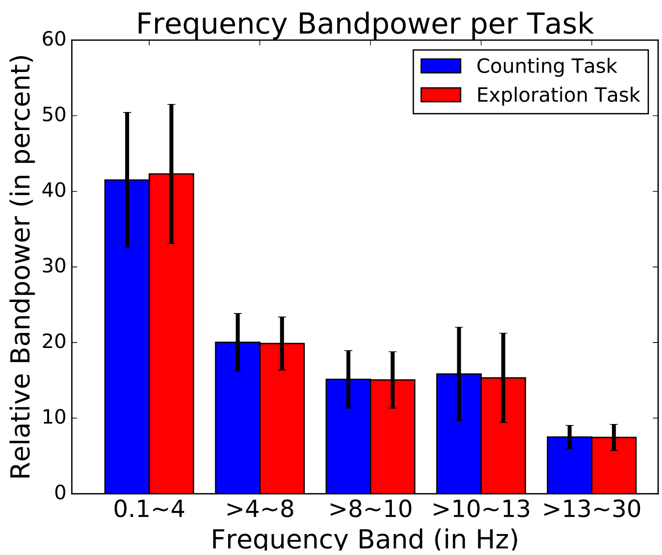


FIGURE 8.6: Relative Frequency Bandpower recorded at the Pz electrode during scenes with respect to scene task. Data used in this figure represent an average over all participants.

While the differences between counting and exploration task in relative power bands may seem negligible when averaging over all subjects, a strong variation remains as shown by the errors bars represented by the standard deviation in Figure 8.6. In particular, the delta (0.1 to 4 Hz) and upper alpha (10 to 13 Hz) bands present a high standard deviation, which lead to different frequency profiles for different subjects. Indeed, when observing individual subjects, these frequency bands show a more noticeable difference between both tasks. This indicates that, while on average, the same attentive state is maintained between tasks, this is not necessarily true for all considered subjects. For each individual participant, differences in the frequency spectrum take different forms depending on the considered task. Interestingly, for certain participants, the counting task presents higher relative frequency power in the upper alpha bands while others show higher values during the free viewing task. Identifying the nature of these differences may help to distinguish different groups of users according to their strategy of approaching an interaction, or more generally, the difficulties a participant has with the visual exploration of cluttered scenes.

In this study, no other manipulation action was logged during the interaction, as there was none during the exploration (beyond the button press ending the exploration). Thus, no segmentation of the EEG data can be investigated using system information and subsequently no non-fixation ERPs can be observed. To gain further insight into the pursued task, the data from both EEG and ET are now combined to detect FRPs.

8.4.3 FRP Results

The signal is segmented into epochs of a duration of 1 second at the onset of a fixation. Similarly to the two previous studies, a strong positive P100 potential is always present at the start of each epoch. This potential occurs in the occipital channels. A negative N100 potential can also be observed in the frontal channel. These potentials are the same as the ones presented in Figure 6.8. The N100 potential will not be further considered in this study. Taking inspiration from Chapter 6, the variation of the P100 potential amplitude is firstly analyzed (via aMLRd) as it relates to the quantity of objects present within the fixation region. These variations are illustrated in Figure 8.7.

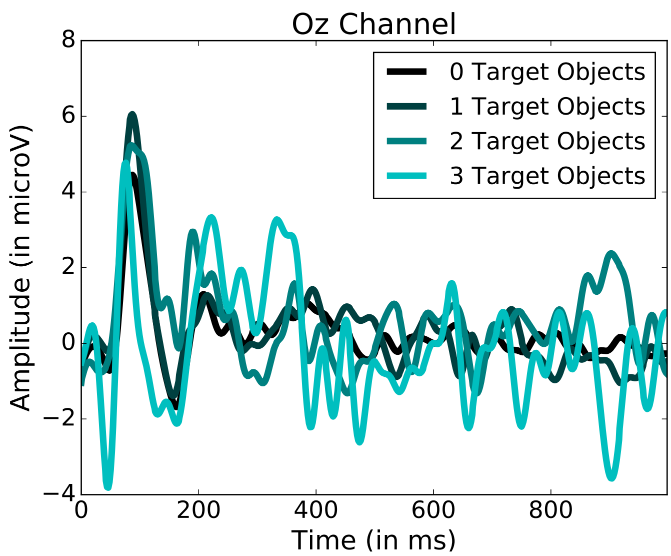


FIGURE 8.7: Averaged FRP at the Oz Channel, for different numbers of target objects present in the foveal region during fixation. This figure was generated using the data from a single participant.

When comparing the grand averages of FRPs, only small differences in P100 amplitude can be observed between the different numbers of target objects contained within the foveal region. In Figure 8.7, only FRPs obtained from the counting task were considered, as the target objects are not present in the free viewing task. The P100 potential amplitude does not seem to increase with the amount of objects present within the fixated visual region. This could be due to the presence of many other objects in the foveal region, which were also fixated at the same time due to the general scene clutter. This indicates that the amplitude of these potentials are linked much more to the quantity of information present in the foveal area rather than the relevance of these objects for a task. This aspect is strengthened by the preattentive nature of the P100 potential.

To examine this idea further, the P100 amplitude can also be related to other factors occurring during exploration such as the presence of target/non-target stimuli or depending on the task the participant had to perform. In both of the categories, subtle changes can be observed as shown in Figures 8.8 and 8.9.

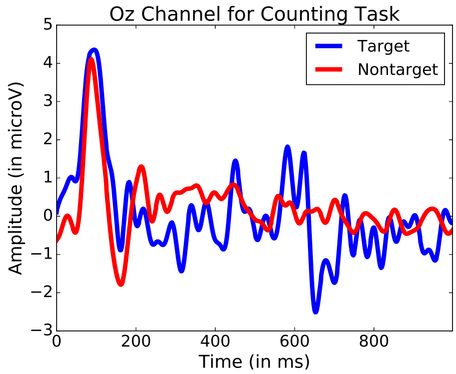


FIGURE 8.8: Grand average FRPs at the Oz Channel according to target fixations. Generated from one participant.

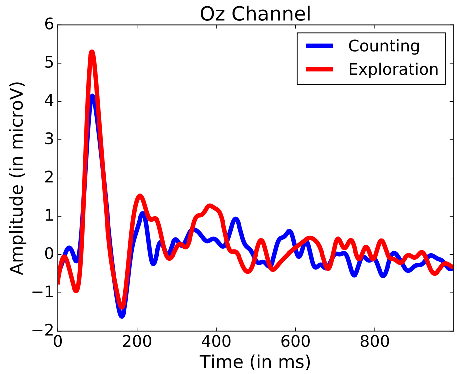


FIGURE 8.9: Grand average FRPs at the Oz Channel according to considered tasks. Generated from one participant.

In the case of fixations onto target objects during counting tasks, the P100 amplitude in target and non-target FRPs are close to each other (as seen in Figure 8.8). On the other hand, when comparing P100 amplitudes between counting and free viewing tasks (illustrated in Figure 8.9), exploration tasks present on average a higher amplitude than counting tasks. More analysis is necessary to explore the meaning of these differences in all considered participants. However, the increased P100 amplitude (in Figure 8.9) in free viewing tasks would indicate a greater attentional effort when freely exploring the scene rather than a guided one through the search of objects. This difference will need to be investigated in more detail to explore the differences in perception between the two tasks.

With these findings from the three considered types of analyses, the objectives we set above for the verification and expansion of our set of BCI methods can now be addressed.

8.5 Validating our Methods

In this study, the focus was put on the variation in tasks proposed to the participant and the presence of scene clutter to achieve it. Both of these aspects occur frequently in real-world scenarios where our proposed BCI setup ultimately aims to be employed. The purpose of this study was to see how well our bi-modal BCI deals with changes relating to both of these aspects. The objectives were thus twofold. At first, the applicability and utility of the methods established in Chapter 6 and 7 are verified. The new UdPV algorithm will discussed in greater detail than the three other methods below. As a second point, the changes in task are explored with our setup to verify if valuable information about task-related difficulties can be obtained.

8.5.1 Validating GDET, SPPQ and SFBA in Changing Interactions

In the experiment of Chapter 6, three methods (GDET, SPPQ and SFBA) were presented with which the combined ET and EEG data were analyzed to extract information about the interaction.

The GDET focuses on the use of relative ET metrics to understand which scenes caused the most difficulties to the user. Similar to applying the GDET to the study of Chapter 7, only two of the proposed metrics are used: the ratio of longer fixations per scenes as well as the ratio of revisits of inspected visual regions. As, again, no spatial segmentation of the scene is present, the last metric needs a modification to reflect task relevant information. Obtaining this information here also requires a slight system adaptation. In the context of this study, the relative amount of time spent looking at target objects takes the role of this last metric. As the presented scenes are static, the location of target objects can be obtained beforehand, minimizing necessary system adaptations. This third metric will also be used in the context of the free viewing task even if objects are not considered targets. Using these three metrics a t-SNE projection can be performed on the participants' data to examine the first intricacies of the collected data. These projections can be seen in Figures 8.10 and 8.11.

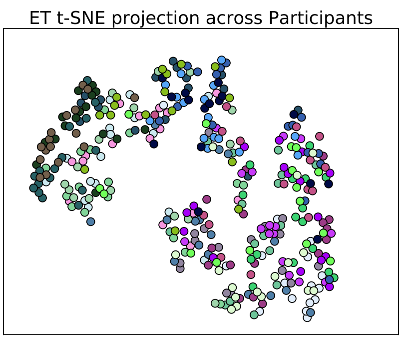


FIGURE 8.10: t-SNE projection for all 23 participants. Each color represents a participant.

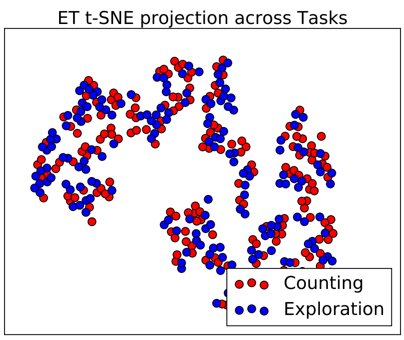


FIGURE 8.11: t-SNE projection for 2 performed task. Each color represents a task.

The resulting projections do not present any clear distinct clusters segregating the data in a meaningful way. Indeed, while ET differences have been shown to be present between counting and exploration scenarios (as discussed in Table 8.1), there is a consequent overlap between both task categories as can be seen in Figure 8.11. However, similar to the previous study discussed in Chapter 7, clusters can be identified more easily on participant level (as illustrated in Figure 8.10 where each color indicates a different participant). This indicates that participants operated the interface quite differently. As such, while remaining applicable, the GDET provides little discriminative power to identify tasks. As mentioned above, this may notably be due to the exploration of cluttered scenes which remains difficult to explore using broad ET metrics regardless of the considered tasks.

The SFBA concerns the identification of other related interaction manipulations. However, as there is no active manipulation within the visual search recordings, the segmentation of the EEG signal into fixation-bound epochs, varying in duration, contains no information relative to other manipulation tasks. While this method also remains applicable, the recorded data lack the information necessary to exploit them.

The SPPQ, the last method established in the initial study, considers the variation of the P100 amplitude to examine the changes in attentional effort provided by participants during an interaction. Here, P100 potential amplitudes have been shown to change between fixations. However, these changes are not correlated with the number of distinct objects contained within the fixated regions (see Figure 8.7). Due to the cluttered nature of the scene, each fixation contains a high amount of information regardless of the number of target objects. To still allow for a verification of the quantity of effort provided by the participants as it relates to visual complexity, each scene can be decomposed according to its visual saliency using the Graph Based Visual Saliency (GBVS) algorithm [Harel, Koch, and Perona, 2007]. The idea being that a non salient region presents a different cognitive exploration cost due to more difficulty to visually separate the objects. Each fixation is thus given a saliency score, averaging the GBVS value of the foveal region. In Figure 8.12, these scores are grouped and compared with the P100 amplitude extracted using the MLRd algorithm. Only two categories have been chosen. One category (named salient) considered fixation on regions with an average score above the mean saliency score of the entire scene. The other category (named non-salient) considered fixation with saliency scores below mean scene saliency. This composition has been made for the counting as well as the exploration tasks.

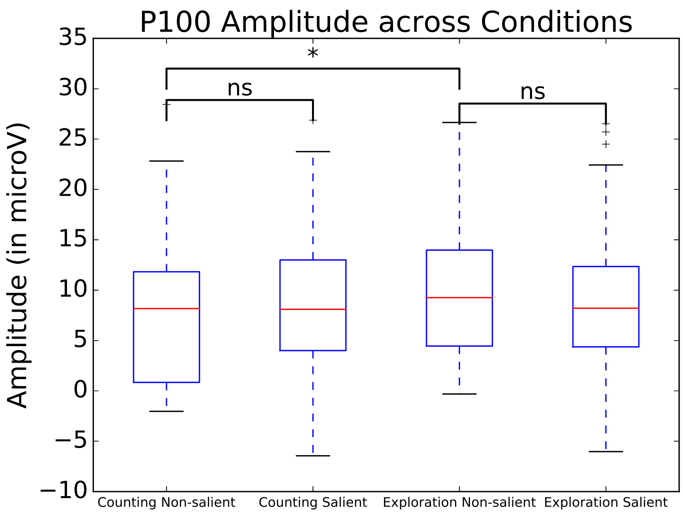


FIGURE 8.12: P100 amplitude depending on the fixation saliency (as determined by GBVS) and compared to median image saliency. 'ns' corresponds to non-significant distribution differences (Student t-test p-value> 0.1). '*' corresponds to a significant distribution difference (Student t-test p-value < 0.05). The present data corresponds to one participant.

The results are shown to vary strongly between participants. This is represented by the boxplot bars showing a 20 to 30 microV difference between second and third quartile of data. In most cases, there is no significant difference between tasks and between saliency groups (Student t-test p-value > 0.1). However, for 6 participants, a small but significant difference occurred between tasks in terms of P100 amplitude (Student t-test p-value < 0.05). Thus, general statements about attentional effort as relating to the visual saliency or the task are difficult to make : certain participants show an influence while others do not. When compared to the results from the two previous studies, the uniformly cluttered scenes seem to reduce variation in P100 during exploration. Indeed, participants need to constantly put effort into parsing the contents of a region regardless of the task, due to the high information density in each fixation. This SPPQ method thus remains applicable as well and still provides valuable information in validating the uniformly cluttered nature of the scene. It also suggests differences between users to be present.

Prior to doing so via the UdPV, we investigate the difference between tasks first. As differences between tasks have only been present to a minor extent (see Figures 8.12 and 8.6), but yet appear to be prominent in ET data (see Table 8.1), a deeper investigation into their differentiation needs to be performed.

8.5.2 Investigating Task difficulties with Our Setup

Small differences have been shown to be present between EEG recordings of the counting and exploration tasks. These differences have been observed in relative bandpower during scenes (see Figure 8.6) as well as in P100 amplitudes for some participants (see Figure 8.9). To quantify these differences in further detail a 5-fold classification has been performed on the collected FRP epochs from each task. The classification was done using a PCA feature extraction and an LDA classification algorithm. The classification has been applied between fixations on target and non-target objects as well as between task categories. As input, both the time and the frequency information has been used. The classification was performed for each participant individually and then averaged together.

TABLE 8.2: AUC ratios obtained after classification of different categories of FRP epochs. The results are averaged over all participants.

	Time Signal	FFT Frequencies
Counting Target v Non-target	50.3% (± 2.5)	63.6% (± 4.6)
Free viewing Target v Non-target	49.3% (± 3.6)	59.6% (± 4.3)
Counting v Free viewing	50.2% (± 2.5)	89% (± 9.3)

As shown in Table 8.2, using the FRP epochs as input features yields areas under curve of the ROC around the chance level (being 50% here). However, when using the FFT frequencies powers as input features, the classification AUC ratios exceeds chance level. This indicates that the properties contained within the frequency measurement present differences between the considered categories while regular time features in the signal do not. One possible explanation, already proposed in the study of Chapter 7, is the differences between ERP amplitudes. Rather than presenting different potentials, the considered classes of FRPs only vary in the amplitude of the already present ERPs. These ERP amplitude differences are however not necessarily limited to the P100 amplitude, as significant differences in P100 amplitude were only present for a small subset of participants, as mentioned above. A couple of facts can be noted from the AUCs. Firstly, there is a slight above average classification rate in the Free viewing task frequency domain. As one would expect chance level classification (as target and non-target have no meaning in free viewing tasks), this difference is likely due to slight variations in saliency as they are observed in Figure 8.12. This classification score slightly increases in the counting task, which is more expected as a cognitive difference is introduced. Lastly, the AUC between tasks is very high (89 percent in Table 8.2) indicating a strong difference in frequency composition of the signal. However, when compared to the results of Figure 8.12, this discriminability seems unrelated to the P100 amplitude. These differences show the need for more effective methods to analyze small potential differences.

When differences are clearly present between tasks (as shown in Table 8.2), the nature of these differences can be investigated graphically to be better understood. For this purpose, a t-SNE projection of the relative band power (obtained via SFBA) over all participants is performed. The results of Figures 8.13 and 8.14 show that differences

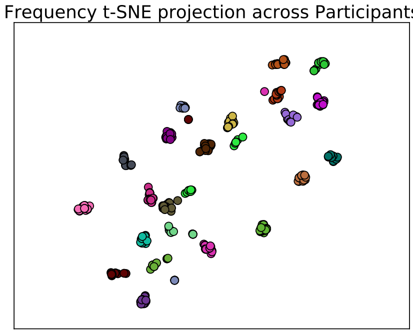


FIGURE 8.13: Frequency projection by participants.

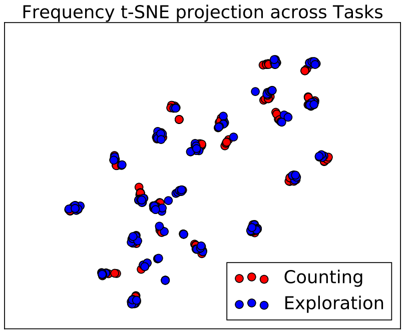


FIGURE 8.14: Frequency projection by task.

in frequency bands between counting and exploration tasks are indeed present. These differences are however not particularly prominent, with only certain clusters dividing themselves between red and blue dots in Figure 8.14). Differences between participants (see Figure 8.13) are however much more prominent. Indeed, clusters of the different participants show that, for some, there is a subdivision between counting and free viewing scenes in this projection (see Figure 8.14). However, as noticed previously, it is not present for all participants for relative frequency bands. Thus, in contrast to the results of Table 8.2, the differences between tasks seem to vanish when using features that are too broad (i.e. SFBA features). Still, we have shown that the proposed setup can be reliably and transposably be used to identify different tasks within one subject. The SFBA method could be expanded to encompass this approach. While further investigations are necessary to determine how differences between tasks within a participants can be detected, certain comments can already be made. The differences in ERP epochs are related to frequency changes, but do not appear to be linked to changes in any of the prominent or studied potentials. Aside from these differences in task, differences between participants are also visible and more prominent in the visualization of Figure 8.13. These can be investigated using the UdPV method.

8.5.3 Transposing the UdPV Method

Similarly to the study of Chapter 7, the differences between participants can be investigated as they relate to changes in the P100 amplitude. When differentiating between both tasks and comparing P100 amplitude standard deviation across all scenes with the completion duration of the entire study, certain linear trends can be observed. The comparison of the study duration with P100 amplitude variation is plotted in the Figures 8.15 and 8.16 for both tasks. These trends are however less prominent than in the previous study, presenting a considerable amount of outlier to the linear regression. This is reflected in the correlation coefficient which is around -0.32 for the counting task (displayed in Figure 8.15). When considering the free viewing task, no correlation can be seen (see Figure 8.16). As such, the variability between participants can change depending the interaction paradigm. It only presents a similar linear trend in difficulty evolution, as was observed in the second study (of Chapter 7), when observing the counting task.

To investigate user difficulty further, a relative metric of performance (rather than study duration) can be chosen. As already seen in Table 8.1, fixation duration is strongly task dependent and changes significantly between participants. Fixation duration informs partially about the ease with which the participant could parse the content of a

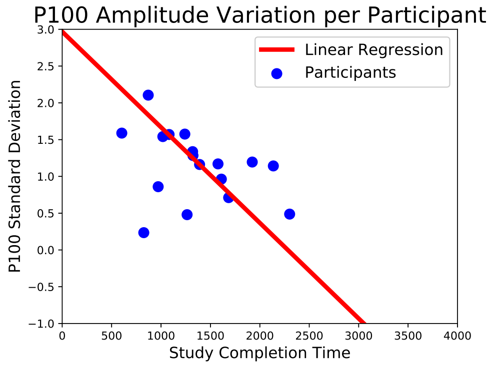


FIGURE 8.15: Counting task P100 variations.

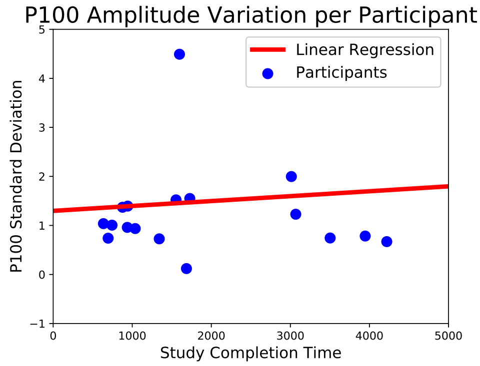


FIGURE 8.16: Exploration task P100 variations.

visual region. When comparing average fixation duration in scenes with the P100 variability, a general trend can be observed more clearly. This is represented in Figure 8.17. Indeed, two central cluster showing a downward trend as fixation duration increases are present. When fitting a linear curve onto the whole data (in black in Figure 8.17) a correlation coefficient of -0.5 is measured. Interestingly, when considering both central clusters separately, the resulting linear regression result in parallel curves, with improved correlation coefficient (-0.68 and -0.89 represented by the red and green curves in Figure 8.17). As proposed previously, the variation in P100 amplitude reflects the selectiveness of the attention of the participant. The presence of these two parallel clusters of performance indicates that a significant part of users may have used their attention in either one of two distinct strategies. These clusters are not present in the exploration task (as seen Figure 8.18), where a linear trend appears more clearly but, similar to the study completion duration, remains rather low. These differences in strategy appear to be linked to the currently pursued task.

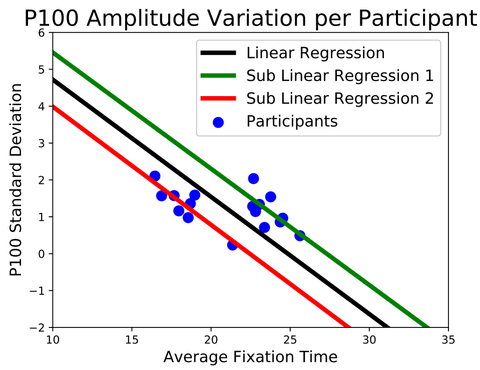


FIGURE 8.17: Counting task P100 variation

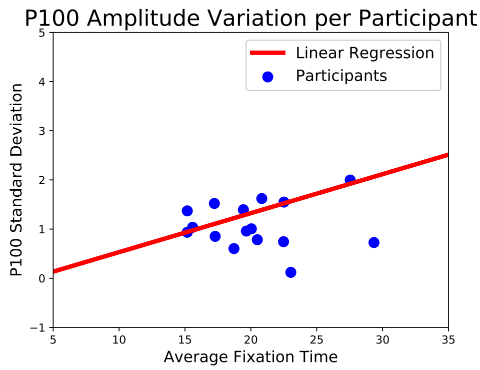


FIGURE 8.18: Exploration task P100 variation

Through both of these comparisons, participants can be grouped based on their counting task performance. These groups could then be analyzed in greater detail in order to identify underlying properties of the interaction which can be used to improve the interface. The UdPV method, putting in relation P100 and task completion metrics established in the previous study thus remains valid and transposable to other scenarios. Our proposed BCI setup allows for a reliable way to gauge user difficulties during interaction. To gain more information from it, still deeper analysis taking into account the specificities of the interaction would have to be considered. However, these types of analyses would be specific to each interface and thus beyond of the scope of this work.

8.6 Conclusions Regarding the State of the BCI

The two objectives from the beginning of this chapter have been validated. The first aim was to verify the transposability and utility of the methods established in the first and second study (i.e. the GDET, the SPPQ, the SFBA and the UdPV). The UdPV has been investigated in greater detail to verify its utility. The second aim was to investigate the differentiation between tasks in natural interactions. Differences were shown to be present in the frequency domain. All of these aspects give a clearer picture about the information that the proposed BCI system can provide to evaluate interfaces.

8.6.1 Final State of the Methods

The study has shown that the four previously established methods exploring scene complexity, task variations, local visual complexity and user difficulties are still applicable in this new natural interaction scenario with changing tasks. However, their utility is again significantly different from the first study. The presence of cluttered scenes has shown that eye movement remain mostly the same across conditions in terms of the established metrics. Looking at more subtle changes in eye motions such as average fixation duration (via GDET) has however shown to point to differences between tasks. As such, the application of ET metrics can still help to identify which task a participant is trying to undertake.

This is strengthened by the application of the SFBA method which relies on EEG frequency bandpower. While very subtle on an averaged level, a different segmentation of the signal, be it across fixation or scene still allows for an adequate separation of tasks and identification of the user's current manipulation goal. Indeed, here, it may be beneficial to amend the SFBA method, as more granular frequency changes, which are likely due to changes in potential amplitude variations, are well distinguishable on an FRP level. This property indicates that many other potentials are present in the signal which can be analyzed. Furthermore, this FRP approach, due to its accuracy, may allow for the creation of adaptive interfaces, identifying and helping specifically with the task at hand. This shows that the BCI setup enables to see fundamental changes in how an interaction is approached by its user. This method still needs to be harmonized and contrasted to the SFBA method.

Task changes have also been explored in deeper detail through the observation of the P100 amplitude via SPPQ, serving as a reference for attentional effort. The variation of these preattentive potentials however is less informative in the exploration of uniformly cluttered scenes. Indeed, as is in our case, attentional effort remains mostly constant regardless of fixated region, both in terms of relevancy and visual salience. This finding helps to validate the observation of the uniform cluttering of the visual scenes. Yet the variation of P100 can still be explored between participant showing the different difficulties they have faced or strategies they have employed.

The variation of P100 amplitudes, via UdPV, has shown to be correlated to the various performance related metrics such as study completion time and average fixation duration. Indeed, these metrics relate to the efficiency of the exploration of the scene when counting objects. The variation of P100 amplitude can give a glimpse at the expected performance of users within the interaction. This aspect could alternatively allow to identify which strategy of exploration a participant is using when considering a task.

Both applications of this correlation can help to quantify users performance and adapt use case scenarios accordingly. Overall, the proposed BCI setup and its methods can thus be applied in a wide range of scenarios and provide a variety of information, about tasks and users. As seen previously, element-wise and scene-wise difficulties can also be investigated when the considered interaction allows for it.

8.6.2 Further Possible Improvements

The exploration of cluttered scenes and task differences should be studied in further detail. Notably, in the proposed scenario, both of these aspects remained static within a scene. In more natural scenarios, a fast and more dynamic change of these aspects can often be present. As it stands however, the proposed methods do offer and improve a solid basis with which the proposed BCI system can provide general insights into various interactions. Similarly, the FRP frequency-based classification method needs to be examined in greater detail. Indeed, due to its use of frequency based features it may be possible to integrate or expand our SFBA method (which relies on relative frequency bands) to encompass this classification procedure. This step is however not performed in this work.

Considering, the results obtained from all previous studies, all proposed methods and the BCI setup in general can be compiled and a final conclusion on its utility with interface evaluation can be achieved.

Chapter 9

Discussion of the BCI and Findings of the Studies

The proposed bi-modal Brain Computer Interface and the set of methods presented in this work were designed to help with the evaluation of interfaces. Following this objective, the BCI and its set of methods were developed to respond to two aspects : being transposable between different interactions scenarios and providing information about difficulties the user is facing within these scenarios on different levels. Adhering to both of these aspects will ensure that our BCI is useful and helpful in assisting designers during the testing phase of HMIs. The implementation of both aspects has been successful but still shows some possibilities for improvement.

9.1 BCI Transposability

One of the main challenges when designing a BCI for the evaluation of interfaces is the considerable range of possible interaction scenarios. Developing a system that improves the understanding of the interaction often requires a significant workload, both to log BCI-relevant information and to change the interaction paradigm to create BCI relevant activity. The proposed BCI system tries to circumvent both of these workload sources by relying on an approach requiring minimal adaptations. The resulting implementation is thus more resource efficient and can be easily employed in many different interactions.

9.1.1 Setup Transposability

Practically, the proposed BCI system relies on three different sensors applied to the user. These are an EEG, an EOG and an eyetracker. All three of these devices are portable systems and can subsequently be attached onto users in either sedentary and or mobile settings [Chynał et al., 2016]. In the context of this work, only sedentary settings were considered. The application of the system in mobile settings is technically feasible but requires a deeper analysis into the resulting artifacts due to movements [Muthukumaraswamy, 2013]. For our purposes, the artifact correction methods detailed in Chapters 4 and 5 for all sensors recordings are satisfactory.

The chosen sensors do not require any cognitive attention from the user to record data. The sensors also do not constrain the user in their manipulation. With the exception to the EEG and EOG electrode-cables, which seldom bother the user's comfort if set up adequately, changes to interaction are negligeable. Other changes to interaction which cannot be avoided, such as the initial calibration of the sensors, were handled by adjustments done prior to the considered interaction. Calibrating the eyetracker is the most noticeable change. Thus, the setup keeps the core interaction paradigm intact, making any results obtained with sensors virtually equivalent to the same interaction performed without the BCI setup. These interaction adaptations for BCI could be reduced even further in the future if the EOG sensors were removed. The usage of EOG for eye movement artifact correction was found to be redundant as the ICA (detailed in Chapter 5) performs most of the necessary artifact correction [Verleger, Gasser, and Möcks, 1982,Wallstrom et al., 2004] to gain a clean signal.

Additionally, by minimizing the information from the studied system-side component of the HMI necessary for our BCI to function, no adaptations to the hardware or software of the system need to be performed. This reduces the overall adaptation workload of the considered interfaces and ensures that the bi-modal BCI can easily be applied in new situations. Recordings of EEG and ET can even be performed on different computers than the ones used to run the studied HMI interface. This was done successfully for the study discussed in Chapter 6 by relying on one of our methods detailed in Chapter 5 : the FCSync method. Relying on the aMLRd, the FCSync only probes the EEG and ET recordings to resynchronize these recordings, allowing our setup to provide appropriate data regardless of the intricacies of the physical setup.

From an overall technical perspective, the proposed BCI setup is applicable in a large scope of situations. In order to prove its utility, the considered analysis methods have to offer valuable information about a user's perception of the interaction. Furthermore, to be valuable across interaction situations, the methods have to remain applicable in the same scope for different interactions scenarios in which our BCI is used.

9.1.2 Method Transposability

The proposed evaluation methods (i.e. the GDET, the SPPQ, the SFBA and UdPV methods) rely exclusively on the observation of three types of signal. These are the EEG signals, the Eye movements extracted from gaze locations and the observation of fixation-segmented EEG epochs (of fixed or changing length) across the study. To properly extract these EEG epochs and analyze their resulting ERPs, timestamps must be shared accurately between the recordings of the EEG and ET. This is achieved through the FCSync method, mentioned above, or equivalent techniques.

In our application of the evaluation methods, we studied interaction on multiple levels. Most frequently, we look at general difficulties between interaction scenes and momentary changes during the same interaction scenes. For observations between scenes (as done in the SFBA and GDET methods), the ET and EEG recordings are observed over the entire duration of a scene exploration. For the creation of such scene level epochs, information from the system side is necessary. However, this type of information is light (i.e. scene start and end points), requiring only a timestamp and a label to enable the general segmentation of the data for broad interaction observations. With respect to frequency observations on scene or task level (as done in the SFBA method), only rough estimates of the starting and end points of a scene or task are required to obtain relevant and interpretable information. Thus, the required system information for using these methods can easily be provided by hand (e.g. through a manual trigger by the designer, not integrated into the system-side component) or through inexpensive system adaptations (e.g. logging scene change timestamps).

When studying more momentary changes due to specific interaction moments, adapting the system to log specific events however becomes necessary. When these events are only contingent on button presses, system adaptations typically remain inexpensive. However, for more subtle ERP studies (i.e. P300 or N400 paradigms) the requirement can quickly increase in workload and complexity. By combining EEG and ET recordings, the studies of FRPs and fixation-bound segments have shown to provide valuable insights into the interaction. Both recordings are part of the transposable BCI setup and do not require system adaptations. Thus, the study of FRPs, segmenting EEG data through the context information provided by fixations, improves the transposability of the methods further. Notably, the FRPs enable the application of the SPPQ method in many different scenarios.

Our last method, the UdPV, relied on comparing FRP changes to whole interaction performance metrics (i.e. completion time and average fixation duration). Similar to the system information required for the SFBA and GDET methods, obtaining these simple performance metrics can be logged by hand or emerge naturally from the recorded data.

Finally, the part requiring most system adaptations is the acquisition of ground truths. These are necessary for an initial categorization of the information present within the recorded data. In our considered studies, all stimulation was based on static images which could be analyzed separately from the interaction without losing information. While a prior labeling and decomposition of the scenes allowed to obtain valuable ground truths, this sort of analysis is not as easily transposable between interactions as other system-side information used above.

Overall, the proposed methods used for interface analysis did only rely to a minimal extent on system-side adaption or information. While their application still needs to be verified in more dynamic and mobile scenarios, their application so far suggests a likely resource efficient implementation in new interaction scenarios. Thus, the transposability of the BCI system is essentially validated. However, to be effective for interface evaluation, the utility of the presented methods has to be demonstrated as well.

9.2 BCI Utility

The series of proposed methods (i.e. GDET, SFBA, SPPQ and UdPV) offer insights into four different aspects of interaction. These correspond to difficulties on a scene element level (SPPQ), difficulties on a whole scene level (GDET), difficulties on the task level (SFBA) and difficulties on the user level (UdPV) respectively. The methods sometimes overlap between categories and have different advantages and inconveniences in their application. The methods are integrated into the data acquisition process proposed by the BCI setup. This is illustrated in Figure 9.1 where the information necessary and provided by each method is detailed. All the methods and their informative utility is summarized below.

9.2.1 Difficulties Relating to Individual Scene Elements

The difficulties relative to individual interface elements are traditionally explored using eyetracking [Greene, 2006]. Notably, through the analysis of fixation duration, specific objects are identified which are of particular relevance for the user. However, as this traditional approach does hold only little information about the individual scene elements, FRPs are also observed for additional information. Most notably, the variation of the occipital P100 potential amplitude gives precious information about the attentional effort invested into parsing the contents of a fixated region.

The P100 amplitude has shown to change depending on the visual complexity of the fixated region. This property was observed through the SPPQ method and allows to identify the regions of high amounts of visual information within the scenes. It is however only usable when major changes are observably present in the scene. Indeed, when scenes either present relevant information very distinctly (as in the study of Chapter 7) or in a uniformly cluttered fashion (as in the study of Chapter 8), the P100 amplitude does not change significantly during the visual search. However, the examination of the

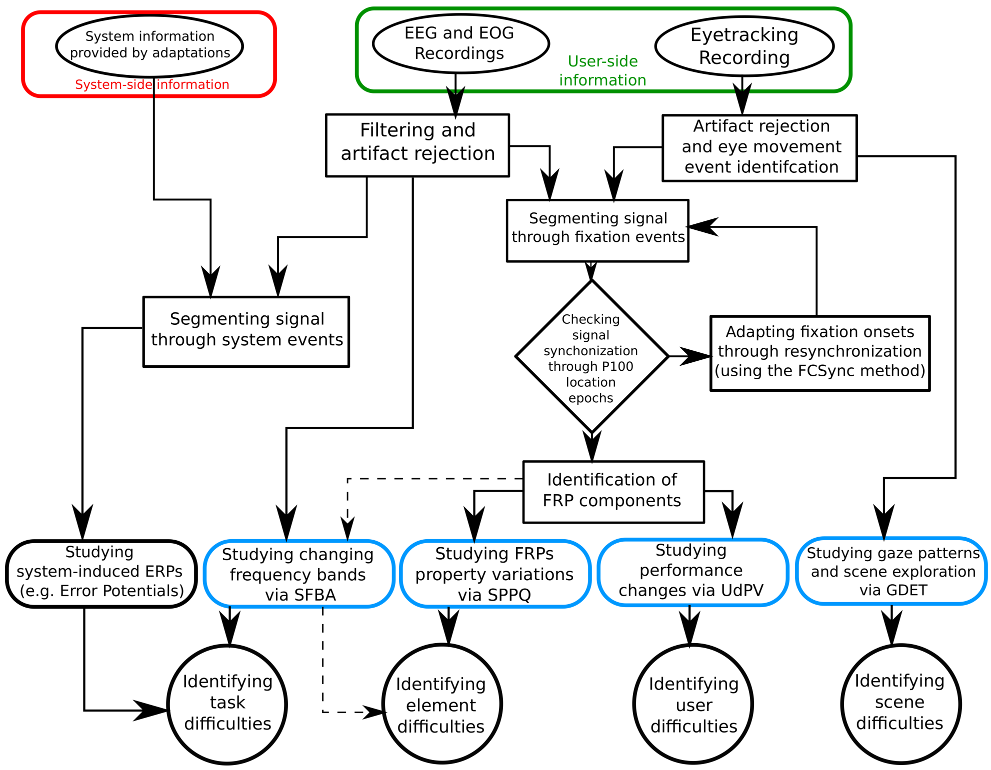


FIGURE 9.1: Process of obtaining biosignal information and using them to obtain information about difficulties user's encountered during manipulation. The blue framed methods are the new methods introduced in this work.

P100 amplitude remains of interest as it provides a way to verify the visual information distribution across the scene. This can give HMIs designers an additional tool to examine the scene composition.

The information contained in one second of FRP epoch can, in some scenarios (as in the study of Chapter 8), still be used to identify perceptual differences between target and non-target objects. These differences appear in the frequency space and can be associated with changes in peak amplitude between target and non-target categories. Although these changes were only slightly above chance level (around 60 % with the help of LDA classification according to Table 8.2), they can be investigated in more detail by exploring the use of methods for peak analysis (such as the aMLRd). A possible improvement of the SFBA method could be considered in the future to work with such frequency variations (as proposed by the dotted arrow in Figure 9.1). In all presented studies, only the P100 potential has been explicitly studied. However, the examination of other EEG potentials (such as the frontal N100) may provide additional information valuable to interface evaluation, potentially through its latency changes [Callaway and Halliday, 1982]. Indeed, other ERPs of interest are likely present in the studied cases.

9.2.2 Difficulties Relating to Entire Scenes

Difficulties relating to the entire scene have been investigated within the study of ET metrics. A set of three ET metrics was chosen to maximize method transposability between subjects and tasks. The fractions of longer fixations, ratios of region revisits and ratios of task relevant activity were practically used as parameters for the analysis. The first two of these metrics represented an abstraction of the temporal and spatial exploration rhythm of the scene. This last metric relies on system-side information, which can pose issues depending on the nature of the studied interaction. In most cases however, only simple system information is required. As described above, this information is most frequently obtained by annotating the scene prior to the study. Afterwards, the t-SNE projection of these metrics provides a graphical distribution of the scenes which can be used for a qualitative categorization according to the scenario type and its perception by the user. The method analyzing the three ET metrics has been named GDET in this work.

This categorization via t-SNE projection is however only possible in situations where major structural differences are present between scenes. When presenting scenes which are clearly organized or cluttered (i.e. the studies of Chapters 7 and 8), these differences cannot be observed or classified using the GDET method. In this regard, scene-wise difficulty identification is similar to the identification of element-wise difficulties seen above. Beyond this application, this way of graphically categorizing can however still be used to identify if differences between users are present. Indeed, as absolute scene duration is not present in the metrics, only the relative gaze metrics remain. Thanks to these abstractions, more subtle but meaningful differences between users become apparent, especially showing that every user presented a very different gaze pattern. Here the t-SNE projection approach can be used to verify if all presented scenes show comparable gaze patterns when explored. If other dimensions, such as user differences are better represented (e.g. like in Figure 8.10), scene-wise difficulties have most likely an insignificant role in the scene exploration. Such conclusions allow to better hone in the specific part of the interaction which needs improving, identifying scenes or other aspects of interest. More studies are required to effectively identify the extent of information which can be obtained through the methods relating to scene difficulty.

9.2.3 Difficulties Relating to the Task

Difficulties relative to the task currently performed by a user have been identified using frequency differences in segmented epochs. The method which uses these differences has been termed SFBA. The segmentation used to identify these differences can be done both at fixation onset, for a one second duration, or along the entire fixation. By comparing the relative frequency bands and performing a classification using FFT frequencies extracted from these epochs, the currently performed task can be determined quite effectively with an average AUC of 80 % (see Table 8.2 of the third study and Figure 6.17 of the first study). This allows to accurately identify what the goal of the user was at a given moment. Potentially, this identification could also be used in a live setting where the interface actively adapts itself to the task of its user.

Depending on the type of segmentation of the data, inferences can be drawn about the origin of the differences between tasks. In the case of fixation-bound epochs varying in length, the observed variation in task can be due to the user performing a motor action (e.g. in the study of Chapter 6). In the case of same-length fixation epochs, the differences can be due to perceptual difference in the early moments of fixation. Each approach helps to gather insights into how the interaction is handled by users and how to improve it. Indeed, a difference in frequencies between fixations indicates how often or significantly users changed their modus operandi. This knowledge can help in the HMI design phase to identify if the user performs actions too often or, on the contrary, not often enough and allows to adjust the flow of the interaction.

Finally, depending on the presence of other small tasks during the interaction, the observation of more noticeable ERPs is possible . This can include potentials like ErrPs or more specific ones such as the P300 or N400 potentials. The presence of such potentials can be particularly informative in assessing how the interface is perceived by the user, as they provide clear indications about a specific type of cognitive process. Situations in which such types of potentials occur are rare and subsequently, seldom usable. Nonetheless, like the less noticeable potentials discussed above, the exploration of these other ERPs can be expanded upon by using property exploration methods such as through the aMLRd technique [Wobrock et al., 2017] if the necessary system-side information is present for their observation.

9.2.4 Difficulties Relating to User Strategies

The detection of difficulties relating to individual users is more challenging. Indeed, these difficulties only become apparent through the previous methods (i.e. GDET, SPPQ and SFBA) when other sources of difficulty are mostly absent (as shown in the second and third study). In these conditions, where strategies of an individual user gain larger prominence, the variation of the P100 potential amplitude during the execution of a task can be considered. These variations of the amplitude have shown to correlate with the completion time of a task, indicating that users with a more selective attention complete the task more quickly. We termed this method the UdPV. This method could be used to identify which users have more difficulties with the interaction scenario. Tailored help can then be provided when noticing the users' less selective attention.

Multiple clusters of strategies can also be identified when comparing P100 amplitude variation to average fixation duration (e.g. see Figure 8.17). This indicates that factors other than selective attention also play a role in task performance during interface exploration. Associating a user to one of these clusters thus enables a deeper analysis of the difficulties the interaction poses to the general groups of users. This aspect was not further investigated but could be considered by monitoring coherence measures between different electrodes [Hanslmayr et al., 2007]. Coherence represents the influence one brain region has on another. Examining and comparing such connectivity metrics between participants of P100 variations cluster may allow to shed light on where the difference in performance lies.

9.3 Expanding on BCI Insights

The aspects of transposability and utility, discussed in the previous sections, are the main results of this work. However, there are more potential applications for our proposed BCI framework. The bi-modal framework offers an abundant amount of data which has to be properly analyzed to extract meaningful data from it. Additional analysis may shed light onto new peculiarities in the interaction. This section offers ideas on how to expand it further.

9.3.1 Comparing Study Scenarios

One of the analyses which was not performed so far is the comparison of interactions and interfaces among the three studies. Due to the transposable nature of the proposed BCI and its methods, this technique can confront interface performances. This can notably be done by comparing P100 amplitudes by averaging over all subjects who participated in the same study of one interface, thus limiting participant variability. Here, we use the data collected from all studies detailed in the previous three chapters. Figure 9.2 shows the averaged Oz channels across all participants of all three different studies.

As can be noted from Figure 9.2, the third study, which only featured a counting task in cluttered scenes, presents the highest average amplitude of the P100 potential among the three studies. As this amplitude reflects the attentional cost of the average fixation,

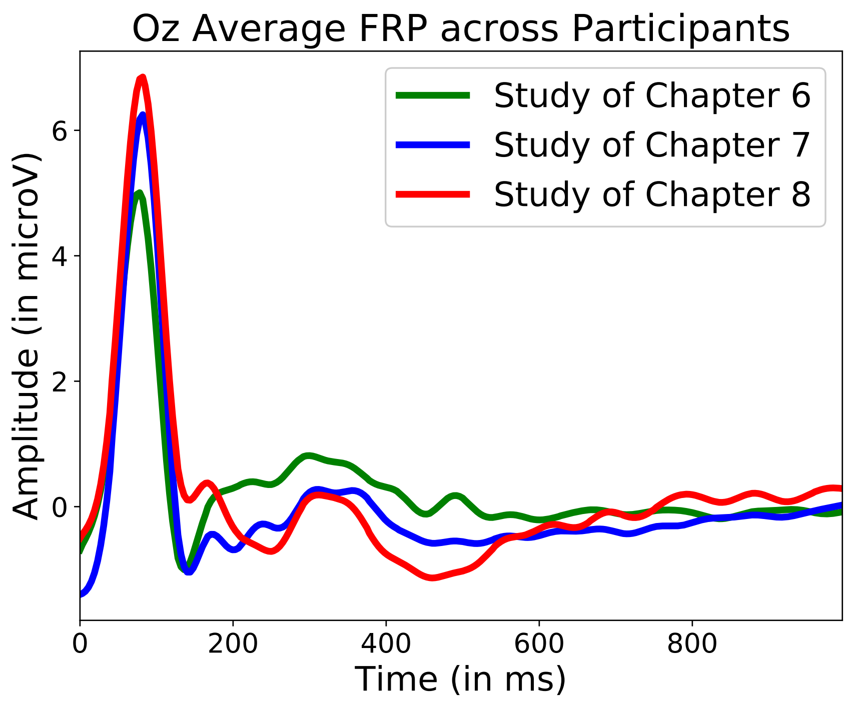


FIGURE 9.2: Averaged FRPs of the Oz Channel taken across all participants for the three presented studies.

the general cognitive demand of each study scenario can be estimated in relation with the others. This might present an additional tool to quantify the effort performed by participants. To investigate this aspect further, additional experiments are necessary.

9.3.2 Overlapping Epochs

Another point of contention concerning this setup deals with the different duration of fixations happening in quick succession to one another. As the general length of a fixation epoch is around 250-300 ms and certain ERPs can appear around a second after stimulus onset, overlap between different potentials can occur frequently. This is particularly problematic as the P100 amplitude is quite low when compared with other potentials. Indeed, the overlap with other more prominent potentials would mask the P100, rendering the previous methods less applicable. As stated in Chapter 5, the aMLRd method is not capable of extracting potentials which are occluded by more prominent ones.

To better attribute the different potentials to the epochs they belong to, an approach of untangling these potentials can be considered. Such an approach has been proposed by Woldorff, 1993. As mentioned above, this approach relies on the identification of all present potentials and their conditions for occurring. From this point, ERPs are recreated by relying on a set of rules established from prior observation. Specific potentials can then be cleanly extracted from epochs even if strong overlap between epochs is present [Ehinger and Dimigen, 2018]. This allows notably for a better extraction of the underlying P100 and thus an improved identification of the attentional effort invested by the user. This method could help broaden the scope of applications for fixation-segmented epochs and FRPs within them. However, this method also requires more extensive system-side adaptations to obtain the relevant ground truths to properly label each epoch.

9.3.3 Other Fixation-related Potentials

The exploration of potentials within this work has almost exclusively occurred in the Oz occipital channel. In the context of our studies this was notably due to the absence of other stable potentials during the entire manipulations. However, as already mentioned, other notable potentials can be present during fixations. Mainly the negative N100 potential can be observed in the frontal channels during the appearance of the P100 potential. These potentials vary similarly to the P100 potential but may be more difficult to observe reliably when observing scenes of varying visual information.

These potentials thus may contain other information which could be of interest for the evaluation of the interaction. Some of these potentials have already shown that they contain information about facial identification or other similar low-level cognitive processes. These aspects, observed in conjunction with Gestalt theory [Smith, 1988] may help to better advance in the exploration of the interaction scene than before.

Finally, other, later-occurring potentials can also be considered. These potentials suffer from similar overlapping problems as discussed in the previous section, but could offer new information into the interaction such as insights about certain specific actions or misunderstandings. These potentials however occur at unknown moments during a fixation-bound epoch and no accurate way to localize them is yet available. New ways of approximating the event onset inducing these potentials are required to continue with their investigation.

With the potentials actively considered within this work, a great number of inferences about the interaction and its quality can already be made. These new results can possibly be integrated in the future into further adaptive systems to improve interaction and make interfaces more usable by an even broader variety of users and systems.

Chapter 10

Conclusion and Perspectives

10.1 Summary

In this work, we aimed to develop and to test a BCI setup and a set of BCI-related methods which help with the evaluation of Human-machine interaction, gathering information about a user's perception. Our BCI setup combined EEG and Eyetracking sensors, allowing for the observation of FRPs. Our set of methods had two objectives : to make the data from these sensors always available and to analyze them in order to provide information about specific difficulties relevant to interaction.

For data availability we proposed one new method : the FCSync method, which ensures synchronization of recording modalities in all situations. For data analysis, we provided four methods : The GDET method, which provides information about difficulties linked to overall scene complexity, the SPPQ method, which provides information about difficulties relating to specific elements the user has perceived, the SFBA method, which provides information about difficulties relating to the task a user had performed and the UdPV method, which provides information about difficulties relating to specific user strategies or bias when interacting.

These methods have been tested in three separate studies where participants were monitored using the bi-modal BCI. The first study proposed a set of scenes varying in the amount and color of shapes present within them. Participants were instructed to find specific shapes within the larger scene. The FCSync method was used in this study and was also tested using the data of the two following studies. The controlled nature of this interaction scenario ensured well established ground truths and introduced difficulties at three different levels : General scene layout, specific location containing more difficult elements to perceive and changes in task between exploring the scene and interaction with objects. Thanks to the presence of these difficulties and ground truths, the GDET, SPPQ and SFBA were tested and found to provide insights not obtainable through classical interaction evaluation methods. Specifically, the GDET could discriminate between scenes according to number and color of shapes. The SPPQ observed P100 amplitude variation, showing that location featuring many objects were perceived differently by the user. Finally, the SFBA was shown to be able to discriminate between fixation containing user input and those which did not contain any. All three methods were also shown to rely on little information specific to the interaction, allowing them to be easily transposed to another interaction scenario. As this study featured a controlled and artificial interaction, these findings had to be explored in a more natural interaction.

The second conducted study proposed a visual search and object selection task in natural kitchen scenes. These scenes showed few objects. These objects could vary in their relevance and novelty to the participant. The three methods established in the first study were tested here. Due to the lack of variation, while still being usable, the three previous methods did not show much variation between the scene. This was to be expected as variation in task, number of elements in a fixation and general scene layout were not noticeable. This dearth in complexity made user differences more apparent. Here the transposable UdPV method was formulated and tested. It showed that the variation in P100 amplitude correlated with performance metrics (i.e. study completion time) of the users. Inferences about the reason of an individual user's performance can thus be made with greater precision. Additionally, FRPs were shown to present different frequencies depending on the task relevance of a fixated objects. These two findings, as well as our three previous methods were tested in another environment ensuring their utility and transposability.

The third and last study of this work proposed cluttered natural scenes which had to be explored either freely or by searching and counting a specific type of object. This type of scene ensured the presence of difficulties as they relate to the whole scene composition, individual scene elements, interaction task and varying user strategies. All four developed methods remained applicable in this new scenario. The GDET, SFBA and SPPQ presented no particular changes as there was no major differences between scenes clutter, no fixations which presented less information and the absence of other interaction events. The SPPQ showed however a small difference in potential when comparing P100 amplitudes between tasks. This observation was confirmed by the frequency comparisons of the previous study. Additionally, the UdPV showed difference between participants also being present in P100 amplitude variations. Here multiple clusters of users were also present, indicating the existence of other factors which influenced performance.

All three of these studies proved that our set of analysis methods, the GDET, SFBA, SPPQ and UdPV are applicable and useful to examine the difficulties present in the interaction. Indeed, they provide information at the four proposed different levels of interaction : at element-level, task-level, scene-level and user-level. Even when the information provided by them is limited, they still allow to observe a uniform difference of difficulty in the entire interaction. The methods are thus useful and transposable, adding value that classical HMI evaluation cannot provide.

10.2 Observations and Criticisms

In this work, we have discussed a bi-modal BCI setup which can be employed to evaluate a Human-machine interaction. This was done with the perspective in mind to improve the usability of the interface by making it more effective, more efficient or more satisfactory to use. To do so, grievances during manipulation had to be identified and associated to a specific source. These sources could be grievances specific to the user, to the interface lay-out, to the current task or to specific elements within the interface layout. One of the major issues in applying BCI to this problem is that the inspected interaction requires adaptation to fit this new way of observation.

Through our BCI approach however, the combination of Eyetracking and EEG has proven to circumvent these necessary adaptations. Indeed, these dual modalities can be recorded separately from the inspected system without changing interaction or losing their informative value. This has been shown by Wobrock et al. (2019) where the FCSync method was tested and shown to be reliable. Furthermore, the presence of Eyetracking technology has been able to identify the moments and locations of interaction that are most relevant to the user. Indeed, the segmentation of the interaction into fixations, punctuates different moments of interaction, allowing to gain context-relevant information about an interaction without requiring to query the underlying system. Thus, the proposed BCI proves to be easily transposable. In this work, this transposability was key goal and the proposed interface analysis methods were subsequently built around this idea. However, eye fixations remain difficult to contextualize on their own, as the information they contain is not specific to the interaction. This can make it difficult to locate other relevant moments of interaction (e.g. feedback communicated during fixation, motor actions performed over multiple fixations, ...) not linked to eye movements. The context information related to eye movements has however been shown to be reliably observable in different tasks. Comparisons between interactions using our methods were shown to be possible in Wobrock et al. (2016) and used in Wobrock et al. (2017).

The fixations used for segmentation of the EEG data occur naturally through visual exploration of the interface and can serve in different ways to analyze interaction. However, before combining the modalities to consider Fixation-related Potentials, the recordings are inspected separately. Using EEG data in its raw form still allow for various insights into the manipulation (notably via SFBA). Most of these insights only concern information about the higher level of grievances such as the general performance of the subject or about the difficulty of certain very distinct tasks or scenes. Most of the time however, the EEG signal alone gives little information unless really significant different levels of difficulty are present. The ET signal follows the same principle (see GDET) but can also be used to visualize gaze heatmaps. These heatmaps enable the identification of specific interface elements grabbing the users attention.

The obtained fixations then also help to punctuate the EEG signal analysis by offering curated segments for analysis. These segments contain the P100 potential, which is created by the fixation paradigm itself. As shown in this work, the amplitude of this potential reflects the attentional effort expended by the subject. While it may not be possible to say what exact P100 amplitude corresponds to what level of difficulty, the P100 amplitude can still be contrasted between different categories of fixations (via SPPQ) to gain a reference point about which categories pose more of a perceptual challenge to the user. For comparisons of full interface lay-outs, tasks or specific individual interface elements, the difference in potential amplitude could also show how the user reacts differently to the interaction. First and foremost however, the P100 amplitude variation helps to understand what attentional strategy of the user yielded a more efficient interaction (via UdPV). Overall, one of the main takeaways of our setup is the utility and reliability of the P100 potential to offer an additional insight into interaction. As shown in the last study, other potentials or frequency variations are also present in the signal which could be used but which have not yet been reliably captured through our methods. However certain of our proposed methods (i.e. aMLRd) could help to create new techniques enabling the exploitation of these EEG potentials (by examining the variation in amplitude, latency or morphology of these new potentials).

Overall, all our methods allow to identify the factors which created grievances to the users. Thanks to this, specific aspects of the interface can be targeted for improvement in order to guide the users toward an optimal use. The iterative design cycle used to improve the interaction can gain quantifiable information more easily through the metrics given by our methods. It is however also notable that the variation between the users is always quite high, making it difficult to find a singular law for improving the interface, even with more quantifiable metrics. Nevertheless, the study of this P100 potential offers a new way to engage with interaction and improve it both in term of efficiency and satisfaction. Our BCI has shown its utility in Wobrock et al. (2020).

The BCI setup also allows for many other types of applications and for the extraction of other properties. To investigate these and gain their informational value for improving human machine interaction, more research is required. Notably, the four analysis methods established during this work need to be tested in further scenarios proposing a more dynamic range of activity. Additionally FRP frequency observation can be improved and integrated into the set of methods, which then may allow to arrive to further conclusions.

With our BCI, a base system as well as a set of methods has been provided which both contribute to a enhancing the evaluation of interactions.

10.3 Perspectives

The exploration of all the above properties has so far only been considered in visual exploration. While essential to interaction, most interactions are performed as a continuous active dialogue between the system and its user. Considering all types of manipulation would, for example, include actions as keyboard presses or additional motion performed freely during interaction. Such actions are essential to common interaction and analyzing them through the proposed BCI setup would offer a new way to advance interaction. A possible future direction would be the adaptation of this setup, both in terms of methods and hardware, to allow the analysis of active manipulation situations. Their evaluation has been proposed here through the analysis of ERPs (e.g. ErrP evaluation in Chapter 6) but more reliable user-side markers would need to be discussed to investigate them more reliably as we assume system-side information can not be always obtained reliably.

Similarly, another common aspect to interaction, especially with digital interfaces is the presence of multiple varying tasks changing very quickly. To investigate these quickly changing tasks properly, an investigation using a more controlled environment may be necessary as a first step. A Virtual Reality (VR) application would present an ideal environment to investigate such a broad variety of interactions due to the full control over the interaction environment it provides. Our proposed setup can easily be used in a VR environment (supposing the presence of VR goggles with integrated eyetracking) and thus would require minimal system adaptation while being applicable to many different complex interaction situations. Indeed, the nature of VR interaction allows for a convenient implementation of the proposed system and permits to easily inspect variations alongside new parameters of interaction. An interaction, such as cooking a dish in the kitchen could be performed in a VR environment and all movements and actions could be logged in detail without influencing the interaction further.

In this regard, another crucial aspect of interaction consists in physical manipulation of the interface and the underlying system, notably as it relates to grasping of objects. This aspect was investigated in parallel to the described studies by Shirley Mey [Mey, Koester, and Schack, 2018a ; Mey, Koester, and Schack, 2019]. The presence of difference in FRPs according to grasping was shown, indicating possibilities of integrating identification of finer motor motions into the evaluation of an interface.

Finally, as the observation of the P100 potential remains a main pillar used in the context of the interaction, a deeper analysis of this potential could be performed by increasing the number of electrodes used for recording. Indeed, when applying a higher spatial surveillance of the occipital region, a more nuanced observation of the P100 potential can be obtained [Robinson et al., 2017]. These observations could then allow for a more precise estimation of the visual attention, potentially leading to the reconstruction of the attentional mapping of the user, as it can be performed through face-relation potentials [Nemrodov et al., 2018]. All these different directions of exploration show that there is much more research to be done before applying this setup to the real-time adaptation of interaction to the needs of the user.

Appendix A

Study Two : Scene Stimuli

Tutorial Scene of Study 1 :

The 40 Scenes of Study 1, in ascending order :

Scene 1 : Throw away packaging.

Scene 2 : Wash dishes in the sink

Scene 3 : Put groceries in the fridge

Scene 4 : Clean cooking plates

Scene 5 : Bake a cake

Scene 6 : Make spaghetti

Scene 7 : Make rice pudding

Scene 8 : Mix a cocktail

Scene 9 : Cut the vegetables

Scene 10 : Make a mushroom omelette

Scene 11 : Make a smoothie

Scene 12 : Make a fruit salad

Scene 13 : Boil an egg

Scene 14 : Grate Cheese

Scene 15 : Make Orange Juice

Scene 16 : Get drunk

Scene 17 : Make Coffee

Scene 18 : Wash dishes in the dishwasher

Scene 19 : Make a pizza

Scene 20 : Make a Hamburger

Scene 21 : Prepare Breakfast

Scene 22 : Make Pancakes

Scene 23 : Set the table for Soup

Scene 24 : Make hot chocolate

Scene 25 : Peel an orange

Scene 26 : Make Cookies

Scene 27 : Make mashed potatoes

Scene 28 : Put away sweets

Scene 29 : Make tea

Scene 30 : Make a sandwich

Scene 31 : Make strawberry marmalade

Scene 32 : Get lemon zest

Scene 33 : Light a candle

Scene 34 : Make Caramel

Scene 35 : Separate Egg Yolk and Egg white

Scene 36 : Make Whip cream

Scene 37 : Toast toast

Scene 38 : Make French fries

Scene 39 : Put marmalade on toast

Scene 40 : Clean the sink.

Appendix B

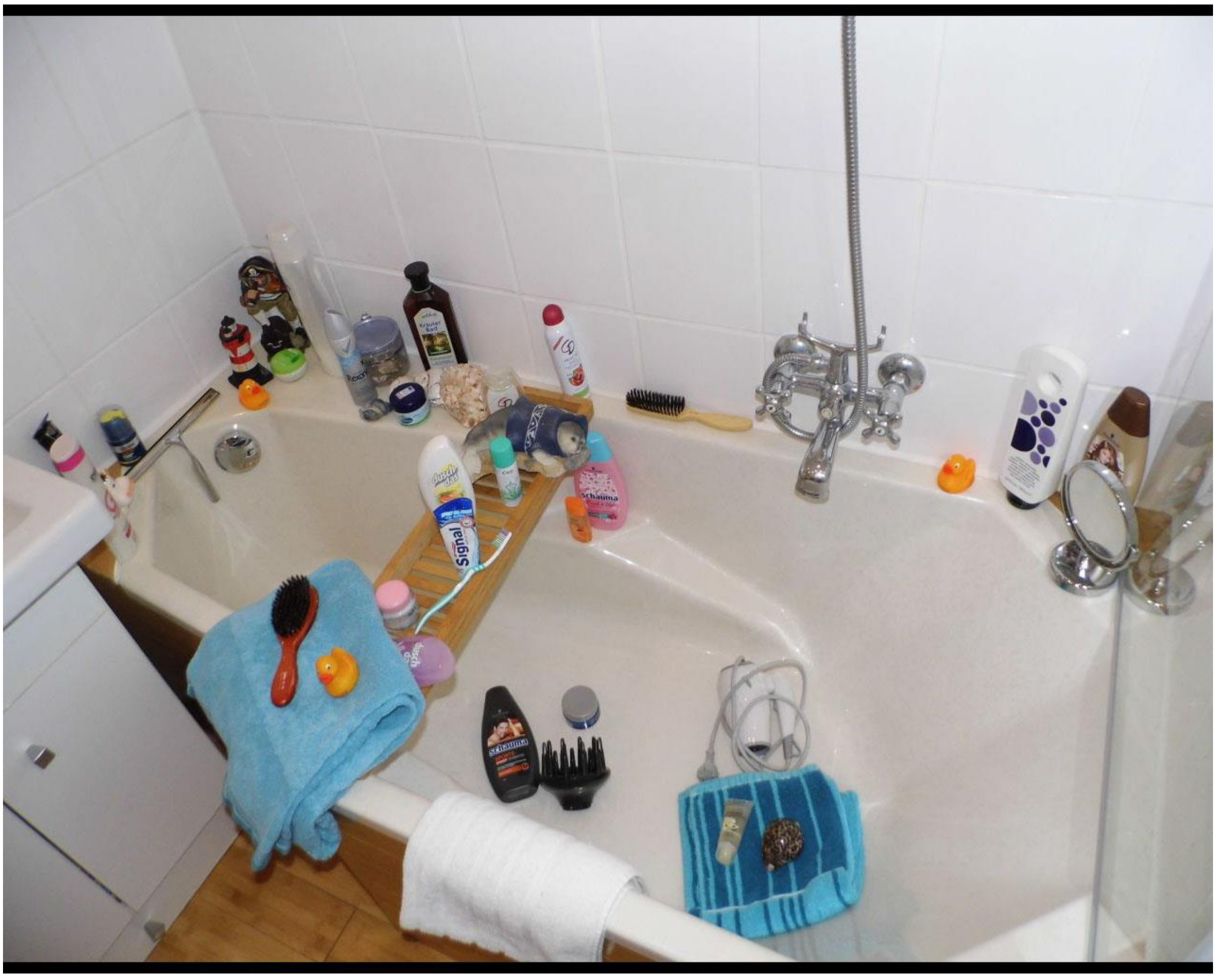
Study Three : Scene Stimuli

Scene 1 : Counting task of either coffee cups and flower pots.

Scene 2 : Counting task of either forks or knifes

Scene 3 : Couting task of either screwdrivers or pliers

Scene 4 : Counting task of either fruits or vegetables



Scene 5 : Counting task of either shampoo bottles or plastic figures

Scene 6 : Counting task of either books or folders

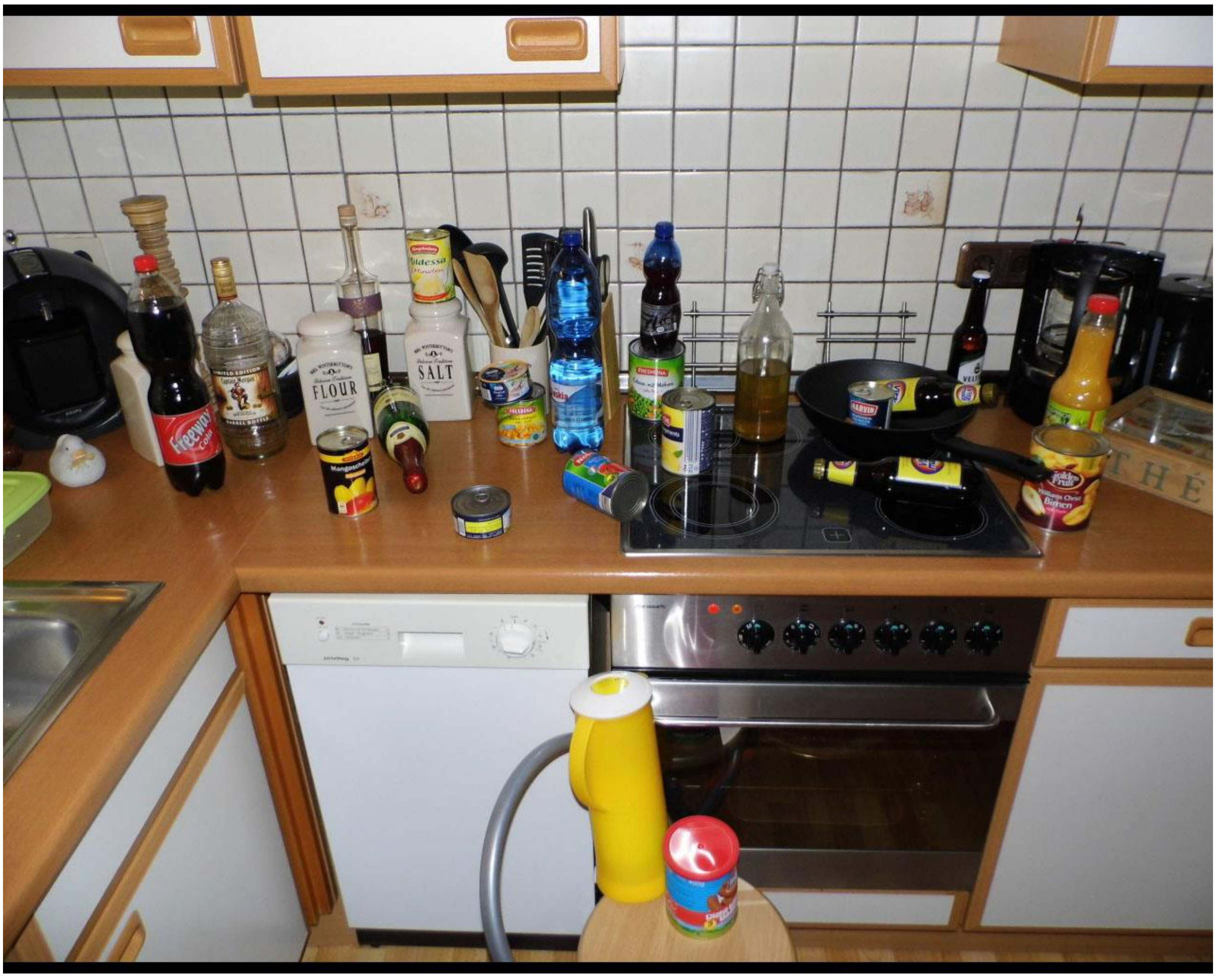
Scene 7 : Counting task of either combs or brushes

Scene 8 : Counting task of either sponges or dish towels

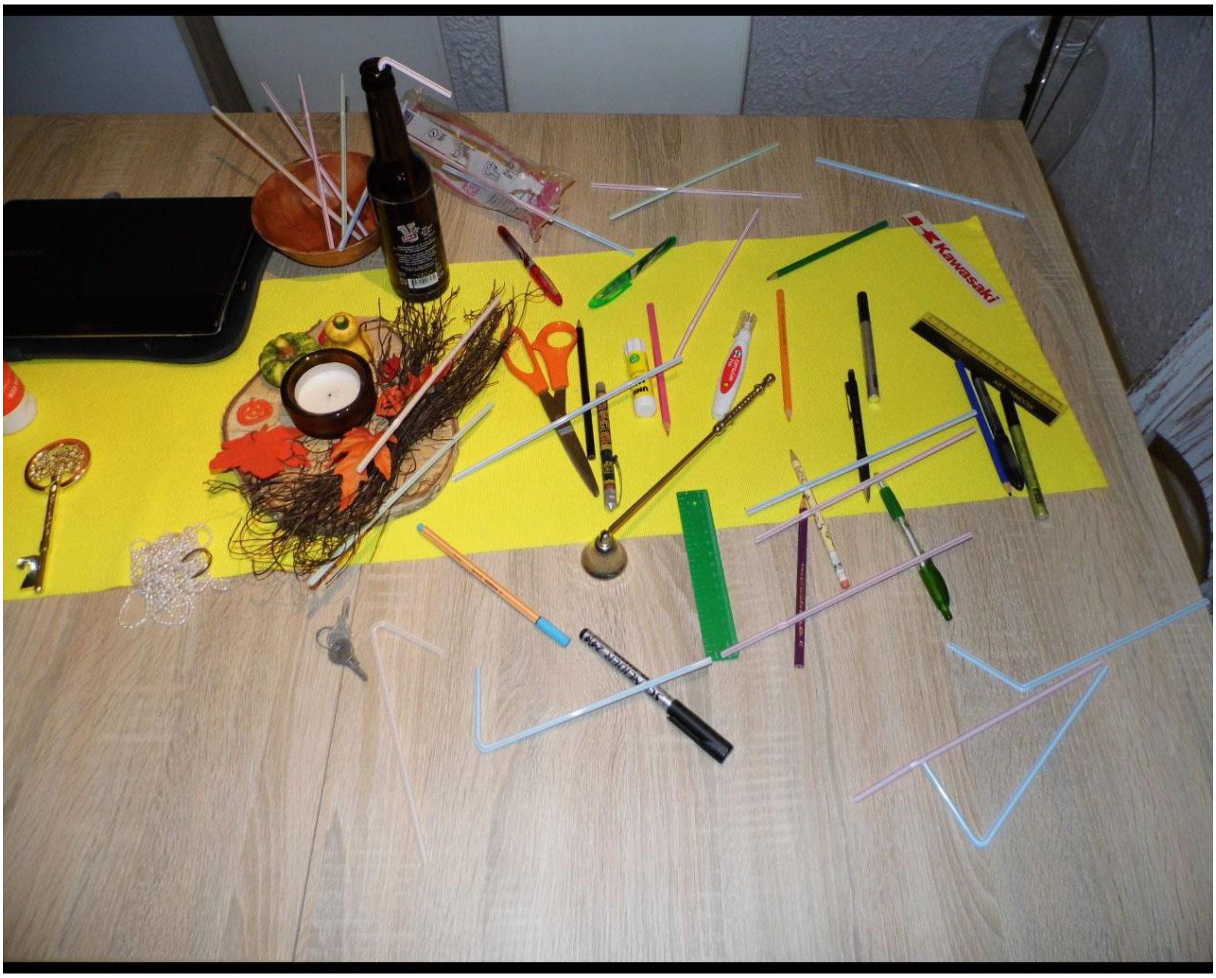
Scene 9 : Counting task of either cushions or towels

Scene 10 : Counting task of either pots or pans

Scene 11 : Counting task of either brooms or shovels



Scene 12 : Counting task of either bottles or cans



Scene 13 : Counting task of either pens or straws

Scene 14 : Counting tasks of either keys or bottle openers

Scene 15 : Counting task of either calculators or remotes

Scene 16 : Counting task of either socks or shoes

Appendix C

Algorithm MatLab Code

C.1 MultiLinear Regression with dispersion term

```
function [Reg,Repochs,ImpPeak] = multipleLinearRegressionwDispersionRegressorsV2(dataRaw,regressorDepth,...
    regressorMeanSize,pcaDims,srate,lCut,hCut,peakSpecificity,regressorSteps)

% Description :
% Performs a variation of the Multiple Linear Regression with dispersion term
% on a set of epochs to determine the properties of extrema in the Grand
% Average within each single Epoch. To determine the location and properties
% of each peak in an epoch properly, the algorithm takes into account the
% epoch variability contained within the given set. Most function arguments
% exists to specify how this variability should be accounted for.
%
% Input :
%
% - dataRaw          : [epochs x samples x channels] Set of Epochs that are
%                      to be analysed
% - regressorDepth   : Number of random permutation within the epochs to
%                      determine the regressors (1 <= x < Inf)
% - regressorMeanSize : Number of epochs that are averaged together for one
%                      variability example used to estimate the regressor
%                      (1 < x < nbEpochs in DataRaw)
% - pcaDims : Number of Principal Components used as regressors for one
%             peak if x >= 1. If x < 1, the value expresses the
%             percentage of variability to be contained within the amount
%             of dimensions
% - srate : Sampling rate of the dataRaw signal
% - lCut : Low cutoff Frequency of filter to smooth outcoming regressor
% - hCut : Highcut cutoff Frequency of filter to smooth regressor
% - peakSpecificity : threshold of possible offset the regressor peak can
%                      have to the closest real epoch peak before being
%                      marked as absent
% - regressorSteps : Number of Steps with which a random permutation of
%                      epochs is looe through (1 <= x < nbEpochs-regressorMeanSize)
%
% Output :
%
% - Reg : Contains information about the raw regressors that were used to
%         find the information in each peak
% - Repochs : contains information about each epoch : their regressed for,
%             the regression coefficients that were used, and the
%             properties of each marked peak in the real and the
```

```
%                 regressed signal
% - ImpPeak : Important peaks that were searched for. Their properties in
%               the ERP and the regressor used to represent them.
%

%% INFORMATION

warning('off','all');


%% Multiple linear regression

data = dataRaw;

%% Calculate mean ERPs
fprintf('Calculating mean ERPs of all Channels given the new Data \n\n');

[nbepochs, nbsamples, nbchannels] = size(data);
ImpPeak = cell(nbchannels,1);
meanERPs = squeeze(mean(data,1));

Repochs = cell(1);

for i = 1:nbchannels
    [~,MaxG] = findpeaks(meanERPs(:,i));
    [~,MinG] = findpeaks(-1 * meanERPs(:,i));
    ImpPeak{i}.generalPeakPositions = sort([MaxG;MinG]);
    ImpPeak{i}.generalPeakType = ismember(sort([MaxG;MinG]),MaxG);
    clear MaxG MinG
end

fprintf('General ERP Analysed \n');


%% Create regressors

% Create regressors for general ERP

fprintf('Creating general Regressor \n');
for i = 1:nbchannels
    fprintf('Doing Channel %i / %i \n',i,nbchannels);

    allRegressors = cell(nbchannels,1);

    steps = 1:regressorSteps:nbepochs;
        nbSteps = length(steps);
    ChannelVariabilityMatrix = zeros(nbSteps*regressorDepth,nbsamples);
    allPeakLocations = cell(nbepochs*regressorDepth,1);

    allRegressors{i}.variaBilityMat = cell(length(ImpPeak{i}.
    generalPeakPositions),1);

    % Create variability changes for this channel
    for v = 1:regressorDepth

        changes = randperm(nbepochs); % FIND BETTER WAY FOR RANDOMIZATION ?
     Variable amount of epochs in mean
        currentChannel = [data(changes,:,i);data(changes(1:
    regressorMeanSize),:,i)];

        for w = 1:nbSteps

            ChannelVariabilityMatrix(nbSteps * (v-1) + w,:) = ...
```

```
            mean(currentChannel(steps(w):(steps(w)+regressorMeanSize),:),1);
        currentMeanEpoch = ChannelVariabilityMatrix(nbSteps * (v-1) + w,:)';


        [~,MaxLocation] = findpeaks(currentMeanEpoch);
        [~,MinLocation] = findpeaks(-1*currentMeanEpoch);
        allPeakLocations{nbSteps * (v-1) + w} = sort([MaxLocation;MinLocation]);

        MaxIds = find(ismember(allPeakLocations{nbSteps * (v-1) + w},MaxLocation));
        MinIds = find(ismember(allPeakLocations{nbSteps * (v-1) + w},MinLocation));

        checker = ones(length([MaxLocation;MinLocation]),1);
        currentminId = 1;

        separatorPoints = find(diff(currentMeanEpoch,2) == 0);
        separatorPoints = sort([separatorPoints;-1;nbsamples-2])';
        separatorPoints = sort(unique([separatorPoints,(find(abs(diff(sign(diff(currentMeanEpoch,2))))==2))']))';
        separatorPoints = separatorPoints + 2;

        separatorPoints3 = allPeakLocations{nbSteps * (v-1) + w};
        separatorPoints3 = unique(sort([separatorPoints3;1;nbsamples])');

        for x = 1:length(ImpPeak{i}.generalPeakPositions) % Go through all potential existing peaks

            currentTime = ImpPeak{i}.generalPeakPositions(x);
            if(ImpPeak{i}.generalPeakType(x) == 1)
                currentArray = MaxLocation;
                currentIds = MaxIds;
            else
                currentArray = MinLocation;
                currentIds = MinIds;
            end

            [times,timeIds] = sort(abs(currentArray-currentTime),'ascend');
            selectionId = 1;

            if(isempty(timeIds))
                continue
            end

            while(timeIds(selectionId) < currentminId) % Check if element hasn't been chosen already
                selectionId = selectionId + 1;
                if(selectionId > length(timeIds))
                    selectionId = selectionId - 1;
                    break
                end
            end

            chosenId = currentIds(timeIds(selectionId));

            if(times(selectionId) <= peakSpecificity && checker(chosenId) ~= 0) % TODO : modular choice

```

```matlab
                    checker(chosenId) = 0;
                    temporary = find(checker == 0);
                    currentminId = temporary(end);

                    variabilityVector = zeros(nbsamples,1);
                    chosenExtremaTime = currentArray(timeIds(selectionId));

                    tmp1 = find(separatorPoints < chosenExtremaTime);
                    lowerSeparatorId = tmp1(end);
                    tmp2 = find(separatorPoints > chosenExtremaTime);
                    upperSeparatorId = tmp2(1);

                    tmp3 = find(separatorPoints3 == chosenExtremaTime);
                    lowerSeparatorId2 = tmp3 - 1;
                    upperSeparatorId2 = tmp3 + 1;

                    lSP = (separatorPoints(lowerSeparatorId) + separatorPoints3(lowerSeparatorId2))/2;
                    uSP = (separatorPoints(upperSeparatorId) + separatorPoints3(upperSeparatorId2))/2;

%                     variabilityVector(separatorPoints(lowerSeparatorId):separatorPoints(upperSeparatorId)) = ...
%                         currentMeanEpoch(separatorPoints(lowerSeparatorId):separatorPoints(upperSeparatorId));

                    variabilityVector(lSP:uSP) = ...
                        currentMeanEpoch(lSP:uSP);

                    allRegressors{i}.variaBilityMat{x} = [allRegressors{i}.variaBilityMat{x}, ...
                        variabilityVector];

                end

            end

        end
    end

    for x  = 1:length(ImpPeak{i}.generalPeakPositions)
        pcaVector = allRegressors{i}.variaBilityMat{x};
        if(size(pcaVector,2) >= floor(nbSteps*regressorDepth/100))
            [pcomps,~,latent] = pca(pcaVector);
            if(pcaDims >= 1)
                nbDims = pcaDims ;
            else
                if(pcaDims == 0)
                    pcaThres = 0.9999;
                else
                    pcaThres = pcaDims;
                end
                pcaDims = sum(cumsum(latent)/sum(latent)< pcaThres);
                nbDims = sum(cumsum(latent)/sum(latent)< pcaThres);
            end
            ImpPeak{i}.regressor{x} = pcaVector * pcomps(:,1:nbDims);
            Reg.channel{i}.peak{x}.regressor = pcaVector * pcomps(:,1:nbDims);
        end

        clear pcaVector

    end
```

```
    clear allRegressors;

end

%% Regress signal
fprintf('Regressing general and finding properties \n');
% Find peaks and extract their properties for each epoch
for j = 1:nbepochs
    fprintf('Doing Epoch %i / %i \n',j,nbepochs);
    for i = 1:nbchannels

        regressors = [];
        for k = 1:length(ImpPeak{i}.regressor)
            regressors = [regressors,ImpPeak{i}.regressor{k}];
            Reg.channel{i}.peak{k} = ImpPeak{i}.regressor{k};
        end
        currentEpoch = data(j,:,i);
        regressionCoefficients = regress(currentEpoch',regressors);
        regression = regressors * regressionCoefficients ;

        Repochs{j}.channel{i}.regression = regression;
        Repochs{j}.channel{i}.regCoeffs = regressionCoefficients;

        Reg.channel{i}.allRegressors = regressors;

        regressorCounter = 0;

        [~,MaxLocsReal] = findpeaks(currentEpoch);
        [~,MinLocsReal] = findpeaks(-1*currentEpoch);

        orderReal = sort([MaxLocsReal,MinLocsReal],'ascend');
        realType = (ismember(orderReal,MaxLocsReal));
        notTaken = ones(length(orderReal),1);

        for k = 1:length(ImpPeak{i}.regressor) % go through all peaks to
    analyse

            if(~isempty(ImpPeak{i}.regressor{k}))

                regressorCounter = regressorCounter + 1;

                localRawRegression = sum(regressors(:,(pcaDims*(
    regressorCounter-1) + 1):(pcaDims*regressorCounter))...
                    .* repmat(regressionCoefficients((pcaDims*(
    regressorCounter-1) + 1):(pcaDims*regressorCounter))',nbsamples,1),2);

                % Find part that is uninteresting

                nullPoints = (localRawRegression == 0);
                counter = 1;
                while(nullPoints(counter) == 1)
                    counter = counter + 1;
                    if(counter > nbsamples)
                        counter = 1;
                        break;
                    end
                end
                startPoint = counter;

                inversion = flipud(nullPoints);
                counter = 1;
                while(inversion(counter) == 1)
```

```
                counter = counter + 1;
                if(counter > nbsamples)
                    counter = 1;
                    break;
                end
            end
            endPoint = nbsamples + 1 - counter;

            temporaryMat = eegfilt_new(localRawRegression,lCut,hCut,srate);
            correctedLocalRegression = zeros(nbsamples,1);
            correctedLocalRegression(startPoint:endPoint) = ...
                temporaryMat(startPoint:endPoint);

            currentType = ImpPeak{i}.generalPeakType(k);
            Repochs{j}.channel{i}.peak{k}.type = currentType;
            if(currentType == 1) %is Maximum
                modifier = 1;
                truePeakLocs = MaxLocsReal;
                %truePeakVals = MaxValsReal;
                truePeakOrder = find(realType == 1);
            else %is Minimum
                modifier = -1;
                truePeakLocs = MinLocsReal;
                %truePeakVals = MinValsReal;
                truePeakOrder = find(realType == 0);
            end

            [~, regressionLocs] = findpeaks(correctedLocalRegression * modifier);

            % Removal of peaks outside of interesting region
            removalPeaks = (regressionLocs > startPoint & regressionLocs < endPoint);
            regressionTimes2 = correctedLocalRegression(regressionLocs(removalPeaks));
            regressionLocs2 = regressionLocs(removalPeaks);

            if(~isempty(regressionTimes2))

                % Complete regression properties
                if(currentType == 1)
                    [~,prominentLocation] = max(regressionTimes2);
                else
                    [~,prominentLocation] = min(regressionTimes2);
                end

                separatorPointsX = find(diff(correctedLocalRegression,2) == 0)';
                separatorPointsX = sort([separatorPointsX,find(abs(diff(sign(diff(correctedLocalRegression,2))))==2)']);
                separatorPointsX = separatorPointsX + 2;
                separatorPointsX = sort(unique([1,separatorPointsX,length(correctedLocalRegression)]));


                tmp1 = find(separatorPointsX < regressionLocs2(prominentLocation));
                lowerSeparatorIdX = separatorPointsX(tmp1(end));
                tmp2 = find(separatorPointsX > regressionLocs2(prominentLocation));
                upperSeparatorIdX = separatorPointsX(tmp2(1));

```

```
                Repochs{j}.channel{i}.peak{k}.localReg.amplitude = regressionTimes2(prominentLocation);
                Repochs{j}.channel{i}.peak{k}.localReg.latency = regressionLocs2(prominentLocation);
                Repochs{j}.channel{i}.peak{k}.localReg.isabsent = 0;
                Repochs{j}.channel{i}.peak{k}.localReg.morphology = upperSeparatorIdX - lowerSeparatorIdX;
                Repochs{j}.channel{i}.peak{k}.localReg.lmorph = regressionLocs2(prominentLocation) - lowerSeparatorIdX;
                Repochs{j}.channel{i}.peak{k}.localReg.rmorph = upperSeparatorIdX - regressionLocs2(prominentLocation);


                %Complete actual properties

                [comVals,closestComparison] = min(abs(truePeakLocs - regressionLocs2(prominentLocation)));

                if(notTaken(truePeakOrder(closestComparison)) == 1 & comVals <= 15) % is not already assigned

                    notTaken(truePeakOrder(closestComparison)) = 0;

                    separatorPointsY = find(diff(currentEpoch,2) == 0)';
                    %size(separatorPointsY)
                    %size(find(abs(diff(sign(diff(currentEpoch,2))))==2)')
                    separatorPointsY = (sort([separatorPointsY;find(abs(diff(sign(diff(currentEpoch,2))))==2)']))';
                    separatorPointsY = separatorPointsY + 2;
                    separatorPointsY = sort(unique([1,separatorPointsY,length(currentEpoch)]));

                    tmp1 = find(separatorPointsY < truePeakLocs(closestComparison));
                    lowerSeparatorIdY = separatorPointsY(tmp1(end));
                    tmp2 = find(separatorPointsY > truePeakLocs(closestComparison));
                    upperSeparatorIdY = separatorPointsY(tmp2(1));


                    Repochs{j}.channel{i}.peak{k}.actualEpoch.amplitude = currentEpoch(truePeakLocs(closestComparison));
                    Repochs{j}.channel{i}.peak{k}.actualEpoch.latency = truePeakLocs(closestComparison);
                    Repochs{j}.channel{i}.peak{k}.actualEpoch.isabsent = 0;
                    Repochs{j}.channel{i}.peak{k}.actualEpoch.morphology = upperSeparatorIdY - lowerSeparatorIdY;
                    Repochs{j}.channel{i}.peak{k}.localReg.lmorph = truePeakLocs(closestComparison) - lowerSeparatorIdY;
                    Repochs{j}.channel{i}.peak{k}.localReg.rmorph = upperSeparatorIdY - truePeakLocs(closestComparison);

                else

                    Repochs{j}.channel{i}.peak{k}.actualEpoch.amplitude = 0;
                    Repochs{j}.channel{i}.peak{k}.actualEpoch.latency = 0;
                    Repochs{j}.channel{i}.peak{k}.actualEpoch.isabsent = 1;
```

```
            Repochs{j}.channel{i}.peak{k}.actualEpoch.morphology = 0;
            Repochs{j}.channel{i}.peak{k}.actualEpoch.lmorph = 0;
            Repochs{j}.channel{i}.peak{k}.actualEpoch.rmorph = 0;

          end

        else

          Repochs{j}.channel{i}.peak{k}.localReg.amplitude = 0;
          Repochs{j}.channel{i}.peak{k}.localReg.latency = 0;
          Repochs{j}.channel{i}.peak{k}.localReg.isabsent = 1;
          Repochs{j}.channel{i}.peak{k}.localReg.morphology = 0;
          Repochs{j}.channel{i}.peak{k}.localReg.lmorph = 0;
          Repochs{j}.channel{i}.peak{k}.localReg.rmorph = 0;

          Repochs{j}.channel{i}.peak{k}.actualEpoch.amplitude = 0;
          Repochs{j}.channel{i}.peak{k}.actualEpoch.latency = 0;
          Repochs{j}.channel{i}.peak{k}.actualEpoch.isabsent = 1;
          Repochs{j}.channel{i}.peak{k}.actualEpoch.morphology = 0;
          Repochs{j}.channel{i}.peak{k}.actualEpoch.lmorph = 0;
          Repochs{j}.channel{i}.peak{k}.actualEpoch.rmorph = 0;

        end

      end

    end

  end

end

end
```

C.2 Fixation-based Component Resynchronization

```
function [adjustedFixationStarts, fullscore] = resynchroniseFRP(continousOzChannel,EEGtimestamp,fixationTimes,EEGSrate,ETSrate)

% Function designed to synchronise an continous EEG signal with a
% eye-tracker recording relying on Fixation related Potentials.
%
% Input variables :
% - contiousOzChannel [nbSamples x 1] : recorded samples at the Oz electrode
% - EEGtimestamp [nbSamples x 1] : timestamps (in ms) of recorded samples
% - fixationTimes [nbFixations x 1] : timestamps (in ms) of fixationstarts
% - srate [1 x 1] : Sampling rate of the EEG recording
%
% Output variables :
% - adjustedFixationStarts [nbFixations x 1] : timecorrected fixations to
%   fit with EEG data
%

  nbSamples = length(continousOzChannel);
  nbFixations = length(fixationTimes);

```

```
  [sortedfix,fixOrder] = sort(fixationTimes,'ascend');
  [~, originalOrder] = sort(fixOrder);
  addedEEG = zeros(nbSamples*2,1);
  offset = round(0.09 * ETSrate);

  baseOffset = EEGtimestamp(1);
  timeDifferences = cumsum([0;diff(sortedfix)]);

  for i = 1:nbFixations
    [~,currentTimePoint] = min(abs(EEGtimestamp-(baseOffset + timeDifferences(i))));
    startPoint = length(continousOzChannel) - currentTimePoint +1;
    endPoint = length(addedEEG) - currentTimePoint;
    addedEEG(startPoint:endPoint) = addedEEG(startPoint:endPoint) + continousOzChannel;
  end

  fullscore = addedEEG/nbFixations;

  addedEEG = addedEEG / nbFixations;
  [~,locationP1] = max(addedEEG(length(continousOzChannel)+1:end));
  actualTime = EEGtimestamp(locationP1);

  newEpochs = zeros(nbFixations,400);
  newTimes = zeros(nbFixations,1);
  for i = 1:nbFixations

    [~,midPoint] = min(abs(EEGtimestamp-(actualTime + timeDifferences(i))));
    newEpochs(i,:) = continousOzChannel(midPoint-200:midPoint+199);
    newTimes(i) = EEGtimestamp(midPoint);
  end

  meanCoefficient = 50;
  locationMap = zeros(nbFixations-meanCoefficient,1);
  meanTimes = zeros(nbFixations-meanCoefficient,1);
  for i = 1:(nbFixations-meanCoefficient)
    meanEpoch = mean(newEpochs(i:i+50,:));
    [~,location] = max(meanEpoch);
    locationMap(i) = location;
    meanTimes(i) = mean(newTimes(i:i+50));
  end
  x = meanTimes;

  y = round((locationMap-200)*(ETSrate/EEGSrate)) + meanTimes;

  reconstructionPolynom = polyfit(x,y,1);
  f = polyval(reconstructionPolynom,x);

  locaFix =  polyval(reconstructionPolynom,actualTime + timeDifferences);
  newFix = locaFix - round(offset);

  newFix = newFix(originalOrder);
  adjustedFixationStarts = newFix;

end
```

Bibliography

Acharya, Jayant N et al. (2016). “American clinical neurophysiology society guideline 2: guidelines for standard electrode position nomenclature”. In: *The Neurodiagnostic Journal* 56.4, pp. 245–252.

Acı, Çiğdem İnan, Murat Kaya, and Yuriy Mishchenko (2019). “Distinguishing mental attention states of humans via an EEG-based passive BCI using machine learning methods”. In: *Expert Systems with Applications* 134, pp. 153–166.

Ahlberg, Christopher and Ben Shneiderman (2003). “Visual information seeking: Tight coupling of dynamic query filters with starfield displays”. In: *The Craft of Information Visualization*. Elsevier, pp. 7–13.

Albert, Guillaume et al. (2005). “Scene perception, gaze behavior, and perceptual learning in virtual environments”. In: *CyberPsychology & Behavior* 8.6, pp. 592–600.

An, Kwang-Ok, Da-Hey Kim, and Jongbae Kim (2013). “Development of a Virtual Keyboard System Using a Bio-signal Interface and Preliminary Usability Test”. In: *International Conference on Human-Computer Interaction*. Springer, pp. 3–9.

Association, World Medical et al. (2001). “World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects.” In: *Bulletin of the World Health Organization* 79.4, p. 373.

Baccino, Thierry and Yves Manunta (2005). “Eye-fixation-related potentials: Insight into parafoveal processing”. In: *Journal of Psychophysiology* 19.3, pp. 204–215.

Beck, Melissa R, Maura C Lohrenz, and J Gregory Trafton (2010). “Measuring search efficiency in complex visual search tasks: Global and local clutter.” In: *Journal of experimental psychology: applied* 16.3, p. 238.

Berg, Patrick and Michael Scherg (1991). “Dipole models of eye movements and blinks”. In: *Electroencephalography and clinical Neurophysiology* 79.1, pp. 36–44.

Bishop, Christopher M (2006). *Pattern recognition and machine learning*. springer.

Bonnet, Michael H and Donna L Arand (2007). “EEG arousal norms by age”. In: *Journal of Clinical Sleep Medicine* 3.03, pp. 271–274.

Bos, Danny Oude et al. (2006). “EEG-based emotion recognition”. In: *The Influence of Visual and Auditory Stimuli* 56.3, pp. 1–17.

Bradley, Margaret M et al. (2008). “The pupil as a measure of emotional arousal and autonomic activation”. In: *Psychophysiology* 45.4, pp. 602–607.

Brouwer, Anne-Marie et al. (2013). “Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials”. In: *Journal of vision* 13.3, pp. 17–17.

Bulling, Andreas et al. (2010). “Eye movement analysis for activity recognition using electrooculography”. In: *IEEE transactions on pattern analysis and machine intelligence* 33.4, pp. 741–753.

Callaway, Enoch and Roy Halliday (1982). “The effect of attentional effort on visual evoked potential N1 latency”. In: *Psychiatry Research* 7.3, pp. 299–308.

Castelhano, Monica S, Michael L Mack, and John M Henderson (2009). “Viewing task influences eye movement control during active scene perception”. In: *Journal of vision* 9.3, pp. 6–6.

Cecotti, Hubert (2010). “A self-paced and calibration-less SSVEP-based brain–computer interface speller”. In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 18.2, pp. 127–133.

Chavarriaga, Ricardo and José del R Millán (2010). “Learning from EEG error-related potentials in noninvasive brain-computer interfaces”. In: *IEEE transactions on neural systems and rehabilitation engineering* 18.4, pp. 381–388.

Chen, Xiaogang et al. (2014). “A high-itr ssvep-based bci speller”. In: *Brain-Computer Interfaces* 1.3-4, pp. 181–191.

Chynał, Piotr et al. (2016). “Shopping behaviour analysis using eyetracking and EEG”. In: *2016 9th International Conference on Human System Interactions (HSI)*. IEEE, pp. 458–464.

Coco, Moreno I, Antje Nuthmann, and Olaf Dimigen (2020). “Fixation-related Brain Potentials during Semantic Integration of Object–Scene Information”. In: *Journal of Cognitive Neuroscience* 32.4, pp. 571–589. DOI: 10.1162/jocn_a_01504.

Cohen, Mark S and Susan Y Bookheimer (1994). “Localization of brain function using magnetic resonance imaging”. In: *Trends in neurosciences* 17.7, pp. 268–277.

Cooper, Alan et al. (2014). *About face: the essentials of interaction design*. John Wiley & Sons.

Coyle, Damien et al. (2011). “EEG-based continuous control of a game using a 3 channel motor imagery BCI: BCI game”. In: *2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*. IEEE, pp. 1–7.

Creel, Donnell Joseph (2019). “Visually evoked potentials”. In: *Handbook of clinical neurology*. Vol. 160. Elsevier, pp. 501–522.

Croft, R. J. and R. J. Barry (2000). “Removal of ocular artifact from the EEG: A review”. In: *Neurophysiologie Clinique* 30.1, pp. 5–19. ISSN: 09877053. DOI: 10.1016/S0987-7053(00)00055-1.

Crossland, Michael D, Gary S Rubin, et al. (2002). “The use of an infrared eyetracker to measure fixation stability”. In: *Optometry and vision science* 79.11, pp. 735–739.

Csikszentmihalyi, Mihaly (1997). *Finding flow: The psychology of engagement with everyday life.* Basic Books.

De Vos, Maarten et al. (2014). “P300 speller BCI with a mobile EEG system: comparison to a traditional amplifier”. In: *Journal of neural engineering* 11.3, p. 036008.

Delac, Kresimir and Mislav Grgic (2004). “A survey of biometric recognition methods”. In: *Proceedings. Elmar-2004. 46th International Symposium on Electronics in Marine*. IEEE, pp. 184–193.

Delorme, Arnaud and Scott Makeig (2004). “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis”. In: *Journal of neuroscience methods* 134.1, pp. 9–21.

Desmedt, John E and Claude Tomberg (1989). “Mapping early somatosensory evoked potentials in selective attention: critical evaluation of control conditions used for titrating by difference the cognitive P30, P40, P100 and N140”. In: *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section* 74.5, pp. 321–346.

Dimigen, Olaf et al. (2011). “Coregistration of eye movements and EEG in natural reading: analyses and review.” In: *Journal of Experimental Psychology: General* 140.4, p. 552.

Donchin, Emanuel and Michael GH Coles (1988). “Is the P300 component a manifestation of context updating?” In: *Behavioral and brain sciences* 11.3, pp. 357–374.

Downing, Ricard E, Joi L Moore, and Steven W Brown (2005). “The effects and interaction of spatial visualization and domain expertise on information seeking”. In: *Computers in Human Behavior* 21.2, pp. 195–209.

Drennan, Judy, Gillian Sullivan, and Josephine Previte (2006). “Privacy, risk perception, and expert online behavior: An exploratory study of household end users”. In: *Journal of Organizational and End User Computing (JOEUC)* 18.1, pp. 1–22.

Duchowski, Andrew T (2007). “Eye tracking methodology”. In: *Theory and practice* 328.614, pp. 2–3.

Dupont, Lien, Marc Antrop, and Veerle Van Eetvelde (2014). “Eye-tracking analysis in landscape perception research: Influence of photograph properties and landscape characteristics”. In: *Landscape Research* 39.4, pp. 417–432.

Earley, P Christopher, Pauline Wojnaroski, and William Prest (1987). “Task planning and energy expended: Exploration of how goals influence performance.” In: *Journal of applied psychology* 72.1, p. 107.

Ehinger, Benedikt V and Olaf Dimigen (2018). “Unfold: An integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis”. In: *BioRxiv*, p. 360156.

Enns, James T and Ronald A Rensink (1990). “Influence of scene-based properties on visual search”. In: *Science* 247.4943, pp. 721–723.

Ewing, Kate C, Stephen H Fairclough, and Kiel Gilleade (2016). “Evaluation of an adaptive game that uses EEG measures validated during the design process as inputs to a biocybernetic loop”. In: *Frontiers in human neuroscience* 10, p. 223.

Farwell, Lawrence Ashley and Emanuel Donchin (1988). “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials”. In: *Electroencephalography and clinical Neurophysiology* 70.6, pp. 510–523.

Faulkner, Laura (2003). “Beyond the five-user assumption: Benefits of increased sample sizes in usability testing”. In: *Behavior Research Methods, Instruments, & Computers* 35.3, pp. 379–383.

Ferrari, Marco and Valentina Quaresima (2012). “A brief review on the history of human functional near-infrared spectroscopy (fNIRS) development and fields of application”. In: *Neuroimage* 63.2, pp. 921–935.

Ferrari, Vincent, Andre Didierjean, and Evelyne Marmeche (2008). “Effect of expertise acquisition on strategic perception: The example of chess”. In: *The Quarterly Journal of Experimental Psychology* 61.8, pp. 1265–1280.

Finke, Andrea et al. (2016). “Toward FRP-based brain-machine interfaces-single-trial classification of fixation-related potentials”. In: *PLoS ONE*. ISSN: 19326203. DOI: 10.1371/journal.pone.0146848.

Frøkjær, Erik, Morten Hertzum, and Kasper Hornbæk (2000). “Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?” In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, pp. 345–352.

Gao, Dashan and Nuno Vasconcelos (2005). “Discriminant saliency for visual recognition from cluttered scenes”. In: *Advances in neural information processing systems*, pp. 481–488.

Goldberg, Joseph H and Xerxes P Kotval (1999). “Computer interface evaluation using eye movements: methods and constructs”. In: *International Journal of Industrial Ergonomics* 24.6, pp. 631–645.

Goldstein, Abraham, Kevin M Spencer, and Emanuel Donchin (2002). “The influence of stimulus deviance and novelty on the P300 and novelty P3”. In: *Psychophysiology* 39.6, pp. 781–790.

Gong, Chen et al. (2015). “Saliency Propagation From Simple to Difficult”. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Gosseries, Olivia et al. (2011). “Automated EEG entropy measurements in coma, vegetative state/unresponsive wakefulness syndrome and minimally conscious state”. In: *Functional neurology* 26.1, p. 25.

Graimann, Bernhard, Brendan Allison, and Gert Pfurtscheller (2010). "Brain–Computer Interfaces: A Gentle Introduction". In: *Brain-Computer Interfaces: Revolutionizing Human-Computer Interaction*. Ed. by Bernhard Graimann, Gert Pfurtscheller, and Brendan Allison. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 1–27. ISBN: 978-3-642-02091-9. DOI: 10.1007/978-3-642-02091-9_1. URL: https://doi.org/10.1007/978-3-642-02091-9_1.

Greene, Harold H (2006). "The control of fixation duration in visual search". In: *Perception* 35.3, pp. 303–315.

Guger, Christoph et al. (1999). "Prosthetic control by an EEG-based brain-computer interface (BCI)". In: *Proc. aaate 5th european conference for the advancement of assistive technology*. Citeseer, pp. 3–6.

Hackos, JoAnn T and Janice Redish (1998). "User and task analysis for interface design". In:

Hämäläinen, Matti et al. (1993). "Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain". In: *Reviews of modern Physics* 65.2, p. 413.

Hanslmayr, Simon et al. (2007). "Prestimulus oscillations predict visual perception performance between and within subjects". In: *Neuroimage* 37.4, pp. 1465–1473.

Harel, Jonathan, Christof Koch, and Pietro Perona (2007). "Graph-based visual saliency". In: *Advances in neural information processing systems*, pp. 545–552.

Hayhoe, Mary and Dana Ballard (2005). "Eye movements in natural behavior". In: 9.4, pp. 188–194. ISSN: 13646613. DOI: 10.1016/j.tics.2005.02.009.

Heger, Dominic, Felix Putze, and Tanja Schultz (2010). "Online workload recognition from EEG data during cognitive tests and human-machine interaction". In: *Annual Conference on Artificial Intelligence*. Springer, pp. 410–417.

Henderson, John M, Phillip A Weeks Jr, and Andrew Hollingworth (1999). "The effects of semantic consistency on eye movements during complex scene viewing." In: *Journal of experimental psychology: Human perception and performance* 25.1, p. 210.

Henderson, John M et al. (2007). "Visual saliency does not account for eye movements during visual search in real-world scenes". In: *Eye movements*. Elsevier, pp. 537–III.

Holcomb, Phillip J (1993). "Semantic priming and stimulus degradation: Implications for the role of the N400 in language processing". In: *Psychophysiology* 30.1, pp. 47–61.

Holmqvist, Kenneth et al. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.

Hooge, Ignace Th C, Casper J Erkelens, et al. (1998). "Adjustment of fixation duration in visual search". In: *Vision research* 38.9, pp. 1295–1302.

Hu, Li et al. (2011). "Taking into account latency, amplitude, and morphology: improved estimation of single-trial ERPs by wavelet filtering and multiple linear regression". In: *Journal of Neurophysiology* 106.6, pp. 3216–3229.

Ionescu, Radu et al. (2016). "How Hard Can It Be? Estimating the Difficulty of Visual Search in an Image". In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Irwin, David E (2004). "Fixation location and fixation duration as indices of cognitive processing". In: *The interface of language, vision, and action: Eye movements and the visual world* 217, pp. 105–133.

Itti, Laurent, Carl Gold, and Christof Koch (2001). "Visual attention and target detection in cluttered natural scenes". In: *Optical Engineering* 40.9, pp. 1784–1793.

Jakob, R (1998). "The use of eye movements in human-computer interaction techniques: what you look at is what you get". In: *Readings in intelligent user interfaces*, pp. 65–83.

Jeffreys, D Aled (1996). "Evoked potential studies of face and object processing". In: *Visual Cognition* 3.1, pp. 1–38.

Jung, DK et al. (2005). "Biosignal monitoring system for mobile telemedicine". In: *Proceedings of 7th International Workshop on Enterprise networking and Computing in Healthcare Industry, 2005. HEALTHCOM 2005.* IEEE, pp. 31–36.

Jung, Tzyy-Ping et al. (2000). "Removing electroencephalographic artifacts by blind source separation". In: *Psychophysiology* 37.2, pp. 163–178.

Just, Marcel Adam and Patricia A Carpenter (1984). "Using eye fixations to study reading comprehension". In: *New methods in reading comprehension research*, pp. 151–182.

Kahneman, Daniel (1973). *Attention and effort*. Vol. 1063. Citeseer.

Kamienkowski, Juan E et al. (2012). "Fixation-related potentials in visual search: a combined EEG and eye tracking study". In: *Journal of vision* 12.7, pp. 4–4.

Kaniusas, Eugenijus (2012). *Biomedical signals and sensors I: Linking physiological phenomena and biosignals*. Springer Science & Business Media.

Kaufman, Arie E, Amit Bandopadhay, and Bernard D Shaviv (1993). "An eye tracking computer user interface". In: *Proceedings of 1993 IEEE Research Properties in Virtual Reality Symposium*. IEEE, pp. 120–121.

Kaunitz, Lisandro N et al. (2014). "Looking for a face in the crowd: fixation-related potentials in an eye-movement visual search task". In: *NeuroImage* 89, pp. 297–305.

Keene, DL, S Whiting, and ECG Ventureyra (2000). "Electrocorticography". In: *Epileptic Disorders* 2.1, pp. 57–64.

Kirakowski, Jurek and Mary Corbett (1993). "SUMI: The software usability measurement inventory". In: *British journal of educational technology* 24.3, pp. 210–212.

Klimesch, Wolfgang et al. (1998). "Induced alpha band power changes in the human EEG and attention". In: *Neuroscience letters* 244.2, pp. 73–76.

Koelsch, Stefan, Erich Schroger, and Thomas C Gunter (2002). "Music matters: Preattentive musicality of the human brain". In: *Psychophysiology* 39.1, pp. 38–48.

Kothe, C (2014). "Lab streaming layer (LSL)". In: *https://github. com/sccn/labstreaminglayer. Accessed on October* 26, p. 2015.

Krystal, Andrew D and Jack D Edinger (2008). "Measuring sleep quality". In: *Sleep medicine* 9, S10–S17.

Kuikkaniemi, Kai et al. (2010). "The influence of implicit and explicit biofeedback in first-person shooter games". In: *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, pp. 859–868.

Kutas, Marta and Kara D Federmeier (2011). "Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP)". In: *Annual review of psychology* 62, pp. 621–647.

Lai, Meng-Lung et al. (2013). "A review of using eye-tracking technology in exploring learning from 2000 to 2012". In: *Educational research review* 10, pp. 90–115.

Lal, Thomas N et al. (2005). "Methods towards invasive human brain computer interfaces". In: *Advances in neural information processing systems*, pp. 737–744.

Land, Michael F and Mary Hayhoe (2001). "In what ways do eye movements contribute to everyday activities?" In: *Vision research* 41.25-26, pp. 3559–3565.

Laurel, Brenda and S Joy Mountford (1990). *The art of human-computer interface design*. Addison-Wesley Longman Publishing Co., Inc.

Lee, Brian B et al. (2011). *Telemetry-synchronized physiological monitoring and therapy delivery systems*. US Patent 7,912,537.

Lester, David (1968). "The effect of fear and anxiety on exploration and curiosity: Toward a theory of exploration". In: *The Journal of general psychology* 79.1, pp. 105–120.

Leuthardt, Eric C et al. (2006). "Electrocorticography-based brain computer interface-the Seattle experience". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 14.2, pp. 194–198.

Lew, Seok et al. (2013). "Effects of sutures and fontanels on MEG and EEG source analysis in a realistic infant head model". In: *Neuroimage* 76, pp. 282–293.
Lewis, James R (1995). "IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use". In: *International Journal of Human-Computer Interaction* 7.1, pp. 57–78.
Lotte, Fabien et al. (2007). "A review of classification algorithms for EEG-based brain–computer interfaces". In: *Journal of neural engineering* 4.2, R1.
Luck, Steven J (2014). *An introduction to the event-related potential technique*. MIT press.
Lüderitz, Berndt (1995). *History of the disorders of cardiac rhythm*. Futura Publishing Company Armonk.
Maaten, Laurens van der and Geoffrey Hinton (2008). "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9.86, pp. 2579–2605. DOI: 10.1.1.457.7213. URL: http://jmlr.org/papers/v9/vandermaaten08a.html.
Mandryk, Regan L et al. (2013). "Games as neurofeedback training for children with FASD". In: *Proceedings of the 12th International Conference on Interaction Design and Children*. ACM, pp. 165–172.
Mann, Derek TY et al. (2007). "Perceptual-cognitive expertise in sport: A meta-analysis". In: *Journal of Sport and Exercise Psychology* 29.4, pp. 457–478.
Meißner, Martin et al. (2017). "Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research". In: *Journal of Business Research*.
Mennes, Maarten et al. (2010). "Validation of ICA as a tool to remove eye movement artifacts from EEG/ERP". In: *Psychophysiology* 47.6, pp. 1142–1150. ISSN: 00485772. DOI: 10.1111/j.1469-8986.2010.01015.x.
Mey, Shirley, Dirk Koester, and Thomas Schack (2018a). "Braintracking! A first step to ideomotor simulation in manual action". In: *HexKop, Aachen*.
— (2018b). "I know what you want! How eye movements tell us our needs". In: *Neurobiology doctoral students workshop, Berlin*.
— (2019). "Ideomotor simulation in manual action, a combined study of EEG and eyetracking". In:
Millán, José del R (2003). "Adaptive brain interfaces". In: *Communications of the ACM* 46.ARTICLE, pp. 74–80.
Minsky, Marvin Lee (1967). *Computation*. Prentice-Hall Englewood Cliffs.
Moacdieh, Nadine M and Nadine B Sarter (2012). "Eye tracking metrics: A toolbox for assessing the effects of clutter on attention allocation". In: *Proceedings of the Human Factors and Ergonomics Society annual meeting*. Vol. 56. 1. SAGE Publications Sage CA: Los Angeles, CA, pp. 1366–1370.
Müller-Putz, Gernot et al. (2015). "Towards noninvasive hybrid brain–computer interfaces: framework, practice, clinical application, and beyond". In: *Proceedings of the IEEE* 103.6, pp. 926–943.
Muller-Putz, Gernot R and Gert Pfurtscheller (2007). "Control of an electrical prosthesis with an SSVEP-based BCI". In: *IEEE Transactions on Biomedical Engineering* 55.1, pp. 361–364.
Müller-Putz, Gernot R et al. (2005). "Steady-state visual evoked potential (SSVEP)-based communication: impact of harmonic frequency components". In: *Journal of neural engineering* 2.4, p. 123.
Muthukumaraswamy, Suresh (2013). "High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations". In: *Frontiers in human neuroscience* 7, p. 138.
Myers, Brad A (1998). "A brief history of human computer interaction technology". In: *interactions* 5.2, pp. 44–54.

Nagashima, Yoichi (2003). "Bio-sensing systems and bio-feedback systems for interactive media arts". In: *Proceedings of the 2003 conference on New interfaces for musical expression*. National University of Singapore, pp. 48–53.

Navarro, Agustin A et al. (2011). "Context-awareness as an enhancement of brain-computer interfaces". In: *International Workshop on Ambient Assisted Living*. Springer, pp. 216–223.

Nemrodov, Dan et al. (2018). "The neural dynamics of facial identity processing: Insights from EEG-based pattern analysis and image reconstruction". In: *Eneuro* 5.1.

Nielsen, Jakob (1993). "Iterative user-interface design". In: *Computer* 26.11, pp. 32–41.

Nielsen, Jakob and Kara Pernice (2010). *Eyetracking web usability*. New Riders.

Nijholt, Anton, Brendan Z Allison, and Rob JK Jacob (2011). "Brain-computer interaction: can multimodality help?" In: *Proceedings of the 13th international conference on multimodal interfaces*. ACM, pp. 35–40.

Norman, Don (2013). *The design of everyday things: Revised and expanded edition*. Basic books.

Norman, Donald A (1999). "Affordance, conventions, and design". In: *interactions* 6.3, pp. 38–43.

Oliveira, Inês, Rui Lopes, and Nuno M Guimarães (2009). "Development of a biosignals framework for usability analysis". In: *Proceedings of the 2009 ACM symposium on Applied Computing*. Citeseer, pp. 156–157.

Paul, Maman and Kanupriya Garg (2012). "The effect of heart rate variability biofeedback on performance psychology of basketball players". In: *Applied psychophysiology and biofeedback* 37.2, pp. 131–144.

Peach, Robert W (1997). *The ISO 9000 handbook*. McGraw-Hill Malaysia.

Pfurtscheller, Gert and FH Lopes Da Silva (1999). "Event-related EEG/MEG synchronization and desynchronization: basic principles". In: *Clinical neurophysiology* 110.11, pp. 1842–1857.

Pfurtscheller, Gert et al. (1997). "EEG-based discrimination between imagination of right and left hand movement". In: *Electroencephalography and clinical Neurophysiology* 103.6, pp. 642–651.

Pomplun, Marc and Sindhura Sunkara (2003). "Pupil dilation as an indicator of cognitive workload in human-computer interaction". In: *Proceedings of the International Conference on HCI*. Vol. 2003.

Pönkänen, Laura M et al. (2010). "Does it make a difference if I have an eye contact with you or with your picture? An ERP study". In: *Social cognitive and affective neuroscience* 6.4, pp. 486–494.

Preece, Jenny et al. (1994). *Human-computer interaction*. Addison-Wesley Longman Ltd.

Pretorius, Marco C, André P Calitz, and Darelle van Greunen (2005). "The added value of eye tracking in the usability evaluation of a network management tool". In: *Proceedings of the 2005 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*. South African Institute for Computer Scientists and Information Technologists, pp. 1–10.

Puce, Aina, Truett Allison, and Gregory McCarthy (1999). "Electrophysiological studies of human face perception. III: Effects of top-down processing on face-specific potentials". In: *Cerebral Cortex* 9.5, pp. 445–458.

Rao, Rajesh PN (2013). *Brain-computer interfacing: an introduction*. Cambridge University Press.

Riechmann, Hannes, Andrea Finke, and Helge Ritter (2015). "Using a cVEP-based Brain-Computer Interface to control a virtual agent". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 24.6, pp. 692–699.

Riseberg, Jocelyn et al. (1998). “Frustrating the user on purpose: Using biosignals in a pilot study to detect the user’s emotional state”. In: *CHI 98 conference summary on Human factors in computing systems*, pp. 227–228.

Robertson, Alan R (1990). “Historical development of CIE recommended color difference equations”. In: *Color Research & Application* 15.3, pp. 167–170.

Robinson, Amanda K et al. (2017). “Very high density EEG elucidates spatiotemporal aspects of early visual processing”. In: *Scientific reports* 7.1, p. 16248.

Rugg, Michael D and Michael GH Coles (1995). *Electrophysiology of mind: Event-related brain potentials and cognition.* Oxford University Press.

Samadani, Uzma et al. (2015). “Eye tracking detects disconjugate eye movements associated with structural traumatic brain injury and concussion”. In: *Journal of neurotrauma* 32.8, pp. 548–556.

Santorio, Santorio, James Keill, and John Quincy (1720). *Medicina statica:: being the aphorisms of Sanctorius, translated into English with large explanations.*. W. and J. Newton.

Scholtz, Jean (2004). “Usability evaluation”. In: *National Institute of Standards and Technology* 1.

Schweinberger, Stefan R, Eva-Maria Pfütze, and Werner Sommer (1995). “Repetition priming and associative priming of face recognition: Evidence from event-related potentials.” In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21.3, p. 722.

Semmlow, John L (2008). *Biosignal and medical image processing*. CRC press.

Silva, Hugo Plácido da, Ana Fred, and Raúl Martins (2014). “Biosignals for everyone”. In: *IEEE Pervasive Computing* 13.4, pp. 64–71.

Silva, Hugo Plácido da et al. (2014). “BIT: biosignal igniter toolkit”. In: *Computer methods and programs in biomedicine* 115.1, pp. 20–32.

Simons, Daniel J and Daniel T Levin (1997). “Change blindness”. In: *Trends in cognitive sciences* 1.7, pp. 261–267.

Singh, Sanjay P (2014). “Magnetoencephalography: basic principles”. In: *Annals of Indian Academy of Neurology* 17.Suppl 1, S107.

Sinha, Gaurav, Rahul Shahi, and Mani Shankar (2010). “Human computer interaction”. In: *2010 3rd International Conference on Emerging Trends in Engineering and Technology*. IEEE, pp. 1–4.

Smith, Barry (1988). “Foundations of Gestalt theory”. In:

Smith, SJM (2005). “EEG in the diagnosis, classification, and management of patients with epilepsy”. In: *Journal of Neurology, Neurosurgery & Psychiatry* 76.suppl 2, pp. ii2–ii7.

Špakov, Oleg and Darius Miniotas (2007). “Visualization of eye gaze data using heat maps”. In: *Elektronika ir elektrotechnika*, pp. 55–58.

Spüler, Martin and Christian Niethammer (2015). “Error-related potentials during continuous feedback: using EEG to detect errors of different type and severity”. In: *Frontiers in human neuroscience* 9, p. 155.

Stenberg, Georg (1992). “Personality and the EEG: Arousal and emotional arousability”. In: *Personality and individual differences* 13.10, pp. 1097–1113.

Stone, Debbie et al. (2005). *User interface design and evaluation*. Elsevier.

Szalma, James L et al. (2004). “Effects of sensory modality and task duration on performance, workload, and stress in sustained attention”. In: *Human factors* 46.2, pp. 219–233.

Takeda, Kotaro, Shuntaro Okazaki, and Junichi Ushiyama (Dec. 2015). “Special Issues No.3 : Measurement Technique for Ergonomics, Section 4 : Measurements and Analyses of Bioelectric Phenomena and Others (3): Noninvasive Techniques for Measuring Brain Activity: EEG, fMRI, and fNIRS”. In: *The Japanese journal of ergonomics* 51, pp. 411–419. DOI: 10.5100/jje.51.411.

Tan, Desney and Anton Nijholt (2010). "Brain-Computer Interfaces and Human-Computer Interaction". In: *Brain-Computer Interfaces: Applying our Minds to Human-Computer Interaction*. Ed. by Desney S. Tan and Anton Nijholt. London: Springer London, pp. 3–19. ISBN: 978-1-84996-272-8. DOI: 10.1007/978-1-84996-272-8_1. URL: https://doi.org/10.1007/978-1-84996-272-8_1.

Teplan, Michal et al. (2002). "Fundamentals of EEG measurement". In: *Measurement science review* 2.2, pp. 1–11. DOI: 10.1.1.534.9267. URL: http://www.edumed.org.br/cursos/neurociencia/MethodsEEGMeasurement.pdf.

Thie, Johnson (2013). *A wireless marker system to enable evoked potential recordings using a wireless EEG system (EPOC) and a portable computer*. Tech. rep. PeerJ PrePrints.

Tudor, Mario, Lorainne Tudor, and Katarina Ivana Tudor (2005). *Hans Berger (1873-1941)–the history of electroencephalography*.

Van Den Broek, Egon L et al. (2009). "Affective man-machine interface: Unveiling human emotions through biosignals". In: *International Joint Conference on Biomedical Engineering Systems and Technologies*. Springer, pp. 21–47.

Verleger, Rolf, Theo Gasser, and Joachim Möcks (1982). "Correction of EOG artifacts in event-related potentials of the EEG: Aspects of reliability and validity". In: *Psychophysiology* 19.4, pp. 472–480.

Walker, Anthony and Craig Barrack (2003). *Method of detecting drift between two clocks*. US Patent App. 10/103,299.

Wallstrom, Garrick L et al. (2004). "Automatic correction of ocular artifacts in the EEG: a comparison of regression-based and component-based methods". In: *International journal of psychophysiology* 53.2, pp. 105–119.

Wenzel, Markus A, Jan-Eike Golenia, and Benjamin Blankertz (2016). "Classification of eye fixation related potentials for variable stimulus saliency". In: *Frontiers in neuroscience* 10, p. 23.

Williams, John R (2008). "The Declaration of Helsinki and public health". In: *Bulletin of the World Health Organization* 86, pp. 650–652.

Winkler, Irene, Stefan Haufe, and Michael Tangermann (2011). "Automatic classification of artifactual ICA-components for artifact removal in EEG signals". In: *Behavioral and Brain Functions* 7.1, p. 30.

Winkler, Irene et al. (2015). "On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, pp. 4101–4105.

Winkler, István et al. (2005). "Preattentive binding of auditory and visual stimulus features". In: *Journal of Cognitive Neuroscience* 17.2, pp. 320–339.

Wintermark, Max et al. (2005). "Comparative overview of brain perfusion imaging techniques". In: *Stroke* 36.9, e83–e99.

Wobrock, Dennis et al. (2015). "Continuous mental effort evaluation during 3d object manipulation tasks based on brain and physiological signals". In: *IFIP Conference on Human-Computer Interaction*. Springer, pp. 472–487.

Wobrock, Dennis et al. (2016). "Assessing the Properties of Single-Trial Fixation-Related Potentials in a Complex Choice Task". In: *International Conference on Neural Information Processing*. Springer. Kyoto, Japan, pp. 555–563.

Wobrock, Dennis et al. (2017). "Fixation-Related Potentials as a Measure for Cognitive Demand in Visual Tasks on Single Trial Level". In: *European Conference on Eye Movements*.

Wobrock, Dennis et al. (2019). "Resynchronizing Recordings Using Fixation-Related Potentials". In: *2019 41th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE.

Wobrock, Dennis et al. (2020). “Using Fixation-Related Potentials for Inspecting Natural Interactions”. In: *Frontiers in human neuroscience* 14, p. 447.

Woldorff, Marty G (1993). “Distortion of ERP averages due to overlap from temporally adjacent ERPs: analysis and correction”. In: *Psychophysiology* 30.1, pp. 98–119.

Zander, Thorsten O and Christian Kothe (2011). “Towards passive brain–computer interfaces: applying brain–computer interface technology to human–machine systems in general”. In: *Journal of neural engineering* 8.2, p. 025005.