Kumulative Dissertation

zur Erlangung des Doktorgrades im

Fachbereich
Mathematik und Informatik

der

Universität Bremen

vorgelegt von

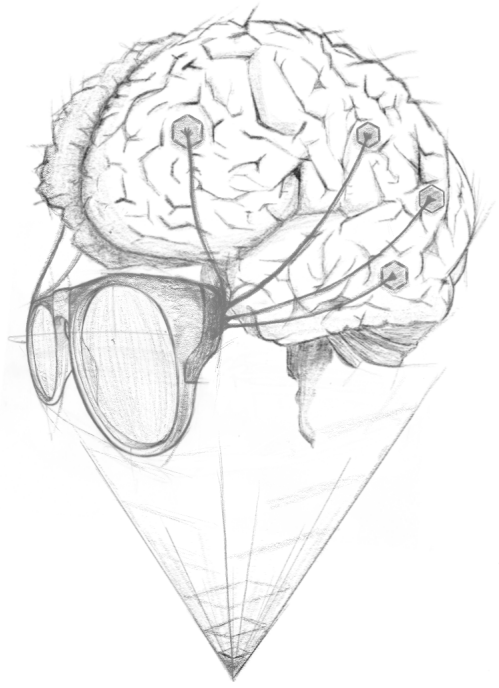
Lisa-Marie Vortmann, M.Sc.

Gutachter:innen	Prof. Dr.-Ing. Tanja Schultz
	Dr. Felix Putze
	Ass.-Prof. Dr. Mathias Benedek
Tag der Abgabe	12.04.2022
Tag des Kolloquiums	12.05.2022

Attention-Aware Interaction Systems for Augmented Reality

by

Lisa-Marie Vortmann



Acknowledgements

First and foremost, I want to express my gratitude to my supervisor, Dr. Felix Putze. I could count on his help and advice from the first to the last day of the past three and a half years - in research, in teaching, and in my own career. He has been an ideal mentor for me, with a healthy mix of criticism and encouragement, motivating me to think further and complete the projects in this dissertation.
Thank you, Felix, for the countless paper revisions, brainstorming sessions, and inspiration for my own teaching, as well as the numerous tips you provided along the way. Thank you for assisting me in navigating the world of science and finding my own path.

Another big thank you goes to Prof. Dr.-Ing. Tanja Schultz, who welcomed me into her team and gave me the opportunity to do this work. During our regular meetings, she always made a point of supporting and encouraging me not only in my research but also in my personal development.
Thank you, Tanja, for everything you have made possible in the last few years, and for everything I have learned during my time on your team.

During my time as a PhD student, I received so much kind help and support from my team members and everyone I had the pleasure of collaborating with.
Thank you for everything you taught me.

In addition, I am eternally grateful for the people who were closest to me during this time - my friends and family.
Especially for Magdi and Eylo who made the pandemic home office so wonderful and allowed me to convert the dining room table into a desk. Even during the most stressful times, they assisted me in maintaining my work-life balance and not losing sight of what was important.
Thank you both, for all of the conversations, and for listening, and showing that you cared. Always. (And thanks for cooking for me when I was sitting at my desk late again.)
I want to thank my parents, who have always supported me unconditionally and lifted every burden off my shoulders wherever they could.
Thank you, Mom and Dad, for always keeping your door open and allowing me to rely on you.

A special thanks also goes to all my friends who spontaneously participated in my experiments or proofread my texts: Svea, Enno, Gen. Your support means so much to me.
Thank you to everyone who helped carrying out the experiments or participated in them.

Last but not least, I would like to thank the two people who always helped me with technical and bureaucratic issues during my time at the Cognitive Systems Lab: Eric and Elke.
You made my work so much easier and were always willing to lend an ear.

Thank you very much for everything.

Zusammenfassung

Die Geschwindigkeit des technischen Fortschritts im 21. Jahrhundert ist kaum zu vergleichen mit allem zuvor Gewesenen. Eine Innovation folgt auf die andere, denn die verfügbare Rechenleistung und Speicherkapazität hat sich in den letzten Jahren stark verbessert. Doch bei all den Weiterentwicklungen darf ein Aspekt nicht aus den Augen verloren werden: Können wir Menschen mit diesen Technologien überhaupt effizient umgehen? Wie kann die Kommunikation zwischen Menschen und Maschinen zuverlässig funktionieren? Traditioneller Weise wurde eine solche Schnittstelle durch Knöpfe, eine Maus oder eine Tastatur geboten. Zuletzt wurden diese Interaktionsmöglichkeiten zunehmend durch Sprach- und Gestensteuerung ersetzt oder erweitert. Diese neuen Methoden sind notwendig in der Mensch-Computer Interaktion, um die Benutzbarkeit der Systeme zu gewährleisten und zu erhöhen.

Augmented Reality (AR) ist eine solche relativ neue Technologie. AR-Brillen erlauben es, virtuelle Inhalte als Hologramme in der echten Welt zu verankern und mit ihnen zu interagieren. Eine bewusste Interaktion mit den virtuellen Elementen kann über die eben erwähnten Interaktionsmöglichkeiten meist flüssig realisiert werden. Nichtsdestotrotz können die zusätzlichen visuellen Reize die mentalen Anforderungen erhöhen und - zur falschen Zeit oder am falschen Ort - sehr ablenkend für Benutzende sein. Ob dies der Fall ist, hängt unter anderem von den aktuellen Aufgaben und auch dem aktuellen Aufmerksamkeitszustand der Nutzenden ab.
Unsere Aufmerksamkeit ist eine Art Filter, mit der wir relevante und irrelevante Informationen voneinander unterscheiden, um unsere mentalen Kapazitäten optimal zu verteilen. Unser Aufmerksamkeitszustand ist sehr vielseitig und es können viele verschiedene Aspekte unterschieden werden.

In dieser Dissertation habe ich mich mit interaktiven Systemen auseinandergesetzt, die den Aufmerksamkeitszustandes des Benutzenden erkennen und Anpassungen am System vornehmen, um die Benutzbarkeit zu erhöhen (z.B. durch das Ausblenden von Informationen oder gezieltes Lenken der Aufmerksamkeit). Hierzu wurden vor allem Augmented Reality Szenarien untersucht und für verschiedene Paradigmen verwendet, da die Technologie hier einen besonders großen Nutzen verspricht.
Bei dem vorgeschlagenen System werden sensorische Benutzerdaten wie Gehirnaktivität oder Augenbewegungen gesammelt und mit Hilfe des maschinellen Lernens (z.B. künstliche neuronale Netze) zur Klassifikation verschiedener Aufmerksamkeitszustände verwendet. Das Hauptziel ist es dabei die multimodalen Daten optimal zu verwerten und darauf basierende effektive Veränderungen am System vorzunehmen. Hierzu wurden im Laufe der Dissertation mehrere Studien durchgeführt, die bisher in 15 Veröffentlichungen und Preprints festgehalten wurden. Diese Arbeiten konzentrieren sich jeweils auf verschiedene Aufmerksamkeitszustände, Paradigmen und Methoden des maschinellen Lernens, um eine möglichst ganzheitliche Sicht auf das hier entworfene Interaktionssystem für Augmented Reality zu ermöglichen. Dabei standen die generelle Unterscheidbarkeit der Zustände, die Zuverlässigkeit und Effektivität, sowie die Benutzbarkeit von End-to-End Systemen im Vordergrund.
Ich schlage in dieser Arbeit also vor, Systeme mit einer hohen Immersivität besser an die Nutzenden anzupassen, indem eine implizite Schnittstelle anhand von passiven, multimodalen Nutzerdaten erstellt wird.

Die Ergebnisse der in dieser Dissertation aufgeführten Studien zeigen, dass die dafür notwendige Bestimmung des Aufmerksamkeitszustandes zuverlässig möglich ist. Die größten Herausforderungen für ein solches System sind das Setup, die Echtzeitfähigkeit und das Sammeln personenbezogener Daten. In mehreren Arbeiten zur Verbesserung der vorgeschlagenen Klassifikationen konnte ich vielversprechende Ergebnisse erzielen und meine beispielhaften End-to-End Systeme bestätigten eine erhöhte Benutzbarkeit gegenüber aufmerksamkeits-unbewussten Anwendungen. Damit konnte ich einerseits die bisherige

Lücke in der Literatur in diesem Bereich schließen und andererseits Ansatzpunkte für neue Fragestellungen liefern.
Diese offenen Fragen konzentrieren sich for allem auf eine weitere Optimierung von Klassifikationsprozessen und System-Setup und stellen einen ethischen Umgang mit resultierenden Problemen und aufgezeichneten Daten in den Vordergrund.

Abstract

The speed of technological progress in the twenty-first century is unmatched in human history. As computing power and storage capacity have increased significantly in recent years, one innovation follows the other. But, despite all of these advances, one aspect must not be overlooked: Can we humans handle these technologies efficiently at all? How can communication between humans and machines work reliably? Buttons, a mouse, or a keyboard were traditionally used to provide such an interface. These interaction possibilities have recently been increasingly replaced or extended by voice and gesture control. Such new interfaces are required in human-computer interaction to ensure and improve system usability.

Another new technology is augmented reality (AR). AR glasses enable virtual content to be anchored in the real world as holograms and interacted with. The above-mentioned interaction possibilities allow for mostly fluent active interaction with virtual elements. Nonetheless, the additional visual stimuli can increase mental demands and be very distracting for users if they occur at the wrong time or in the wrong place. Whether this is the case is dependent, among other things, on the current tasks as well as the users' current attentional state.
In order to distribute our mental capacities optimally, we use our attention as a kind of filter, distinguishing between relevant and irrelevant information. Our attentional state is extremely versatile, with many distinct aspects.

In this dissertation, I investigated interactive systems that are aware of the user's attentional state and make system adjustments to improve usability (e.g., by hiding information or selectively directing attention). In particular, augmented reality scenarios were investigated and used for different paradigms in this project because AR can benefit a lot from the suggested technology.
Sensory data such as brain activity or eye movements are collected and machine learning (e.g., artificial neural networks) is used to classify different attentional states in the proposed attention-aware interaction system. The main goal here is to make the best use of the multimodal data and to make effective changes to the system based on them. Several studies were conducted during the course of the dissertation, and the results have been documented in 15 publications and preprints. These papers each focus on different attentional states, paradigms, and machine learning methods in order to provide the most comprehensive view of the augmented reality interaction system suggested here.
The main focus was on the general discriminability of the states, reliability and effectiveness, and usability of end-to-end systems. In essence, in this thesis, I propose an implicit interface based on passive, multimodal user data to adapt systems with high immersiveness to users.

The findings of the studies cited in this dissertation indicate that attentional states can be reliably detected. The main challenges for such a system are the setup, real-time capability, and personal data collection. Several works on improving the proposed classifications yielded promising results, and exemplary end-to-end systems confirmed increased usability over attention-unaware applications. Thus, on the one hand, I was able to fill a previous gap in the literature in this area, while also providing guidelines and inspiration for future work.
Future research in this area should prioritize ethical handling of resulting problems and recorded data, as well as further optimization of classification processes and system setup.

Contents

List of Figures

List of Tables

1 Introduction

The Golden Touch
A Tale from Ancient Greece

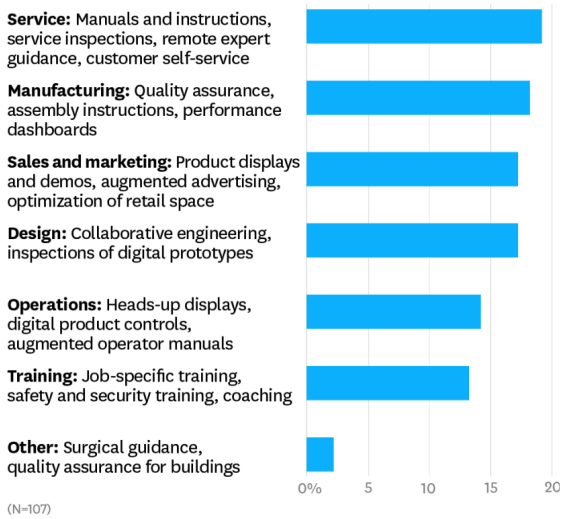
There was once a king named Midas who did a good deed for a Satyr and was granted a wish by the God of wine, Dionysus.
For his wish, Midas asked that whatever he touched would turn to gold. Although Dionysus tried to dissuade him, Midas insisted that the wish was an excellent one, and it was granted!
Excitedly, Midas went about touching all sorts of things, turning them into gold.

Soon Midas became hungry.
He picked up a piece of food, but he couldn't eat it, for it had turned to gold in his hand!
"I'll starve," moaned Midas, "Perhaps this was not such a good wish after all!"
Midas' beloved daughter, seeing his dismay, threw her arms about him to comfort him, and, she too turned to gold!
"The golden touch is no blessing," cried Midas.

He went to the river and wept.
The sand of that river turned as yellow as "fool's gold" for it is there, they say, that King Midas washed away the curse of the golden touch with his own tears.

- [175]

Occasionally, an idea sounds very appealing because of the numerous advantages it offers, such as an unlimited supply of gold. As a result, we are seduced by the desire to always have it at our fingertips, for instance, with a single touch. As the example of King Midas demonstrates, such a blessing can quickly devolve into a curse. In this instance, as a result of a lack of control. Indeed, everything he touches turns to gold, and he wishes his ability could be turned on and off automatically at certain times. Only the things he wishes to transform into gold should do so.
Similarly to the "golden touch", the ease with which information is now accessible via the internet suggests an analogous situation. It appears alluring to have an almost infinite amount of information about every thing and task. Better yet: the information is not only accessible, but also prominently displayed and intuitively integrated into our environment.
That, after all, is the purpose of head-mounted augmented reality devices. Available information is continuously embedded in our real world surroundings based on spatial locations or detected objects. However, when all objects trigger the display of new information as the gaze wanders, such augmented reality devices easily turn into prime examples of the Midas Touch Problem. This metaphor is also frequently employed in eye tracking-based user interfaces, because the fundamental function of the eye, which is to look and perceive visual information, must be distinguished from deliberate system interaction [51, 135, 160, 181].



(a) **Enterprise roles**. Percentage of surveyed developers creating AR experiences in each use category. Source: Harvard Business Review [76]

(b) **Strategic goals**. Percentage of respondents citing each reason as the primary goal of their AR development program. Source: Harvard Business Review [76]

Figure 1.1: Enterprise roles and strategic goals of AR development

1.1 Incentive to use Augmented Reality

In general, augmented reality (AR) refers to the display of virtual objects or bits of information into our physical environment (see Section 2.4). This can be accomplished through the use of a display and a camera (for example, a smartphone) or through the use of a head-mounted display (HMD). In contrast to virtual reality (VR), projections do not encompass all visual content, but rather merely a portion of it. Three-dimensional virtual objects or information boards are projected into our environment and anchored at strategic locations to appear as if they are a part of this environment. There is even one significant advantage of virtual objects over physical ones: they are much more adaptable and can be customized for each of us. Such holograms can be shown on a continuous display in front of our eyes, for instance disguised as pair of glasses. We are no longer required to operate our smartphone or computer in order for it to display the appropriate information. It is always available. When we travel, for example, we no longer need to use our translator to understand the menu; instead, we are shown the menu in the language we understand via a virtual overlay of the text. Or, when assembling a piece of furniture, we no longer have to toggle between the countless small parts and the instructions; instead, the parts to be used are virtually highlighted for us, and the steps to complete are visualized.
The technology of AR is gaining traction and getting more popular in a variety of industries, including gaming, healthcare, engineering, video entertainment, real estate, manufacturing, and the military. This is understandable given the enticing advantages and prospects described above.

1.1.1 Current Use

The Harvard Business Review performed a survey to find the main enterprise roles for AR technology developments. They identified service, manufacturing, sales and marketing, design, operations, and training as the main categories, as can be seen in Figure 1.1a.

Figure 1.2: Ikea Place App [86]

A well-known example of the application of smartphone-based augmented reality is Niantic's Pokémon GO, which was the main topic of conversation among gamers when it was released in 2016. While the gaming sector is probably more concerned with the improved enjoyment factor associated with immersiveness, the majority of other sectors are more concerned with enhancing the interface between humans and machines and the attendant increases in productivity and effectiveness. Through the intelligent use of AR, the efficient utilization of materials, personnel, and time may be increased. These objectives are also reflected in the Harvard Business Review poll (see Figure 1.1b). The aims were described as follows: Improved manufacturing quality, increased sales revenue, simplified development cycle, and improved monitoring capabilities.
For instance, BMW tested the usage of smart AR glasses for their workers. They displayed information in the field of view of the employees and barcode scans allowed them to interface with the warehouse management system. Over the course of an eight-hour shift, there was a 22% decrease in inventory identification time and a 33% reduction in errors [189]. Boeing adopted a head-mounted display for the assembly of wire harnesses for commercial airplanes, which resulted in a 25% reduction in manufacturing time and a near-zero mistake rate, all while boosting safety and uniformity of their standard operating procedures [179].

In other circumstances, however, the client or end user is the central incentive for using augmented reality. Enhanced end-user experience, product differentiation, and improved service are all highlighted as strategic aims in this section. A well-known example of this user-centered approach is the IKEA Place app, which enables customers to project furniture straight into their own four walls before making a purchase (see Figure 1.2).

1.1.2 Predicted Future Relevance

Augmented reality is a comparatively young but very promising technique and it is already on the rise in many industries. In an era when many individuals fear losing their jobs to intelligent machines, augmented reality technology enables the use of this intelligence to augment rather than replace human capabilities. As reported by the Harvard Business Review in 2017 [76], most industries were already planning to significantly increase their investment in augmented reality technology (see Figure 1.3).

IDG Research Services and PTC conducted a study of German-speaking decision-makers and executives from Germany, Austria and Switzerland on the use and investments into AR and VR [118]. They

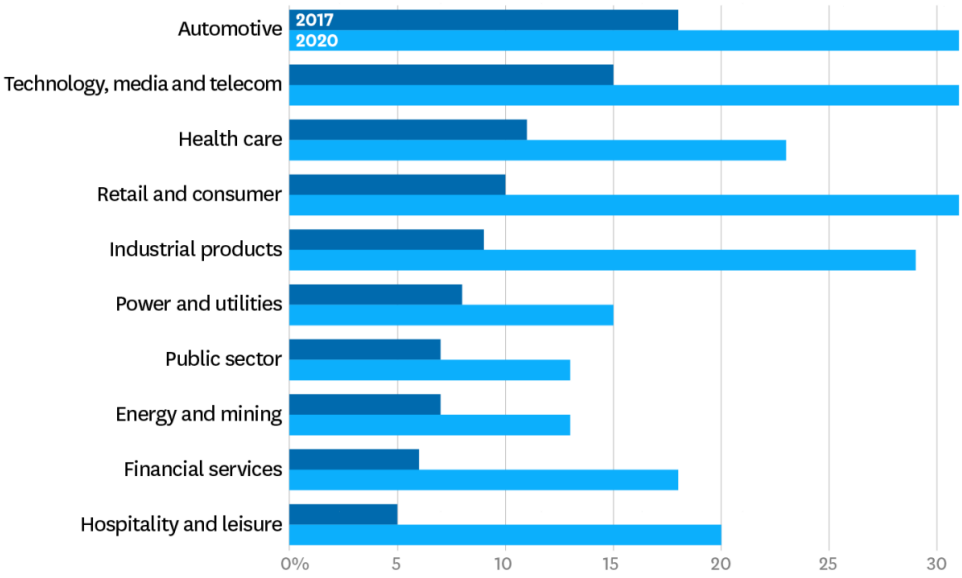


Figure 1.3: **Investments** Percentage of executives in each industry who say they were making substantial investments in AR in 2017 (year of the survey), and percentage anticipated for 2020. Data: PWC 2017 global digital IQ survey, taken by 2216 business and IT executives from 53 countries. Source: Harvard Business Review [76]

found that almost 75% of German businesses are utilizing or planning to utilize VR or AR. 77% of businesses reported that their VR/AR projects are meeting their objectives and over a third of businesses that use AR/VR technology have chosen remote assistance applications.
Pulse and CGS recently surveyed 100 enterprise-level IT leaders to ascertain the extent to which they have implemented AR into their operations. An overwhelming majority (86%) of respondents agreed with the statement "AR solutions will assist firms in better meeting the rising demands of their customers." Furthermore, they discovered that 73% of respondents have boosted their spending on augmented reality technology as a result of the pandemic. Approximately 76% of respondents agreed that AR assisted their organizations in recovering from the effects of COVID-19 [81].

More precisely, several big businesses specifically plan to incorporate the use of AR in their digitalization strategies, for instance DHL: In a trend research report it was predicted that drivers spend 40% to 60% of their days searching the suitable boxes for their next deliveries. There is an obvious need to reduce this wasted time. For instance, using augmented reality in logistics, crates might be overlaid with digital information highlighting the cargo scheduled for the next delivery [44].

These projections and outlooks for the industrial application of augmented reality are just one area that justifies ongoing research to advance this technology. Another noteworthy example demonstrating AR's relevance is Meta's (formerly Facebook) announcement of the Metaverse. The Metaverse is a hypothetical version of the internet that would allow for the creation and maintenance of persistent online 3-D virtual environments. Whatever form it takes, VR, AR, or just on a screen, the Metaverse has the potential to provide a larger overlap between our digital and physical lives in terms of wealth, socializing, productivity, shopping, and entertainment, among other things.

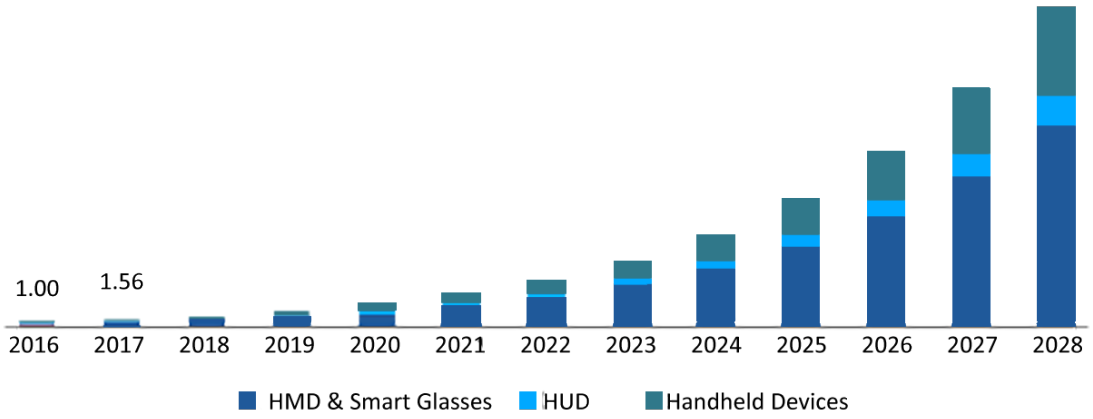


Figure 1.4: Anticipated United States AR market size for 3 different display types. From 2016 to 2028. In Billion US Dollar. Source: Grand View Research [67]

All the aforementioned projects and studies show that the technology of augmented reality is on the rise with an increasing interest and market value. The market size for the United States is even anticipated to increase exponentially until 2028 with the biggest market share for HMDs, as shown in Figure 1.4.

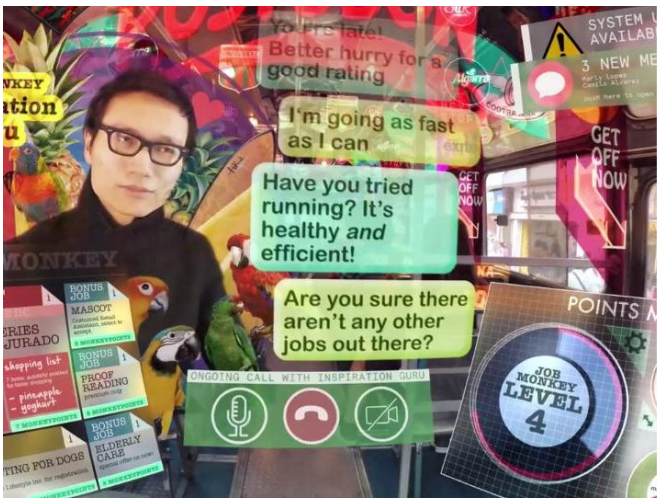


Figure 1.5: Hyper Reality by Matsuda [120]. Scan the QR Code to watch the video.

1.2 Problem Statement

AR comes with many exciting opportunities and could enhance human-machine interaction. However, just as King Midas suddenly became incapable of touching anything without it turning to gold, virtual visual content that does not truly support goal-oriented user behavior in the moment would quickly become a major issue. If virtual visual content is displayed whenever it is available, instead of when it is task-related and congruent to the user's current intentions and state, it can become overwhelming and defeats the purpose of easily accessible aid and information. Currently, with most of our technological gadgets (e.g., smartphones), we can interrupt the interaction immediately by putting them down or turning our head away from the screen. Unfortunately, using head-mounted displays or similar devices, this is not as simple.
But AR is here to stay. Due to its immersive, persuasive and ubiquitous nature, especially for HMDs, the technology requests for new ways of human-machine interaction that go beyond traditional interfaces. To avoid information overload and the resulting distraction and workload increase, attention-awareness should be added to the systems for a better user-adaptation.

1.2.1 Information Overload and the Need for Attention

If displayed objects and information are adapted solely to the environment and current task, rather than to the user's state and needs, there is a risk of sensory overload and an abundance of information that is nearly impossible to manage. Keiichi Matsuda published a short movie titled "Hyper Reality" [120] in May 2016 that illustrates what an excessive display of augmented reality content could look like (see Figure 1.5).

Even without the added visual information provided by augmented reality, we are constantly surrounded with a large number of sensory stimuli. We are conscious of some of them, such as our counterpart's perfume, the approaching ambulance's horn, or the small stone in our shoe. Others, on the other hand, we tend to ignore, such as our own odor, the ticking of the wall clock, or the feel of our clothing against our skin. We "choose" which perceptions enter our consciousness using our attention. Without an attentional filter, we would be unable to act purposefully. If we were aware of every sensation, every smell, and especially all visual information we would quickly get overwhelmed. Thus, attention is a nec-

essary mechanism that has evolved to perform task-orientedly. It is becoming increasingly recognized as a highly complex process that is inextricably linked to perception, memory, and action [34]. Apart from the fundamental concepts underlying attention, current research examines how attention interacts with other cognitive processes, how attention functions across multiple modalities, and what the neural basis of attention may be. Numerous other cognitive functions are influenced and modified by a variety of different attentional states and activities, and attention itself is multilayered (see Chapter 2.1).
The capacity of humans to use their attention purposefully has been optimized over time, both for survival and for effective problem solving. However, evolution has not prepared us for the additional virtual content that is presented to us when we use augmented reality, and as a result, we quickly become distracted by virtual stimuli. Such an increased demand on our attention can result in an increased workload, resulting in suboptimal performance and task processing delays. As discussed before, the desired effect of augmented reality over conventional technologies is an increased efficiency. But adapting to the increased sensory input is difficult for humans due to limited attentional capacities. Thus, it would be best if the AR system adapted to the user instead. This adaptation includes prioritizing and restraining information and content presentation in accordance with the user state and the task to aid goal-oriented stimulus selection and hence, help the user. In general, the usability of an augmented reality system would be significantly improved if the system took the user's attention state into account. This could not only avert an increase in workload, but also result in a significant decrease thereof.

1.2.2 Attention-Awareness for Usability Improvements

Interpersonal communication demonstrates that it is significantly improved when we can determine whether the person with whom we are conversing and interacting is focused on the subject at hand instead of thinking about something else. When we are aware that we are diverting the other person's attention away from something, we are more effective at interrupting communication. On the other hand, if our message is important, we may attempt to make ourselves more noticeable. This way, we can draw the attention to ourselves. Likewise, during a conversation, we hope to determine whether our counterpart is still following us or is overwhelmed by the information we provide. Humans communicate a great deal through facial expressions and even reaction times. The central message here is that some sensitivity to the state of attention enhances interaction so that, in a metaphorical sense, we are no longer speaking past one another. Our communication is better suited for the situation and counterpart. We adapt to each other.

In the same manner, human-machine interaction requires a mutual understanding. Obviously, when we use a system, it is helpful when we understand its state: whether it is operating normally, whether it has hung up, whether it is low on battery, whether it is responding to our commands, and so forth. On the other hand, it is just as important for the system to be informed about our state, all the more so if such system is ubiquitous and immersive. A system's ability to distinguish between states of distraction, focused thought, and also purposeful absorption of the information it provides us with is necessary for effective interaction.
In the best-case scenario, the machine not only detects these states of attention and adjusts its behavior appropriately in the moment, but also learns something about the user over time. Each individual is unique and responds to stimuli differently. While one person may perceive a stimulus as extremely intense, another may have a much higher threshold. Interpersonal communication is frequently easier when dealing with people we know well or have dealt with frequently, because we learn from previous reactions and encounters. Similarly, if a system is aware of the correlation between its performance and

a specific user's attention, that may improve the system's ability to adapt the user interfaces and interactions during long time use.

Exemplary Problem Scenario

Consider the following scenario: a manufacturing employee is tasked with the responsibility of assembling furniture. To assist and guide the employee, job-related information is displayed on their head-mounted augmented reality display. Certain parts of the process that are more mechanical in nature can be completed quickly and with little time or attention required. Others may involve technical or software enhancements, as well as the handling of highly sensitive items. In the latter case, a break in concentrated attention or an exogenously driven attention shift may result in errors or even damage, costing the factory time and money. If the augmented reality system was aware of the user's attentional state and attention shifts, it could adjust the salience and timeliness of the instruction displayed in augmented reality.
In an alternative situation, the employee may be interrupted by their employer or coworker and shift their attention from task-relevant attentiveness to task-irrelevant attentiveness. The attention-aware interaction system should detect that the worker is still alert but not focused on the current task and should pause its instructions.
Another possibility is that a worker loses focus on the primary task and enters a state of task-unrelated mind wandering. This period stalls their progress and costs the factory time and money once more. When a system detects prolonged periods of internal attention that are not required, it can alert the worker to the need to refocus on their task. Other possible applications include adaptive car driving assistants, learning and teaching, medical personnel training, augmented reality guidance and simultaneous attention monitoring of a doctor during complicated surgeries, or therapy sessions for patients with attention-related deficits.

In summary, I argue that a system's usability can be significantly improved if the system is both momentarily and long-term aware of the user's state. To achieve this level of awareness, I suggest to supply the system with a combination of several different modalities of user data, such as interaction data, eye gaze behavior or brain activity. In the proposed system, the multimodal data can then be used to classify the attentional state and this information can be used to enhance task-relevant content and reduce task-irrelevant content optimally (see Figure 1.6).

1.2.3 Suitable Human-Machine Interaction Techniques

Active communication is one of the most direct forms of interaction. This occurs between two people, for example, through spoken or written language and is comparable to explicit user input in human-machine interaction. The human-machine interface can be accomplished through the use of voice commands or, as still predominantely used at the moment, by pressing buttons and flipping switches. Consider now communicating the state of attention: Active communication is only possible if the user is consistently aware of their attentional state and the communication thereof does not interfere with the current task. The impracticality of this for the intended application is immediately apparent in a simple example: If we were aware and capable of communicating our own mind wandering, the system would probably no longer need to prevent us from doing so. In another situation, if the user is simply overwhelmed by the flood of information, they can pause the system consciously to process everything before moving on. While conscious input of the current user state may provide the system with useful additional information, it is not a holistic solution to the described problem.
A significant portion of human interaction is based on nonverbal signals such as facial expressions, reaction time, and even eye contact [101]. The communication partner can interpret this additional infor-

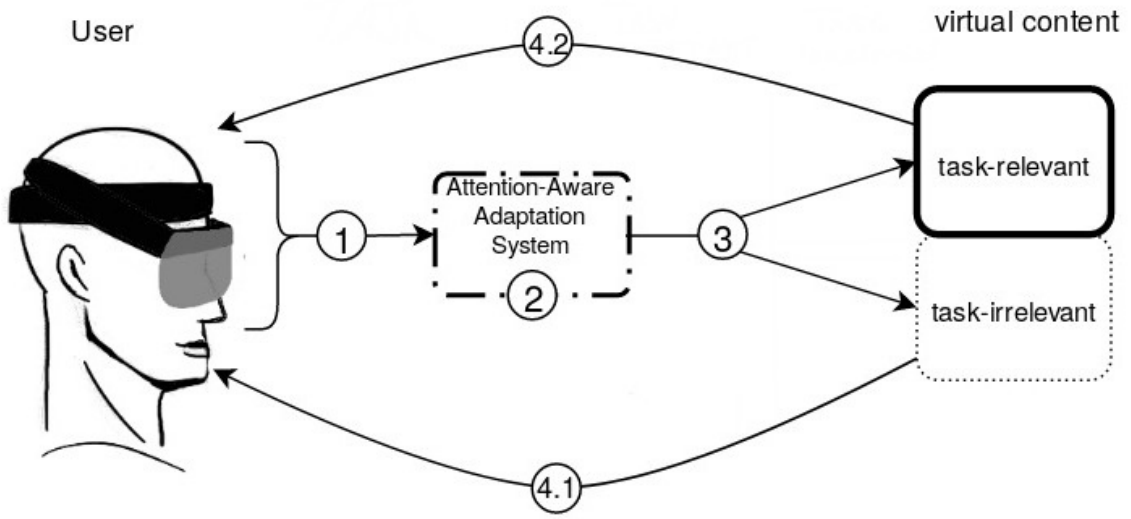


Figure 1.6: Proposed system overview: Adding attention-awareness for system adaptations to enhance task performance in AR. Multimodal user data can be used to classify the attentional state of the user in real time. (1) multimodal data recording (2) real-time attentional state classification (3) task-dependent adaptation (4) task performance improvement through (4.1) less distraction and (4.2) more focus.

mation and draw conclusions about the counterpart's state of attention. Similarly, machine learning can be used to classify the state of attention based on available user data. The data (or features extracted from it) are used as input for training a model that is then capable of predicting a user state in the classification problem described above.

The aforementioned eye movements are an excellent example of user data. Users' eye movements can be recorded using so-called eye trackers. For example, in smartphone-based augmented reality systems, the front-facing camera could be used directly to record eye movements. Additionally, with head-mounted displays, the integration of eye trackers is intuitive and feasible without the user requiring additional devices. High-quality HMD augmented reality devices, such as Microsoft's Hololens 2, include them as standard, because eye movements are also used as an explicit active interface in this case (e.g., to control the cursor). However, as will be discussed in Chapter 2.2, eye movements allow inferences about much more than the direction of visual attention. In this case, machines may even have an interactional advantage over humans. They may be able to extract information from eye movements that is not detected by humans during human-to-human communication and cannot be used to determine a person's state of attention.

On top of that, machines have the advantage of being able to detect and utilize biosignals that are not normally available for interpersonal communication, most notably brain activity. Numerous studies on the neural basis of attentional processes have been conducted to identify relevant brain regions and processes. There are several techniques for measuring brain activity, the most popular of which are functional magnetic resonance imaging (fMRI) and electroencephalography (EEG). For the application described here, it is critical for the system to have a high temporal resolution and mobility, which the EEG is better suited for than the fMRI (see Chapter 2.3).
The recorded data can be used to implement passive brain-computer interfaces (BCIs, see Chapter 2.6)

where the user does not have to provide explicit mental state feedback. Such a passive BCI, in addition to the eye tracking based classification, is an appropriate solution approach for designing a system with attention awareness.

In summary, our environment contains a vast amount of data in comparison to what is considered behaviorally significant. Attention is the cognitive process of subjectively or objectively selecting components of perceivable information in order to cope with information overload. I suggest that a reliable assessment of a user's attentional state would benefit both scientific research and consumer-oriented augmented reality products. I identified attention-awareness as critical for seamlessly integrating AR content into the environment and for adapting the user interface to the current user state in order to compensate for the additional visual information. This dissertation explores this novel idea and the components of the proposed system.

1.3 Objective

To realize the full potential of augmented reality applications, we must address the challenge of optimizing user interfaces (UIs) under aspects of mobility, constant context changes, and continued multitasking, as well as the blending of virtual and real content. Distracting content should be avoided, and the methodologies for determining distracting content should not be constrained by the aforementioned obstacles or the benefits of augmented reality.

The objective of this PhD thesis is to explore all the components necessary to develop an end-to-end mobile interaction system for augmented reality that is capable of adapting its behavior and the UI to the user's attentional state and is thus attention-aware.
On the assumption that the user's attentional state can be deduced from recorded biosignals, a system with knowledge of the current context and task would be able to adjust the augmented reality device's displayed content to optimally assist the user in their task while remaining as unobtrusive as possible.
Precisely, the two underlying hypotheses for this work are as follows:

H1 Several relevant aspects of the user's attentional state can be classified from available biosignal data.

H2 The attention-adaptation of the AR system will improve the usability and reduce the distraction.

The central tasks were to (1) determine which recording devices are most appropriate, including the identification of appropriate biosignals and recording device setups, as well as the capturing and processing of the biosignals; (2) construct experimental paradigms for different attentional states with reasonable applications for real-time attentional state assessment; (3) train, evaluate, and optimize classification algorithms for unimodal and multimodal biosignal data; (4) design task-dependent adaptation mechanisms and test their usability improvements by comparing attention-aware and attention-unaware system versions; and (5) integrate all components into an end-to-end closed-loop real-time system.

The intersection of augmented reality and brain-computer interfaces is a relevant issue at the moment. Although BCIs have been a primary focus of research for over two decades, the proposed research area of attention-awareness has received only little attention. Merging many information sources for such attention-driven augmented reality interfaces has not been researched systematically previous to this thesis.
The devices that display augmented reality have developed significantly in recent years, and when combined with the high processing capacity of today's computers, the objective of developing a real-time adaptive system for augmented reality has become much more attainable. The core idea of this thesis that links up the two main hypotheses is that the developed models classify the available multimodal biosignal data in order to produce an accurate prediction about a user's attentional state. This prediction will then be utilized to modify the present UI of an AR system, either by avoiding immersive information display or by emphasizing information that may be overlooked. Such a closed-loop system will regulate its behavior automatically without the need for manual input from the user.

1.3.1 The Proposed System

There are technical, design-related and algorithmic aspects to the proposed system. An attention-aware AR-Interface system would incorporate several subsolutions that respond to different aspects of attention. For this purpose, different biosignal recording devices and different output devices for the AR content can be used, as well as different machine learning algorithms and strategies to perform the classification. Importantly, context information about the augmented reality scene provides critical a priori

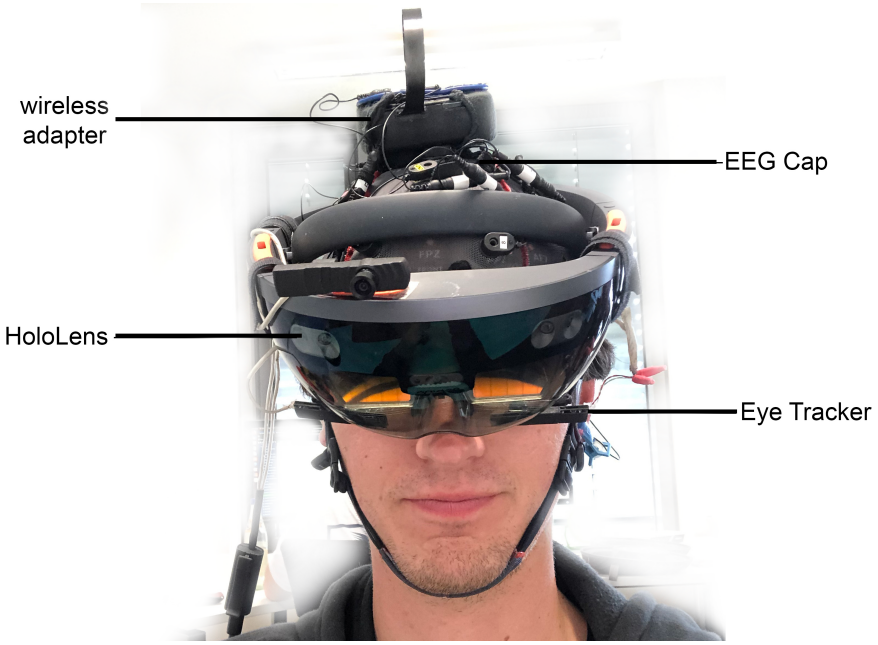


Figure 1.7: Exemplary Setup using a mobile EEG cap, an HMD AR Device (Microsoft's Hololens 1), and an additional binocular eye tracker.

knowledge about likely attention targets. Hence, the suggested system is highly application dependent and all performed studies rather explore a frame of possible system setups than one final fixed attention-aware interaction system. An exemplary setup that was used in several of the performed experiments can be seen in Figure 1.7.

System Components

The proposed systems consists of three main components: a recording component, a processing component and a component to display the AR content. If and in how far these components can be combined will be discussed in the course of this thesis. Currently, no device is capable of performing all necessary steps by itself and therefor the systems setup is usually a compartment of three or more components.

The recording components capture the biosignals and user data that the attentional state classification is based on. Eye tracking and EEG are two critical complementing approaches for automatically determining a person's focus of attention. The measurement of gaze behavior can be performed with eye-tracking devices or built-in cameras, facing the user. From the variety of brain signal measuring devices, the focus in this project will be on EEG recordings for the aforementioned reasons. Both research level EEG headsets and light weight consumer-grade EEG measuring devices were considered. Synchronized with behavioral data, the digitalized biosignal data will be transmitted to the processing component for the classification.
For the processing component, the main requirement is a high computational power. This is for instance fulfilled by desktop computers and high-end smartphones. Using the processing component, the biosignal data is transformed and classified. During this transformation process the digitalized data is taken as input and several steps are performed to output a prediction for the attentional state, including the filtering, optional feature extraction, and multimodal data fusion. The signal transformation is necessary

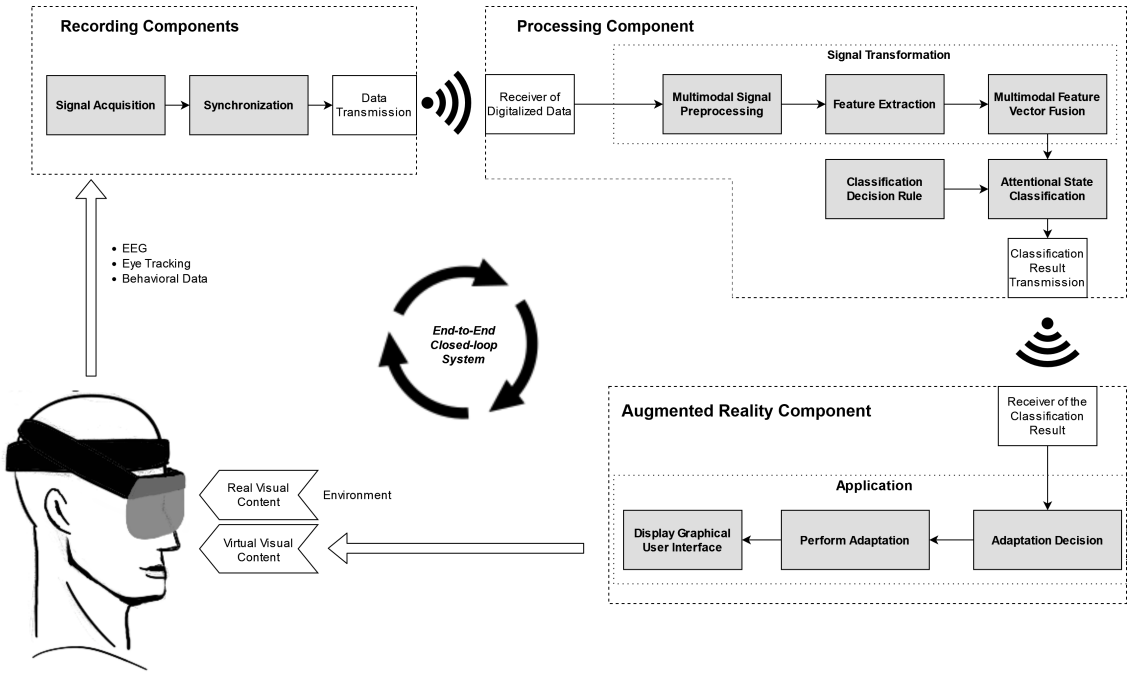


Figure 1.8: The suggested end-to-end closed-loop system with the relevant components on separate devices for recording, processing, and AR content display.

to clean the biosignal data from noise and provide optimal input features for the classifier. Depending on the decision rules and application purposes, the results of the classification process are transmitted to the AR component.
The AR component receives the classification result and displays an adapted UI based on predefined adaptation rules. The focus of this doctoral thesis is the optimization of the user interface for HMDs, also referred to as smart glasses. This sort of AR device is especially prone to distracting the user and adding workload, due to the constant visibility in the visual field and missing flexibility in removing the device spontaneously and for short times. However, at the moment many AR applications are still smartphone based and thus, they were also explored. Figure 1.8 illustrates the proposed end-to-end closed-loop system.

Classification Techniques

Machine learning algorithms and their application offer several options for adjustments. After the biosignal modality choice and optimal preprocessing, the focus is on finding optimal feature extraction algorithms and feature selection methods. In case of multimodal recordings, the feature fusion approach needs to be chosen. That choice is also highly dependent on the employed classification algorithm. For this thesis, several algorithms were tested and compared and the main focus was on Linear Discriminant Analysis (LDA) and Neural Networks.
Once basic discriminability of the attentional states is possible, more effort needs to be put into improving the reliability and efficiency of the classification and setup. It is desirable to apply classifiers that need little to no training in advance and that are able to perform the classification as fast as possible but accurately. One solution for this is training person-independent classifiers that work sufficiently well on all participants with fixed features and the same decision boundaries for everyone. This option

was explored in detail in the course of this work but reduces the classification accuracy compared to person-dependently trained models because of interpersonal differences. Another solution to reducing the training time to a minimum are online learning approaches where the classifier is personalized during the active usage of the system. Finally, the prediction of the attentional state can be consciously biased towards a specific prediction, depending on the use case.

Paradigms and Adaptation

The final usability improvement is reached by applying appropriate changes to the AR application in the closed-loop system, thus, without explicit user input to alter the systems state. How the obtained classification results are included and handled in the user interface highly depends on the paradigm and the classified attentional state. For this PhD thesis, the focus was on paradigms including steady-state visually-evoked potential stimuli (that are very common for BCI setups) and other paradigms that represent typical AR applications.
The alterations and adaptations can contain changes of the salience of virtually displayed objects, making them more or less distracting. Going one step further, the classification result could even cause the new presentation or deletion of certain objects. Moreover, the classification result could affect the timing of interface changes, or result in different placements of virtual objects in the real surroundings.
By performing these paradigm-specific adaptations based on the biosignal data classification result and not a manual user input to alter the systems state, attention-awareness of the system can be achieved without the need for attention-awareness of the user. Hence, the information about the attentional state and applied attentional mechanisms can be updated frequently and simultaneously and are then used change the systems behavior to optimally support the user in a task-oriented behavior with minimal distractions.

2 Background and Related Work

"Knowledge is of no value unless you put it into practice"

-Anton Chekhov

Numerous technologies and aspects must be combined in order to conduct research on attention-aware interaction systems for augmented reality. This is already evident from the problem's description and the proposed system. Such a cognitive system exists at the intersection of computer science, neuroscience, and psychology. To approach this work methodically and with a sound foundation, it is necessary to have a thorough understanding of the methods used. Additionally, there is a great deal to be learned from previously published work.

The topics addressed in this thesis are very diverse. While augmented reality and brain-computer interfaces are relatively new and technology-heavy scientific fields, the study of attention has captivated psychologists for centuries.
As described before, the aim is to build multimodal BCIs that classify attentional states and adapt the behavior and UI of AR systems. Clearly, one of the presumptions that needs to be fulfilled for an attention-aware interactive system based on multiple modalities is a measurable difference in the neurophysiological correlates of attentional states, such that a classifier is able to differentiate user states. There is a number of studies which deal with neural correlates of attention in the EEG signal or distinguishable eye-tracking features that allow for the derivation of the hypotheses that EEG and eye tracking features have high predictive power for the envisioned classification task. The combination of EEG features and eye-tracking features has been investigated and results show that both feature sets complement each other well [28, 29, 58, 97, 151, 163].

All of these topics, as well as the most significant machine learning techniques used in this work, will be discussed in the following chapter.
The specific objective of designing attention-aware interaction systems for AR has not been addressed in this manner in scientific works prior to this dissertation, so it is a novel and innovative concept, but numerous related works exist in all of the fields mentioned previously. To benefit from the findings of these previous or concurrent works and to incorporate them into the context of this thesis, they will be included in the explanation of the key concepts.

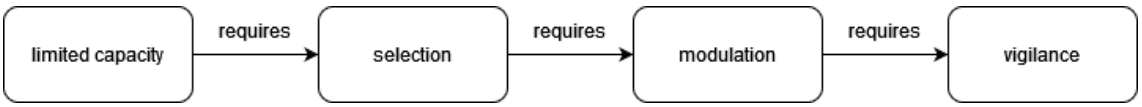


Figure 2.1: The basic functions and characteristics of attention, according to Chun et al. [34]

2.1 Attention

Frequently, in our day to day routine, we perform actions almost "blindly" and will later have no recollection of the process. For instance, parking in the driveway we realize we do not remember how we got there. This happens because we were not concentrating on the task of driving. While we can usually recall our current focus of attention, there are times when our mind wanders, we are unable to remain attentive, or something diverts our attention.

2.1.1 Definition

For centuries, the general concept of attention has been studied. Essentially, the sensory input at any given time is too large for us to process effectively and efficiently interact with our environment. Our brain is forced to be selective as a result of this limited processing capacity. For instance, in 1890, psychologist William James defined attention in his famous "the Principles of Psychology" [91] as follows:

> *"Attention [...] is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought, localization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others, and is a condition which has a real opposite in the confused, dazed, scatter brained state which in French is called distraction, and Zerstreutheit in German."*

While this is a frequently cited definition, it is not universal. The ambiguous framework described by "attention" lacks agreement on the majority of fundamental terms and concepts. Shiffrin and Czerwinski [166] defined it in a broader sense, focusing on the processing capacity limit:

> *"Attention has been used to refer to all those aspects of human cognition that the subject can control (...) and to all aspects of cognition having to do with limited resources or capacity, and methods of dealing with such constraints."*

These implicitly defined concepts share a common premise: the amount of information available in our environment vastly exceeds what is considered behaviorally relevant. Attention is the cognitive process of selecting aspects of perceivable information (subjectively or objectively). Even minor modifications to the definition carry underlying implications and assumptions, most notably about the role and significance of consciousness, concentration, willingness, resource allocation, memory, and vigilance. Chun et al. [34] summarized the fundamental functions and characteristics of attention as the limited capacity for perceptual information processing necessitates a selection and focus on environmental input and internal options [43, 147]. Following selection, attentional mechanisms must immediately bias information processing and modulate the effect and response optimally [168]. Finally, one must be able to maintain this level of focus for the duration of the required time [22]. These functions are summarized in Figure 2.1. It should be noted here, that attention has slightly different definitions and can be used synonymously for arousal, alertness, or vigilance depending on the research field. For instance, the field of machine learning has adapted the term for limited resource allocation mechanisms [113].

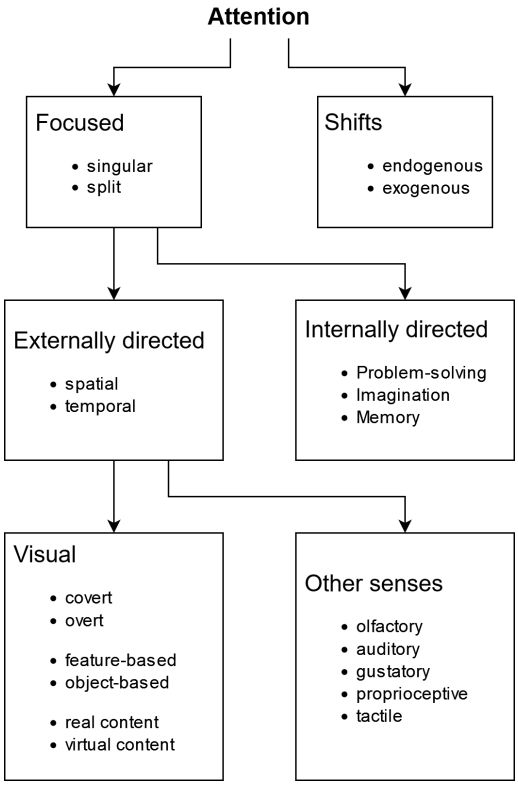


Figure 2.2: Taxonomy of attentional states

2.1.2 Attentional States

Attention is not a binary process: it has multiple dimensions, layers, and aspects to consider in addition to "attentive" and "not attentive" [145, 147]. This complexity is for example reflected in the range of attention that a person can maintain - from "paying no attention" to "being aware of an object or situation" to "completely and exclusively focusing on it." However, in addition to the pure level of attention, the attentional state taxonomy can be described by the direction of attention on tasks, senses, and objects (see Figure 2.2). The aspects important for this thesis will be explained in more detail in the following.
A target of attention can be either external (produced in the world and recorded by our senses) or internal (directed solely within ourselves) [92].
When attention is directed externally, it can be focused on a single object or divided between several objects, as well as be selective about the type of sensory input received - visual, gustatory, olfactory, tactile, or auditory.
Internally directed attention can also be evaluated further: Attention can be directed toward planning, imagination, memory, or any other internal task.

Visual Attention

Visual attention is more easily accessed and quantified than, for example, auditory or tactile attention because it is highly correlated with gaze. The allocation of information-processing resources in space in the context of vision can be classified as "overt" or "covert."

Overt spatial attention is the change in the focus of the eyes that is visible and trackable. When we scan our environment with our eyes, we concentrate the narrow, high-resolution foveal visual field on our center of attention in order to maximize detail sensitivity. This focal point can be an object, a feature, or simply a particular space in our environment. We can either consciously direct the view to a particular point of interest or it will follow our attention unconsciously. The degree to which these "bottom-up" and "top-down" processes cooperate and/or compete is one of numerous unanswered questions. The tracking of eye movements as a representation of visual attention processes is a very straightforward analysis framework with a high practical value. Additionally, in this context, brain imaging techniques are primarily used to analyze the primary visual cortex, downstream visual areas, and neural pathways for vision processing.

Covert spatial attention is a mental shift in focus that occurs without visible eye movements. As a result, it cannot be detected through eye movements and must be inferred via behavioral or neurophysiological changes. Experiments examining covert spatial attention require a careful design that maintains a stable fixation while altering the peripheral field of view to simulate the missing gaze correction to the center of attention. Posner [149] was able to bring this spatial attention under sufficient control to demonstrate that covert orienting governs attention to specific objects and thus can affect the output of perceptual processes. Some theories hypothesize that covert visual spatial attention guides overt attention. The pre-motor theory of attention, in particular, postulates that the same brain circuits that govern saccades also control covert spatial attention [161].

Rather than attenting to a spatial location, humans can also attend to a specific visual feature such as a particular hue or a certain shape. Sometimes, this can involve attending to an entire feature dimension. In contrast to spatial attention, feature-based attention is spatially global but spatial and feature attention appear to have cumulative effects [74].
Object-based attention is a closely connected concept to feature-based attention. In this case, attention is not directed toward an abstract feature in advance of a visual stimulus, but rather a specific object in the visual scene [33]. Objects are sorted out from their backgrounds pre-attentively across the visual field during the pass of activity through the visual hierarchy, which requires recurrent and serial processing to recognize distinct items in cluttered visual situations [106]. Serial processing entails redistributing scarce attentional resources within an image by alternating either covert or overt spatial attention [30].

In summary, gaze behavior is a result of an interaction between covert (detecting objects in the peripheral vision) and overt attention (eye-movement to focus on the object of interest).

Attention Shifts

When attention is relocated from one target to another, it can be classified according to how it was shifted: an endogenous shift of attention is a "desired" shift that occurs top-down, whereas an exogenous shift of attention is caused by sensory input (typically with a high saliency, diverting attention away from previous targets) and is thus classified as a bottom-up process [149].

2.2 Eye Tracking

Eye tracking is a type of sensor technology that detects and tracks a person's gaze in real time. The system translates oculomotor movements to a data stream that can include blinks, pupil position and size, gaze vectors for each eye, and gaze point information. Essentially, the technology decodes and converts eye movements into insights that may be used in a variety of applications or as a secondary input modality. Eye trackers are utilized in visual system research, psychology, psycholinguistics, marketing, as a human-computer interface input device, and in product design. The most often used variation extracts the eye position from video images. Other techniques rely on search coils or the electrooculogram (EOG) [162].

In the nineteenth century, investigations of eye movement were conducted through direct observation. Later, Edmund Huey invented the first eye tracker, utilizing a type of contact lens with a pupil hole and an attached aluminum pointer that moved in sync with the eye's movement. Guy Thomas Buswell invented the first non-intrusive eye trackers in the early twentieth century, employing light beams that were reflected off the eye and then recorded on film. In the 1970s, eye-tracking research, particularly in reading, grew quickly. Just and Carpenter established the popular "strong eye-mind theory" in 1980, stating that there is no discernible lag between what is in the center of focus and what is processed [7]. If this hypothesis is right, then when a person looks at a word or object, they also processes cognitively for the duration of the fixation. However, current research suggests that there are several limitations to this hypothesis [7].

2.2.1 Oculomotor Movements

Our visual field is constrained in size. If humans were unable to move their eyes, they would be unable to process information outside of foveal vision. We cannot capture all visibly available information at once. When a scene contains several items, only a subset of those things can appear in our field of view at any given time. The light that reaches the eye is focused by the cornea, and the iris controls the amount of light that enters through the pupil by changing its size. The size of the pupil is not only affected by brightness of the surroundings but has been shown to be mental state dependent as well [65, 79]. The lens that is located behind the pupil accommodates to the incoming light and further focuses it before it hits the retina at the back of the eye. This is the light sensitive area where the information is translated to electrical signals that are then transferred (mainly) to the visual cortex via the optic nerve.
The fovea is the area of the retina with the best resolution (see Figure 2.3). Due to the retina's acuity limitations, the human ability to differentiate fine detail declines significantly outside the fovea, in the parafovea (about 5 degrees outside the fixation center), and in the periphery [156]. Our eyes must continually be redirected and, in some ways, refocused in order to comprehend different stimuli.

Saccades

When human eyes are examining their immediate surroundings or reading, they make saccadic motions and halt multiple times, moving very swiftly between each stop.
A saccade is a rapid, simultaneous movement of both eyes in the same direction between fixations. The rate of movement with each saccade is uncontrollable; the eyes move as quickly as they can. By repositioning the eye in such a way that minor details of a scene can be felt more precisely, body resources can be utilised more efficiently. Saccades are one of the quickest movements the human body can generate with a peak angular velocity of up to 700 degrees/s for large saccades. For unexpected input, the onset of a saccade is delayed by about 200 ms and last between 20-200 ms, depending on their amplitude [63]. A saccade's amplitude is the angular distance traveled by the eye during the movement. Saccades are generated by a neural process that avoids lengthy pathways and directly activates the eye muscles [59].

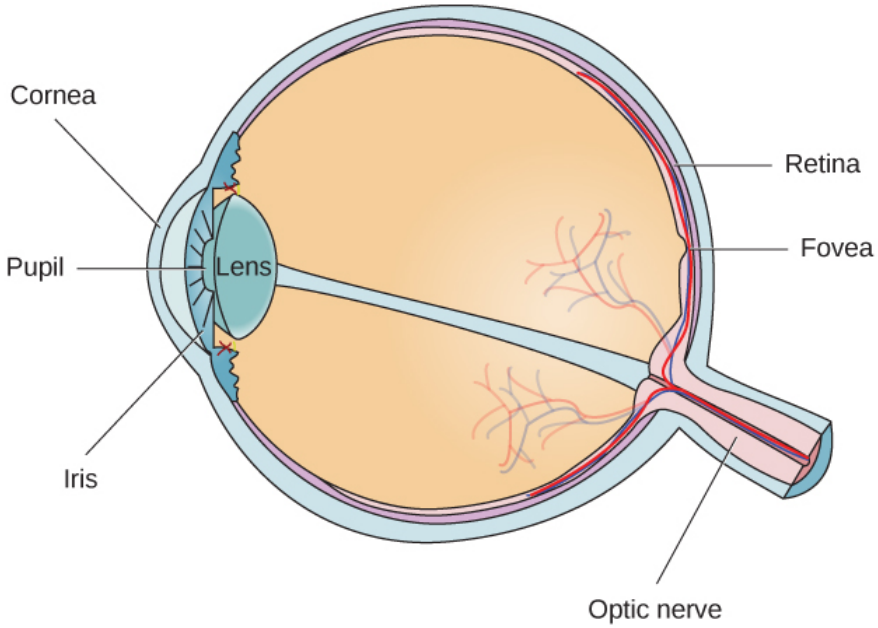


Figure 2.3: Relevant anatomy of the human eye from *Vision* [183]

To detect saccades in eye tracking data, velocity-based algorithms are popular but acceleration-based methods tend to be more exact [15].
When participants must perform many saccades in rapid succession, they often avoid returning to previously visited sites, causing delays in response. This is referred to as inhibition of return. As a result, the visual system is prompted to examine areas outside the image regions that were initially deemed important. This implies that the mechanism that generates saccades must have memory, which is thought to be achieved by temporarily suppressing the representation of recently visited sites [87].

Fixations

Visual fixations refer to the act of focusing the visual gaze on a single point. It can refer to either the location in time and space of the fixation or to the act of fixating. Fixation, or the act of fixating, is the interval between two saccades during which the eyes remain relatively immobile and almost all visual input occurs. A fixation lasts approximately 250 ms but there is a very high variance. It can range from a few milliseconds, which is too brief for conscious control, to several seconds [64].
Without retinal jitter, sensations tend to fade away swiftly. Thus, fixational eye movements are necessary.

Fixational Eye Movements

To sustain sight, the neurological system employs a technique known as fixational eye movement, which excites neurons in the brain's early visual areas in response to brief inputs. Microsaccades, ocular drifts, and ocular microtremors are three types of fixational eye movements. While the existence of these movements has been known since the 1950s, their functions are still studied.

Microsaccades are saccades that occur involuntarily during fixations. They are the most extensive and rapid fixational eye movements and a far more frequent subject of eye tracking than ocular drifts,

and ocular microtremor. Microsaccades, like saccades, are often binocular and consist of congruent amplitudes and directions in both eyes. In the 1960s, experts proposed that the maximum amplitude for microsaccades be 12 arcminutes to differentiate them from saccades [38]. But it has since been shown that microsaccades can easily exceed this number [177]. Thus, microsaccades and saccades cannot be distinguished by amplitude but only by their temporal relation to the fixation. While regular saccades are produced during active eye exploration, microsaccades are produced exclusively during fixations.
When the eye is fixed on an item, **ocular drift** is a smoother, slower, wandering motion of the eye [4].
Ocular microtremors are brief, synchronized oscillations of the eyes that occur at frequencies ranging from 40–100 Hz, however they normally occur at around 90 Hz in a healthy individual [144].

Vergence

Fixations require binocularly synchronized eye movements to fully grasp the object horizontally and vertically, but also in depth. Vergence is the concurrent movement of both eyes in different directions in order to achieve or sustain focused vision.
Convergence is performed when looking at a close item, the eyes rotate toward one another, whereas divergence refers to the eyes rotating away from one another when looking at distant objects. Cross-eyed viewing is a term that refers to excessive convergence. When the eyes diverge into the distance, they become parallel, essentially fixating the same spot [180].
Shifting the focus of the eyes to look at objects at different distances automatically results in simultaneous vergence and accommodation, a phenomenon referred to as the accommodation-convergence reflex [173]. In comparison to saccade movements, vergence movements occur at a much slower rate of roughly 50–200 degrees/s [53].

Enright [52] demonstrated that an upward or vertical saccade is often associated with eye divergence, whereas a downward saccade is associated with eye convergence. When an upward saccade is performed, the eyes diverge to align with the most likely uncrossed disparity in that region of the visual field. When executing a downward saccade, on the other hand, the eyes converge to enable alignment with crossing disparity in that region of the field. The phenomena might be viewed as quick binocular eye movements adapting to the statistics of the three-dimensional environment in order to decrease the need for corrective vergence movements at the end of saccades.

Blinks

Blinking is a physiological activity: the fast closure (100–400 ms) of the eyelids in a semi-autonomous manner. It is a necessary function of the eye that aids in the dispersal of tears and the removal of irritants from the cornea's surface but results in a loss of visual information. The presence or absence of eye blinks indicates a person's level of attentiveness and assists us in disengaging our attention. Following the commencement of the blink, brain activity weakens in the dorsal network and increases in the default-mode network, which is involved with internal processing [138].
Additionally, research indicates that we blink when we do not anticipate missing out on critical information. However, we are not continually aware of the optimal time to blink [8]. Blink rate can be influenced by a variety of factors, including weariness, eye damage, medicine, and disease.

2.2.2 Technology

A very frequent form of eye tracking that was also used for this thesis is based on optical tracking of the eye. An optical eye tracking system typically consists of one or more cameras, certain light sources (i.e.

infra red), and processing capabilities. With the use of machine learning and powerful image processing, algorithms convert the camera stream to data points. Typically, the corneal reflection and the pupil center are tracked over time. The vector formed by the two points can be used to determine the point of view on the surface or the direction of look. Before using the eye tracker, it is normally necessary to perform a simple calibration procedure on the individual. For more sensitive eye trackers, features from the front of the cornea and the back of the lens are tracked or even features from within the eye, such as the retinal blood vessels, are captured and tracked as the eye rotates [72].
Optical techniques, particularly those based on video recording, are frequently utilized for gaze tracking and are preferred due to their non-invasive nature and low cost.

Eye-tracking setups are as varied as the eye tracking technologies. Some are head-mounted, others require the head to be stable (for example, using a chin rest), and others operate remotely, following the head's movement automatically.
The majority operate at a sampling rate of 30–60 Hz and video-based eye trackers can operate at up to 1250 Hz. This frequency is required to accurately capture the fixational eye movements and saccade dynamics described before.

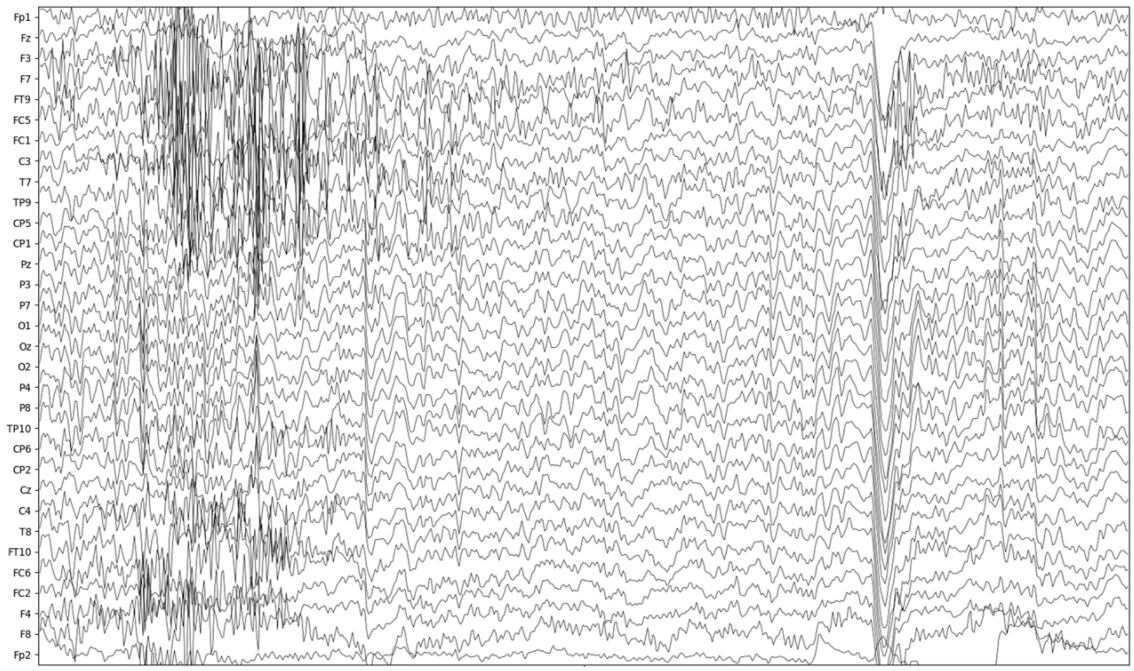


Figure 2.4: Example EEG data of 32 channels (according to the 10-20 system); Time on the x-axis.

2.3 Electroencephalography

As a measure of brain activity, Electroencephalography (EEG) was chosen for the user data collection in this dissertation. It is a non-invasive technique with a high temporal accuracy, with possible mobile setups. Communication between neurons takes place through the transmission of neurotransmitters or ions and the excitement of action potentials. Ions are electrically charged atoms that maintain the balance of positive and negative charges. Thus, this firing of neurons in the brain results in an electrical impulse and current change [39]. In 1875, Richard Caton made the first recordings of electrical brain activity in monkeys and rabbits by placing electrodes on the brain. In 1924, Hans Berger was able to measure human brain activity using electrodes on the scalp and named the invented device "electroencephalogram" [70]. The electrical signals in the brain are generated by millions of electrically charged neurons that can cause an ion wave. When an ion wave collides with an electrode, the wave can either transfer or withdraw electrons from the electrode's metal. The difference between the transfer and withdrawal of electrons can be expressed as an electrical voltage using a voltmeter. EEG is the term used to describe the process of measuring these voltages over time [104].

2.3.1 Technology and Devices

EEG electrodes are typically placed in standardized positions on a cap, such as in the 10-20 system, to conduct the measurement. The elicited electrical signals range between -100 and 100μV if measured at the scalp. Figure 2.4 shows a typical EEG measurement with the time on the x-axis and the voltage on the y-axis for each electrode.
The Fourier transform can be used to convert EEG data to the frequency domain and the resulting power spectra represent supposedly meaningful information about the current neural activity. This can be analyzed for different brain regions and different time points (see Figure 2.5). Different frequency bands

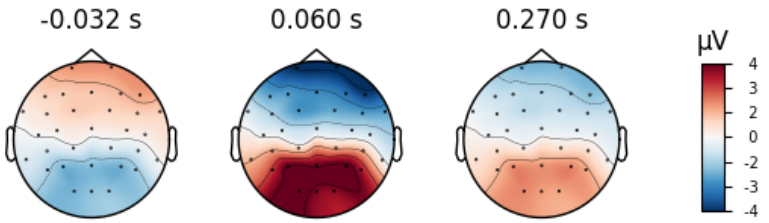


Figure 2.5: Exemplary topographies showing different frequency patterns at three different time points relative to a visual stimulus

are frequently considered when evaluating the EEG in the frequency domain: Delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-14 Hz), beta (14-30 Hz), and gamma (30-70 Hz). These frequency bands are based on experimental findings and vary slightly in the literature [1].

The advantages of EEG over other techniques for measuring brain activity are its low acquisition cost and small size. Additionally, EEG is rather insensitive to body motion, and techniques for automatically removing motion artifacts are available. Unfortunately, EEG has a low spatial resolution because it does not measure individual neurons' activity but rather the synchronous activity of thousands or millions of neurons in aggregate. Measuring the activity of neurons in inner brain areas is more challenging than measuring activity in outer brain regions near the scalp and EEG recordings are typically only used for inferences about activity in the neocortex. Medical and high density research EEG measurements are timely to prepare due to placement of electrodes and the necessity to use of gel to improve conductivity [157].
On top of the aforementioned analysis of general brain activity patterns over time and brain regions, specific neuronal phenomena can be analyzed looking at time-locked or evoked potentials.

2.3.2 Evoked Potentials

When our senses are stimulated, our nervous system generates the described electrical signal in response. This signal is referred to as the sensory evoked potential (SEP) - an electrical potential that occurs in response to the presentation of sensory stimuli and is detectable using EEG. The EEG measures changes in SEP amplitudes that range from less than a microvolt to several microvolts [110].

Each sensory organ can theoretically be stimulated to produce an evoked potential. In practice, mainly auditory EPs (response to an auditory sense stimulation), somatosensory EPs (response elicited by tactile or electrical stimulation of a peripheral sensory nerve), and visual EPs (response elicited by visual stimulation) are used. Following the objective of this dissertation, the main focus will be on visually evoked potentials. Due to the location of the primary visual cortex, the electrical response can be recorded primarily in the occipital region using EEG [110].

Steady-State Visually Evoked Potentials

The stimulus's repeated appearance induces a phenomenon known as a steady-state evoked potential (SSEP). A constant frequency stimulation induces a harmonic electrical response in the brain. In other

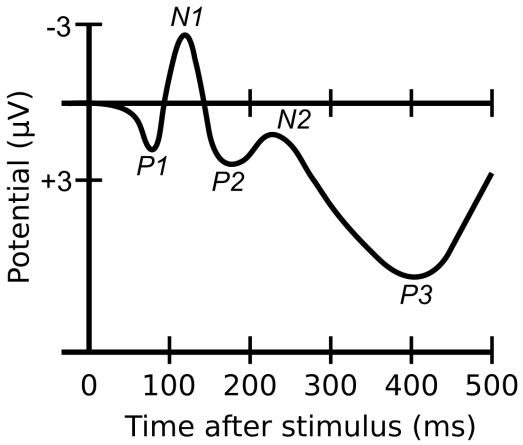


Figure 2.6: Schematic waveform representation of several important ERP components including the N100 (N1) and P300 (P3) from Cipresso et al. [35]

words, rhythmic sensory stimulation induces a rhythmic neural response. When the amplitude and phase of the frequency components remain constant, the brain enters a steady-state regime with the same frequency [159].

As the term suggests, steady-state visually-evoked potentials (SSVEPs) can be summarized as elicited steady neural responses to a constant visual stimulation at a specific frequency. Thus, by flickering the luminance of an object on a computer screen at a specific frequency, we anticipate detecting the same frequency in the EEG data. This concept is applied in neurology and neuroscience research with stimulus frequencies ranging from 3.5 to 75 Hz using EEG techniques [23]. SSVEPs have been reported to have a high signal-to-noise ratio and a high resistance to artifacts [188].

2.3.3 Event-Related Potentials

Event-related potentials (ERPs) are brain electrical potentials associated with the individual's internal or external stimuli, such as sensory, cognitive or motor events. Precisely, ERPs are electrical changes in the brain time-locked to an event. ERPs are analyzed with a focus on time-domain monitoring of the EEG signal at a specific interval after the stimulus onset. ERP signal waveforms are composed of a succession of positive and negative voltage deflections that correspond to a collection of underlying components. The trigger for the primary stimulus is a crucial factor in understanding these potentials, which are validated in terms of their amplitude and latency measures, and may be an intra- or inter-individual assessment. The majority of ERPs are denoted by a letter indicating their polarity (n = negative; p = positive), followed by a number indicating either the component's latency in milliseconds or its ordinal position in the waveform. For example, a negative-going peak that is the first significant peak in the waveform that frequently arrives approximately 100 milliseconds after a stimulus is introduced is frequently referred to as the N100 or N1 peak (see Figure 2.6). ERP component latencies are extremely varied, particularly for the later components associated with cognitive processing of the stimuli. The P300 component induced by the presentation of low-probability targets, for instance, may peak between 250 and 700 ms [12].

Error-Related Potentials

Error-related potentials can occur as a result of user perception of an error or unexpected behavior. Perceiving errors results in quantifiable electrical signals being generated in the medial frontal and central

brain regions [194].
These signals can be identified in EEG data by a specific characteristic: A negative deflection occurs between 50 and 200 ms after the error occurs, followed by a positive deflection between 200 and 500 ms. Nevertheless, some variations were observed across experiments. For instance, Error-Related Potential signals with reversed polarity, resulting in a positive deflection followed by a negative deflection. Additionally, the presence of more than two peaks with unequal latencies is a variation.
The literature makes distinctions between various types of error-related potentials [103]:

- When users are prompted to respond as quickly as possible to a question, they are referred to as "Response Error-Related Potentials".
- "Interaction Error-Related Potentials" occur during machine-user interaction, when the machine misinterprets a command.
- "Observation Error-Related Potentials" occur when a machine or external system detects an error.
- "Feedback Error-Related Potentials" are generated when an incorrect piece of feedback is detected on a task.

Despite their dissimilar apices, polarities, and latencies, error-related potentials exhibit comparable signal distributions in the frontal and parietal lobes. When features in the spectral domain are examined, an increase in activity is particularly noticeable in the theta-frequency band (4–7 Hz), as well as in the alpha-band (10–14 Hz). A combination of temporal and spectral features has been shown to be effective for detecting error-related potentials [103].

Fixation-Related Potentials

Fixation-related potentials (FRPs) - also called eye-fixation-related potentials (EFRPs) - rely on events associated with eye fixation to contextualize brain activity. They are also a subtype of ERPs, with the eye fixation serving as the event upon which the neural activity is time-locked. The recording requires both EEG and eye tracking technology, and a precise synchronization. In a first step, eye fixations need to be detected so that in the second step, the recorded EEG data can be related to the fixation onset. FRPs can be used to get insight into the perceived difficulty of a task in general, to identify periods of increased attentional effort, and to distinguish between periods of exploration and periods of active interaction [190].

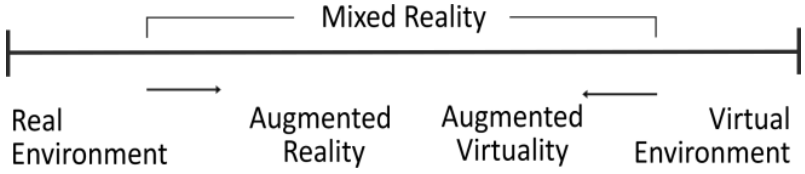


Figure 2.7: Milgram's Reality-Virtuality Continuum adapted from Milgram et al. [129]

2.4 Augmented Reality

There is a fundamental gap between the vast amount of digital data at our disposal and the actual world in which we use it. While reality is three-dimensional, the wealth of data that currently informs our judgments and actions is locked on two-dimensional pages and displays. This divide between the physical and digital worlds impairs our capacity to leverage the flood of data and insights generated by billions of smart, connected products globally. As briefly described before, augmented reality is an attempt to bridge this gap. It has been suggested as early as the 1950s when cinematographer Morton Heilig suggested "Sensorama" - the cinema of the future that takes in all the senses of the viewer [78, 93].

2.4.1 Definition

Augmented reality delivers information as a three-dimensional "experience" superimposed on the real object, rather than as a two-dimensional page on a screen. Thus, what the user sees is a hybrid of the real and the virtual, blurring the line between them. Accordingly, the term "augmented reality" describes a direct or indirect view of a physical real-world environment that has been enhanced/augmented by the inclusion of virtual computer-generated information in a real time, interactive manner [31].

In 1994, Milgram's reality-virtuality continuum was first presented (see Figure 2.7, [129]). It describes the transitioning from a real to virtual environment as a continuum that is called mixed reality. Mixed reality consists of augmented reality and augmented virtuality (AV), with AR closer to the real world and AV closer to a purely virtual environment. The virtual environment is most generally referred to as virtual reality, in which the users are in a synthetic world without being exposed to the real world. VR renders the real world virtually undetectable, in contrast to AR, which augments the perceptible real world.
One common key aspect of augmented and virtual reality is that they must be real-time in order to provide an authentic illusion of virtual content's physical existence. When mixed reality is not responsive, it loses its desirable and authentic façade, which is critical for its believability and interactivity.
Milgram's reality-virtuality continuum was recently revisited by Skarbez et al. [169] to include the research experiences of the past 25 years into the idea. One major criticism is that it is in fact not a continuum, because true pure virtuality can never be achieved as long as there is a self-perception of the user involved. Every suggested system presents mixed exteroceptive and interoceptive stimuli to the user. Only if these interoceptive stimuli were also virtually generated, a true virtual reality system could be implemented. Most importantly, because of this limitation, they adapt the key dimensions of mixed reality systems to be "Extent of World Knowledge" and "Immersion".
On top of that, the current description of mixed reality was too centered around visual stimuli and information, and other exteroceptive senses should be considered in the definitions. This includes auditory stimuli [115, 165], tactile and haptic information [13, 164], taste experiences [89, 132], and olfactory presentations as they were already suggested in Heilig's Sensorama [60, 78, 193]. However, for this thesis, the focus is also on augmented reality Systems that are mainly concerned with visual input. Current AR

(a) Google's Google Glasses (b) Microsoft's Hololens 2

Figure 2.8: Example AR HMD devices

technique centers around virtual display of visual information and, as it will be argued later, we are highly biased towards and reliant on our vision. It is the most prominent factor for our attention.

2.4.2 Technology and Devices

The technological aspects of AR systems compromise the display, the tracking mechanisms and sensors, and the computer. The specifics are not necessary for the scope of this thesis; but a basic understanding, especially of the displays, is required.
The tracking mechanisms and sensors are necessary for the aforementioned world knowledge. As the user moves, the AR contents' size and orientation need to be automatically adjusted to the changing context. New graphical or textual information is displayed while other information is hidden. This is achieved using "world cameras", GPS, compasses, and possibly other movement- or orientation-related sensors such as accelerometers and gyroscopes [31].
Additional meta world-knowledge must be supplied by the application itself, or through various user input mechanisms (gesture recognition, voice commands, mechanical buttons). In industrial environments, users with varying roles, such as a machine operator and a maintenance specialist, can view the same object but receive customized augmented reality experiences.

Devices

Broadly speaking, AR displays can be divided into three categories: Head-mounted displays, handheld devices and spatial displays. An HMD is worn on the head, like glasses or a helmet, and a screen is placed automatically in front of the user's eyes. This screen can either be optical see-through (OST; transparent, so the real-world content stays visible), or a video see-through (VST; displays real and virtual content on the non-transparent screen). An early consumer-oriented HMD AR device was Google's Google Glasses (see Figure 2.8a) and the current state-of-the-art HMD AR device is the Microsoft Hololens 2 that was released in 2019 (see Figure 2.8b). Both systems use the optical see-through technology that is considered to be more natural, because the real world-surroundings are not influenced by the display quality.
Instead, for handheld devices, the display is typically a smartphone- or tablet-like portable device, as the name suggests. With the recent advances for smartphone's computing powers, they have become a popular AR displaying device, using all the inbuilt sensors and cameras, as well as the big screens to display the

AR content. They are considered video see-through, because the displayed content is a representation of the real world recorded through the camera lens. They are very wide spread but reduce the immersiveness of the content and have a smaller display size which inhibits the presentation of 3D virtual content (see Figure 1.2 for an example application).
Spatial AR systems do not require additional displays because they project the virtual content directly onto objects. This content is less portable and not user specific. One example for such spatial AR are head-up displays, which are typically found in airplane or automobile windows. Stationary video see-through or optical see-through displays are also considered spatial AR devices and can be a lot bigger than handheld or HMD displays.
The advantages of HMDs for AR are the screen's broader visual coverage, which enables a better sense of immersion and the free-hands, while the obvious advantage of smartphones as AR displays are their availability and size.

2.4.3 Example AR Applications

Smartphone-based AR applications are more widespread than HMD-based applications, simply because of their high availability.
One of the most widely-known instances of AR is the game Pokémon GO (Niantic, 2016), which uses AR to transfer the game's creatures onto the real world using 3D models. This is accomplished by enhancing the camera feed on the smartphone. Other examples of augmented reality include the IKEA Place App, which enables virtual placement and testing of furniture at home (see Figure 1.2); Sephora's Virtual Artist, which enables virtual testing of makeup products; and Google's augmented reality translator, which replaces text with augmented reality translations. These are just a few examples of the numerous augmented reality apps already available. AR also has a variety of applications in sectors such as psychology, medicine, education, and the military [31].

2.5 Machine Learning

As the term suggests, machine learning (ML) is a subfield of artificial intelligence (AI) and computer science where algorithms improve through experience - the machines learn. It is concerned with the use of data and algorithms to mimic the way humans learn, gradually improving its performance. Algorithms are trained to produce classifications or predictions using statistical approaches, resulting in the discovery of critical insights in the data.
It's a rapidly growing discipline with applications in almost every industry and field that can benefit from data analysis; for instance, weather forecasting, self-driving cars, business intelligence and customer relationship management, and disease spreading forecasts as for the ongoing Covid-19 pandemic.
In the context of this thesis, machine learning is used to classify the attentional state of the user based on the available biosignal data.

Generally speaking, an algorithm will generate an estimate about a pattern in the available input data, which can be labeled or unlabeled. An error function is used to evaluate a model's prediction. If there are known examples, an error function can be used to determine the model's accuracy. If the model fits the data points in the training set better than the known example, weights are modified to lessen the difference between the known example and the model estimate. The algorithm will repeat this "assess and optimize" procedure, automatically updating weights to improve the performance [117].

There are different approaches how the algorithm improves it's predictions:
During **supervised learning** labeled datasets are used to train the algorithms. With the available ground truth classes for each sample, a model's accuracy can be calculated and used to evaluate the performance. As input data is supplied into the model, it modifies it's weights until it is properly fitted. Large amounts of data and an appropriate number of model parameters are necessary to prevent the model from underfitting. Overfitting on the other hand can be avoided by splitting the data into specific training and test sets or using cross-validation techniques.
Unsupervised learning analyzes and clusters unlabeled datasets using machine learning techniques. Without human assistance, these algorithms uncover hidden patterns or data groupings.
Semi-supervised learning is a viable alternative to both supervised and unsupervised learning. It uses a smaller labeled data set to aid classification and feature extraction from a larger unlabeled data set during training. Semi-supervised learning can help alleviate the problem of not having enough labeled data to train a supervised learning algorithm (or of not being able to afford to label enough data).
Reinforcement machine learning is a type of behavioral machine learning similar to supervised learning, except that the algorithm is not trained on sample data. This model is self-learning through trial and error. A series of successful outcomes will be reinforced in order to generate the most appropriate model or policy for a particular problem.

In this thesis, the classification problem of generating attention labels for biosignal data can be tackled using supervised learning approaches. All of the machine learning tasks required during the work on this thesis are considered binary or sometimes multi-class classification problems. Thus, the data needs to be divided into two or more categories. The observed set of user data x is used as the input and the known label y is the current attentional state of the user. The classification problem is then to find a good model for the class y to make a correct prediction for any given observation of x that was not previously used to train the classifier. Finding suitable algorithms and appropriate classification strategies is one of the main challenges when designing attention-aware interaction systems for AR. Many problems, such as training-free person-independent models, real-time aspects, feature extraction methods, and combining

multimodal datasets need to be addressed. Among other algorithms for comparison, most of the studies for this dissertation focused on Linear Discriminant Analysis because of it's rather straight-forward interpretability, and Convolutional Neural Networks because of the power of deep learning approaches.

2.5.1 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a generalization of Fisher's linear discriminant [36] that is applicable to supervised learning and classification tasks. Thus, class labels have to be known a priori. Its objective is to discover a linear combination of features that defines or distinguishes two or more classes of datapoints. As in this thesis, the resulting combination can be employed as a linear classifier, but it can also be utilized for dimensionality reduction prior to subsequent classification. Methodologically, it is closely related to analysis of relevance (ANOVA), regression analysis and principle component analysis (PCA) [119]. The LDA procedure attempts to describe categorical dependent variables (here: the attentional state label) as a linear combination of continuous independent variables (here: the user data).
Assuming multivariate normality and homoscedasticity, thus a normal distribution for the conditional probability density functions and homogeneity of variance and covariance among the groups, and following the Bayesian rule for class allocation the following discriminant function can be obtained:

$$\delta_K(x) = x^T \sum^{-1} \mu_K - \frac{1}{2} \mu_K^T \sum^{-1} \mu_K + log \pi_K$$

which is a linear function in x with K as the number of classes, μ as the class mean, the sum representing the covariance, and π as the class prior probabilities. By randomly choosing one class as the base class, LDA requires just $K - 1$ such discriminant functions (subtracting the base class likelihood from all other classes). In geometrical terms this means that the criterion of an observation of x being classified as a class y if x_i it is located on a certain side of a hyperplane.
When the number of features substantially exceeds the number of samples, the covariance matrix may be incorrectly calculated, requiring the use of shrinking. This means that, via a penalty parameter, the individual covariance matrix shrinks toward a common pooled covariance matrix [14].

Despite its simplicity, LDA frequently yields classification results that are robust, reasonable, and interpretable. When handling real-world classification problems, LDA is often used as a starting point and benchmark method before moving on to more intricate and flexible methods.

2.5.2 Neural Networks

Deep learning is a sophisticated form of machine learning that has attracted widespread attention for its great results, for instance, in computer vision (e.g., image recognition) or natural language processing. An artificial neural network is inspired by biological systems' information processing and communication nodes. By design, input data is routed through a network's layers, each of which has numerous nodes n representing "neurons"; each of which is connected to the others and has a weight w and threshold linked with it. The neural network is typically composed of an input layer, one or more hidden layers, and an output layer (see Figure 2.9). For the described classification problems, the input layer has the size of the feature vectors and the output layer's size is equivalent to the number of possible classes K. The term "deep" neural network is used when a neural network has more than three layers — including the input and output layer [90].

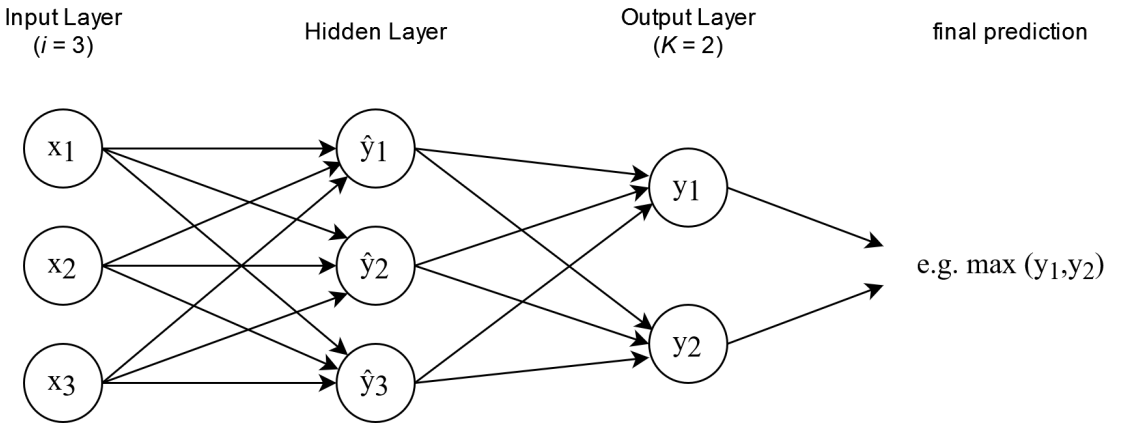


Figure 2.9: Exemplary artificial neural network for a binary classification problem with an input feature vector of size 3 and one hidden layer.

The specific representation of the data is achieved by adjusting the weights using forward propagation and backpropagation. Forward propagation is the prediction step where the network takes the input data x_i (i being the number of input features) to give a final output value y. The input is passed through each layer by computing the weighted sum and using that as the input for a (typically nonlinear) activation function $f(\cdot)$:

$$\hat{y}_n = f(w * x + b)$$

This is repeated for each node in each layer using $\hat{y}_n$ as the input for the next layer until arriving at the final output y. For classification purposes, there is usually a final output y for each class representing a confidence with which the input data belonged to the class. For the final prediction, the class with the highest confidence is chosen. Thus, the output is a weighted set of classes, and is frequently limited to the class with the highest weight.
Backpropagation is the training step in which the network compares y with the ground truth by computing the loss and afterwards adjusting the weights to improve the accuracy. A loss function's slope is recursively calculated with respect to each parameter and an optimization function is used to update the parameters based on the gradient information [90].

Neural networks that strictly pass information forward during the forward propagation are commonly called feedforward neural networks. Several types of layers can be combined in a neural network to specifically suit the problem at hand. Each layer performs a specific function and contributes to the achievement of a specific classification result.

Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are a type of Feedforward Neural Network that have one or more convolutional layers. They, along with the pooling layers, form the basis for CNNs. Typically, they are used for multidimensional input data (i.e. images) for automatic feature extraction [155]. Meanwhile, CNNs have successfully been used in EEG-based brain-computer interfaces [196].

Specifically, convolutional layers act as feature extractors, and their neurons are arranged in feature maps. Each neuron in a feature map has a receptive field that represents the connection between the neu-

ron and its neighbors from the previous layer via trainable weights. The convolution operation produces a feature map by applying a filter composed of a set of trainable weights to the input. The convolutional result is expressed using a nonlinear activation function (i.e. Rectified Linear Units (ReLU) or Sigmoid). A convolutional layer is capable of employing a large number of filters and learns them during training. Filters are well-known in the field of image processing, where they are used to extract information from images, for instance, to determine the presence of vertical or horizontal lines. Due to the convolutional layer's ability to learn the filters independently and to apply multiple filters in a single layer, it is possible to detect complex features in an image [155]. The computation of Y_k feature map can be formally calculated as follows:

$$Y_k = f(W_k * x)$$

Again, the input is denoted as x and the nonlinear activation function is denoted as $f(\cdot)$. Additionally, the filter to this feature map is denoted as W_k [108].
The pooling layers are the second critical component of CNNs, as they are used to reduce the spatial resolution of feature maps. This is to ensure that the output data is spatially invariant to distortions and shifts in the input data. A max pooling layer that determines the maximum value within a receptive field and forwards it to the next layer in the network is frequently used for this purpose [155].

Overfitting is one of the difficulties associated with training CNNs. In this case, the network no longer generalizes sufficiently, but has learned to recognize existing noise in individual training data as features, making it less capable of classifying previously unseen data [155]. Regularization techniques to countermeasure overfitting in CNNs are for example dropout layers, stochastic pooling, and special activation functions. Dropout is frequently successful, blocking some neurons in the hidden layers randomly during each training run. This forces the network to learn specific features through the use of additional neurons and associated parameters, frequently improving the network's generalization after training [66].

Numerous well-known network architectures are now available for use with similar problems or as a guide for developing your own network architecture. For instance, the LeNet, which LeCun invented in 1998 and is widely used in the field of image recognition [109].

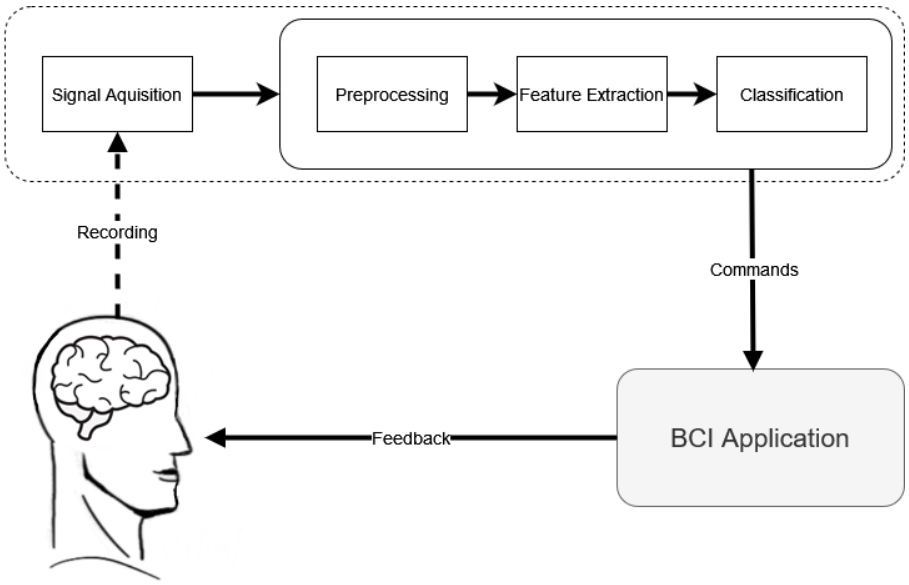


Figure 2.10: Generalized Brain-Computer Interface components

2.6 Brain-Computer Interface

In daily life, interactions with technological systems are primarily achieved by electrical signals transmitted from the brain to muscles, resulting in direct physical contact with input interfaces (e.g., pressing a button).
Through brain-computer interfaces or brain-machine interfaces, direct communication with a computer can be established utilizing cerebral activity as an input signal. In 1973, Vidal [182] suggested one of the first BCI projects. The idea was to employ neural impulses in a person-computer interaction, allowing computers to act as a prosthetic extension of the brain. Previously, in 1969, Fetz [57] demonstrated that impulses from single cortical neurons in monkeys may be utilized to drive a meter needle. However, comprehensive human studies did not begin until the 1970s and progressed slowly, constrained by computer capabilities and missing knowledge of brain physiology. Elbert et al. [50] published a study in 1980 that showed individuals given biofeedback sessions of slow cortical potentials in EEG activity can control the vertical movement of a rocket image traveling across a television screen. The field of brain-computer interface research is advancing at a tremendous speed since then, as computational power has increased and brain data recording equipment have become increasingly available.
While some BCIs use blood oxygenation to measure brain activity, the majority detect electrical brain signals (through EEG) or the magnetic fields generated by them [139]. As stated before, depending on the location of the signal and the band frequency, brainwaves reflect various aspects of our cognition and mental states. Hence, the obtained data can be evaluated using machine learning to provide an input for an application without requiring explicit user feedback.
This permits control that is not dependent on speech or movement, which is particularly advantageous for physically disabled individuals; for example, to control prosthetic limbs or to generate speech [191]. On top of that, BCIs can be utilized for non-medical purposes such as monitoring states of awareness and emotion, controlling vehicles and robotics, and engaging with virtual environments in games [148]. For the purpose of this thesis, BCIs are used to detect users' attentional states.

2.6.1 Definition

There is not one agreed-upon definition of what a brain-computer-interface is because the focus on the components might slightly differ, but the general notion is usually shared. The most important aspects are the data recording, the following processing of the data and how that information is used in the system. Nam et al. [139] put the focus on the data acquisition and modality, defining a BCI as

> *"a communication system in which messages or commands that an individual [140] sends to the external world do not pass through the brain's normal output pathways of peripheral nerves and muscles [191]"*

Shih et al. [167] define BCI systems using a more holistic point of view:

> *"A BCI is a computer-based system that acquires brain signals, analyzes them, and translates them into commands that are relayed to an output device to carry out a desired action."*

Taken together, BCIs bypass the brain's usual output channels of peripheral nerves and muscles, thus detecting and utilizing signals produced by the central nervous system. Following this definition, a user interface based on eye tracking is no BCI. An EEG is not a BCI in itself because it just records brain signals and does not provide an output that affects the user's environment; thus, not fulfilling the second part of the definition. A BCI systems needs to generate a command for an application (see Figure 2.10). Hence, the EEG recording is just one component of the BCI system. Further, the data needs to be processed and used as an input command for an application.
Brain-computer interfaces do not extract information from unsuspecting or unwilling users; rather, they enable users to interact with the world through the use of brain signals rather than muscles.

2.6.2 Devices

The BCI system typically consists of a recording device, a computing device for the signal processing and an application device to perform the feedback. These can be separate devices that transfer information or they can be built as one device with all these components.
BCIs can be grouped as invasive or non-invasive according to their data recording procedure. Invasive BCIs require surgery to use electrocorticography (ECoG) or an intracortical electrode array. The intracortical electrode array is inserted directly into the cerebral cortex. Such techniques typically provide an improved signal-to-noise ratio and a higher spatial resolution than non-invasive techniques.
Non-invasive BCIs do not require surgery and are therefor easier to install but have a worse signal-to-noise-ratio. EEG-based BCIs are an example for non-invasive BCIs because the electrodes are placed on the scalp [139] and are of the most popular types of BCI, being used in 60% of the BCI research from 2007 to 2011 [85]. As a result, for everyday life applications requiring widespread use, such as the proposed system, non-invasive BCIs should be preferred despite the lower signal quality.

BCI projects for everyday use can be found in the gaming, education, and wellness sectors [148]. Current state-of-the-art examples are products from the company NeuroSky, Inc. [49] and Link from Neuralink. The latter has garnered widespread attention due to its popular founder Elon Musk and is developing an invasive BCI to control a cursor on smartphones and computers [137].

2.6.3 Types of BCIs

As previously stated, signal processing and feedback mechanisms are an integral part of the BCI system, in addition to data recording. These steps can vary significantly depending on the application's context.

Preprocessing signals entails the removal of artifacts and the application of spatial and temporal filtering techniques, such as a notch filter to eliminate power line noise. This is followed by the extraction and classification of features that are, again, context- and data-dependent.
The signal can be processed concurrently with or after the acquisition. The vast majority of BCI research is conducted offline, which enables more sophisticated processing and increases classification accuracy. On the other hand, a real-time BCI processes data synchronously with the system's operation. These types of BCI systems are more appropriate for most actual use cases, and need to be considered for consumer-oriented applications and systems where the feedback needs to be integrated into the usage. For this thesis, several studies were performed using offline data processing and additionally, a lot of effort was put into developing real-time BCIs as proposed for the overall system.

User Intent

Another aspect that affects the BCI is the necessity and incorporation of user intent.
BCI applications can be active, reactive, or passive. The first two rely on deliberate modulation of brain signals to communicate with the BCI. Active BCIs require users to alter their thinking intentionally in order to elicit specific brain signals that the BCI picks up on, such as motor imagery to move a prosthetic limb. External stimulation (e.g. flickering images) is used to modify detectable brain signals that serve as the basis for communication in reactive BCIs. SSVEP-based systems are an example of a frequently used reactive BCI [195]. By contrast, when a passive BCI is used, users are not required to act or react in any particular way. It is intended to be an implicit interaction in which the passive BCI monitors cognitive and affective states in the background by detecting automatic, spontaneous brain activity [139].

As discussed before, if the system is capable of predicting the state of humans via passive BCI, it will be able to adapt proactively to the human's needs. Most application cases for passive BCIs are accompanied by a complex, noisy environment. Additionally, the rich forms of real-world interaction and missing user intent results in diversified user state [195]. This can result in decreased brain data quality and is one of the main reasons research on such BCIs has mainly been performed in controlled laboratory environments and is only recently tested more frequently in real world settings [11].

Communication Direction

Furthermore, BCIs distinguish between unidirectional and bidirectional information exchange between the brain and the computer.
Unidirectional BCIs evaluate brain signals and create commands, such as directing the movement of a prosthetic limb. Bidirectional BCIs add feedback into the communication process, allowing the computer to directly send information to the brain (i.e. the temperature of an object that is touched using a prosthetic limb) [3].
For the presented work, only unidirectional BCIs are required. The detected user state will only influence the system behavior and user interfaces.

2.6.4 Challenges

Due to the complexity of the technological challenges that were described, the technology of BCIs is still on the verge of market maturity. The requirement to have high quality brain data brings along challenges such as interoperability, inconvenience for users, low accessibility, and high prices [148]. More-

over, reliable command recognition for active BCIs requires extensive user training and a high level of concentration. Another challenge is the transition from the controlled laboratory environment to the real world, where users are constantly confronted with external distractions requiring the system to cope with ambient noise and unidentified interfering signals [154].

2.6.5 Related BCI Research

The objective of this dissertation is the exploration of attention-aware AR applications. One possible interaction method is the use of BCIs that passively detect the attentional state of the user to adapt the system behavior. This is an innovative field with no previous work to build up on, but there are related BCI studies that either combine BCIs and AR or aim at adapting a non AR system to the attentional state via BCI or gaze. Although some previous works have used other brain data acquisition methods, such as fNIRS [19], this work will focus on related EEG-based BCI systems and augmented reality.

The first studies on AR-BCI systems were already published in 2010 [95, 111]. In 2011, Takano et al. [174] tested a P300-based BCI that could be operated using a see-through HMD and found that the participants could successfully use the system without intensive training.
A general feasibility assessment and a current state-of-the-art report for combining BCIs and AR were done by Si-Mohammed et al. [133, 134]. They classified systems that integrate augmented reality and BCI according to their application domains of medicine, robotics, home automation, and visualization of brain activity. The survey found that the majority of earlier works used P300 or SSVEP paradigms in conjunction with EEG in VST systems, and that robotics is the primary application field with the greatest number of current systems [133]. Further, they evaluated the potential of employing BCI in augmented reality environments OST-HMDs. The experimental results demonstrated that an EEG-based BCI and an OST-HMD are very compatible and that minor head motions may be tolerated. Next, they defined a design space for SSVEPs in which they could evaluate orientation, frame-of-reference, anchoring, size, and explicitness in the context of a mobile robot arm. Based on their findings, they developed an operational prototype of a real mobile robot that is operated in augmented reality via a BCI and a HoloLens headset [134].

Indeed, a large number of studies on SSVEP-based AR BCI systems can be found. The performance of AR-SSVEP studies achieve similar levels as computer-based SSVEP studies [198]. This statement was also supported by the results of Wang et al. [187] who found that wearing an HMD like the HoloLens does not significantly affect the EEG data acquisition.
Faller et al. [55] used an SSVEP-based BCI as a silent, hands-free input channel to an HMD. Similarly, Kishore et al. [100] compared an SSVEP-based BCI with a gaze-based input mechanism to a head-mounted display for controlling a robot. Wang et al. [186] developed a wearable SSVEP-based BCI that provides three-dimensional navigation of quadcopter flying with immersive first-person visual feedback via an HMD, and Ke et al. [98] presented an online SSVEP-based BCI using an OST-HMD that successfully differentiated between eight different classes of targets. The SSVEP-based BCI paradigm used in Park et al. [146] was tested with three different types of visual stimuli on an HMD for an online home appliance control system. The results showed an accuracy of over 90% with an information transfer rate of 37.4 bits/minute, which the authors claim outperforms previous works. However, Liu et al. [114] also achieved over 90% classification accuracy for their HMD SSVEP-BCI paradigm with eight classes and a much higher information transfer rate of over 60 bits/minute. The problem of multi target classification in such SSVEP settings for AR was also addressed by Zhao et al. [197], who suggest using CNN-based

classification instead of traditional canoncial correlation analysis or similar SSVEP specific methods.

As mentioned by Si-Mohammed et al. [133], the ERP P300 was also used frequently for AR-BCI systems. In a recent example, Kim et al. [99] tested P300 components for drone control in AR settings. The authors found that there are no significant usage differences for their proposed system for AR and VR.
Alternatively, in Mercier-Ganady et al. [126, 128], the authors used an augmented reality devices to show the user the abstract output of a BCI.

2.7 Attention-based Research and Applications

Due to its broad scope and complexity, attention has developed into a major area of research in philosophy, psychology, (cognitive) neuroscience, artificial intelligence, and neuropsychology. In medical research, there is an increased interest in diagnosing attention-related symptoms. Neuropsychological, psychophysical, neuroimaging, and electrophysiological techniques are used to elucidate how relevant and irrelevant information compete for processing resources that influence our behavior. Our senses inspire the diversity of scientific attention. On top of that, the extent to which sensory cues and signals in our environment influence and direct our attention is a very active area of research [172].

Neurophysiological research has demonstrated that a variety of modalities contain information about the current attentional state, and are thus suitable for real-time decoding of these biosignals in order to determine the focus of attention. Oculometric data, heart rate, or brain activity - inferred from electrical activity or physiological changes in the brain - are some exemplary modalities that have been frequently suggested and used.
Our understanding of the brain has advanced tremendously over the last few decades. With advancements in imaging techniques, we are increasingly able to visualize, interpret, and analyze brain activity, as well as to comprehend its anatomy, assign functionality to various areas, and decode signals into mental processes.
Recently, the research field devoted to single-trial classification of attention-related data has expanded. Nowadays, such data analysis can be performed immediately, allowing for real-time processing. The increased interest in and technical capabilities for implementing and executing machine learning algorithms enable the classification of recorded data with only a short delay.

2.7.1 Attentional Effects on Gaze Behavior

The detection of attentional states based on eye tracking data has been the center of several research works. A major advantage of eye tracking recordings compared to other biosignals (e.g. EEG, fMRI, fNIRS) is the fast and unintrusive setup that does not influence attention. Even webcams can be used to record reliable eye tracking data [124].
Although all sensory inputs have a major effect on attention, there is a natural inclination in attention research to highlight outcomes obtained with visual stimuli. According to Hutmacher [83], vision is socially and culturally dominant over other senses.
Human gaze behavior is influenced by both, the visual input (bottom-up) and the cognitive state or task (top-down). Researchers identified visual patterns that naturally draw attention by tracking people's eye movements when they were presented with various images. Oriented edges, spatial frequency, color contrast, intensity, or motion are considered "salient" patterns [88]. In general, saliency refers to image regions that draw attention and are most likely the product of the visual system's built-in feature detectors. When asked to saccade to a specific visual target humans often mistakenly saccade to a particularly prominent distractor [200] but these are usually short-term effects and not long-lived [47]. Exogenous attention has a brief duration, affecting covert attention for approximately 80—130 ms after the distractor appears [10].

When people do real-world tasks, the natural pattern of their eye movements can provide insight into underlying cognitive processes [75] and, as discussed before, the current locus of gaze can be used for overt attention evaluation, indicating attended objects or directions.

Hoffman and Subramaniam [82] showed a strong relationship between covert spatial attention and saccadic eye movements.
The eye gaze behavior during internally and externally directed attention was investigated and analyzed in detail in Benedek et al. [18] and Walcher et al. [185]. The authors found significant differences during goal-directed internal attention and external attention. Specifically, internally directed attention was associated with fewer and longer fixations, higher variability in pupil diameter, more and longer blinks, lower microsaccade frequency, more saccades and saccades with higher amplitudes, and a lower angle of eye vergence compared to times of externally directed attention.

2.7.2 Attention in the Brain

As mentioned previously, psychophysiological and electrophysiological methods can be used to investigate the neurophysiological correlates of (visual) attention in greater detail. Attention's effect on the response properties of receptive neuron fields has been precisely quantified and a great extend of related work is available. For this thesis, the focus will be on basic effects and concepts of attention in the brain, and neuronal activity that would be measurable using EEG, as suggested for the attention-aware interaction system. While much of the research on the neural correlates of sensory attention has been on the cortex, it appears that subcortical areas also contribute significantly to the control and performance benefits of attention (the superior colliculus as suggested by Krauzlis et al. [102] or the pulvinar as suggested by Zhou et al. [199]).
As described before, attention has many facets but the main focus will again be on visual attention, the effects of internally and externally directed attention and top-down or bottom-up attention regulation and shifts. The reported findings are consistent across multiple experiments (on humans or monkeys) and can be interpreted as quantifiable neural correlates.

Corbetta and Shulman [41] proposed the concept of the human brain having two physically and functionally different attention systems. A dorsal system (Dorsal Attention Network, DAN) was postulated to facilitate top-down guided voluntary attention allocation to locations or features, whilst a ventral system (Ventral Attention Network, VAN) was proposed to be involved in detecting unattended or unexpected inputs and prompting attention shifts. For EEG recordings, the DAN is represented in the superior parietal, occipital, and frontal brain electrodes. Supposedly, the VAN controls voluntary attention by releasing norepinephrine in non-cortical areas including the anterior insula and temporoparietal junction, and cortical areas like the anterior cingulate cortex and pre-fronts [150].
Vossel et al. [184] conclude from several related studies that neither of the two networks regulates attentional processes independently, but that their flexible interplay permits dynamic attention management in response to top-down goals and bottom-up sensory stimulation (see Figure 2.11).
Examining different cortical areas revealed altered firing rates in early vision-related processing areas such as V1 and V4 [121], as well as increased activity in areas of the prefrontal cortex and the intraparietal sulcus [61, 158].

Attentional regulation is recognized to reduce trial-to-trial variability for neuronal firing rates and noise correlations between pairs of neurons. Particularly, it alters the electrophysiological properties of neurons, reducing their chance of firing in bursts and the height of individual action potentials [6].
Hillyard and Anllo-Vento [80] conducted ERP studies to demonstrate that attended stimuli elicit larger potential amplitudes and shorter potential latencies than unattended stimuli. These findings were corroborated by those of Treue and Trujillo [176], who concluded that attention had a multiplicative effect on the tuning curves of neurons (the amplitude of the tuning curve increase, but not the width and

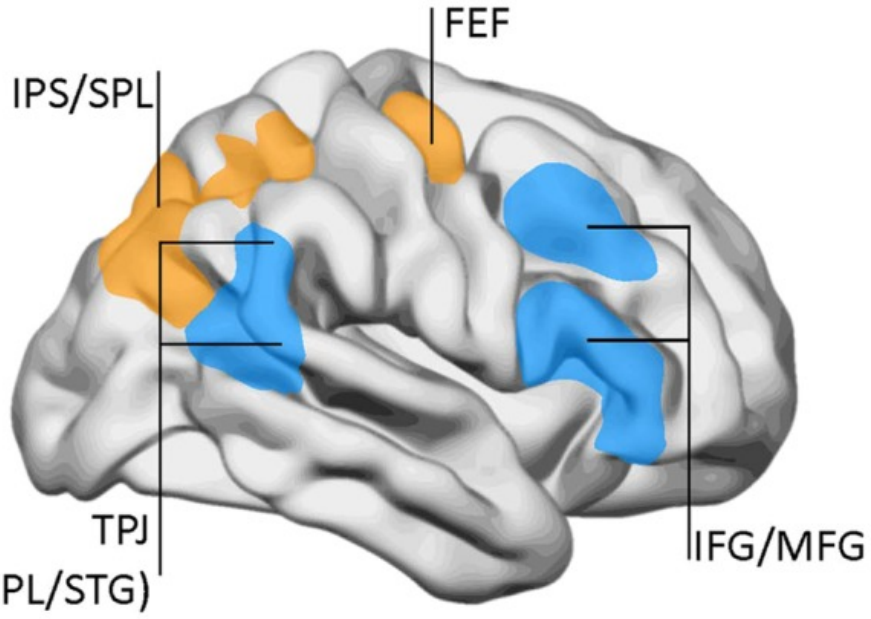


Figure 2.11: The dorsal attentional network (DAN, blue) and the ventral attention network (VAN, yellow). FEF = frontal eye fields; IPS = inferior parietal sulcus; SPL = superior parietal lobe; IFG = inferior frontal gyrus; IPL = inferior parietal lobe (posterior aspect); MFG = middle frontal gyrus; TPJ = temporo-parietal junction; STG = superior temporal gyrus. Figure taken from Aboitiz et al. [2], original data from Corbetta and Shulman [41].

shape).
The modifications associated with attention raise the signal-to-noise ratio of the neurons representing an attended stimulus and affect communication between brain areas by influencing neuronal synchronization. Specifically, Fries et al. [62] demonstrated that attention improves spike coherence in the gamma band. Synchronous firing of a group of neurons enhances their potential to influence shared downstream locations. Attention has been demonstrated to promote synchrony between the neurons in areas V1 and V4 that reflect attended stimuli [27].

Miller and Buschman [130] propose that the prefrontal cortex is a critical structure for top-down attention and Noudoost et al. [142] hypothesize that neurons in the visual system exhibit spatial attention correlates as a result of a distributed network of structures that is involved in the programming of saccadic eye movements. Thus, they exert a significant influence on neuronal activity along all visual pathways but the most significant increase in neuronal firing can be examined in later areas. In multitasking settings that require increased top-down attention, this increase reaches up to 20—30% [131].
Bottom-up attention processing appears to culminate in the formation of a saliency map in the lateral intraparietal area. The cells in this region respond when salient stimuli enter their receptive field [105].

In terms of frequency-bands, attention levels were connected with alpha-band (α) oscillations as early as 1929 by Hans Berger [21]. Particularly upper α frequencies (11.1—13 Hz) have been identified as the primary physiological indicator of a low anxiety state. A shift in EEG power toward low-frequency bands, and an increase in low α (8—11 Hz) waves is associated with increased fatigue [112]. These oscillations are particularly prominent in frontal, frontotemporal, and visual regions, with the highest frequencies

often declining in amplitude [68, 107].

More specifically, α-power increases in right parietal cortex and a lower α desynchronization reflect focused internal attention [17, 32]. Cooper et al. [40] also presented evidence that increased α-power is a frequency marker for the active suppression of external stimuli during times of internally directed attention. In Benedek et al. [16] internally directed attention was linked with greater activity in the right anterior inferior parietal lobe and prolonged inactivation of areas representing the dorsal attention network in the superior parietal and occipital lobes. Additionally, the right anterior inferior parietal lobe demonstrated enhanced connection with occipital regions, implying an active top-down strategy for protecting ongoing internal processes from distracting sensory stimuli.

2.7.3 Attention Classification using EEG and Eye Tracking

So far, neurophysiological and eye gaze correlates of attentional states were discussed. The measurable differences for both modalities suggest that the targeted classification task for this project is reasonable and similar classification studies have been performed in related works.

These works have shown that eye gaze behavior can be used to classify cognitive or psychological disorders (e.g., Autism [143]). Machine learning is utilized to classify the eye movement patterns, and the predictions are employed in a variety of applications [56, 125, 178]. Hutt et al. [84] classified mind wandering during lecture viewing significantly better than chance level and discussed the advantages these methods have for massive open online courses. For the classification, they extracted features, such as the number of fixations and the fixation duration and classified them using Bayesian networks. Xuelin Huang et al. [192] follow a similar motivation and detect internal thought during e-learning. Their features are based on eye-vergence information that is available when a binocular eye tracker is used. The obtained results are reliable in times of math-solving, daily computer-based activities (coding, browsing, reading), and free viewing of online lectures.
In Annerer-Walcher et al. [9], an LSTM was successfully used to classify the raw eye tracking signals in times of internally and externally directed attention with an accuracy of 75% on a single trial basis.

Recently, Souza and Naves [170] provided a review of works published between 2010 and 2020 on attention detection using EEG specifically for virtual environments. They summarize that it is critical to avoid exaggerating the outcomes of particular machine learning algorithms or extrapolating to a single, universal model. Accordingly, the future of machine learning and models for detecting attention is more personalized. The authors also note that there is still a dearth of high-resolution EEG techniques that visualize brain connections in order to obtain information about immersion, attention, and cognitive load. Attention perception in an EEG signal is more dependent on the electrode placement than on the analytical procedure used to derive signal features. They note that under long-term uses, the durability of constituent neuronal properties, particularly under stress and exhaustion, remains unknown. Diverse approaches were used to examine attention allocation under various levels of immersion and cognitive load circumstances. Additionally, there is a difference when processing the EEG signal under controlled settings versus when it must be placed in everyday applications, in touch with natural stimuli from its environment, in the context of free viewing.
Benedek et al. [17] were able to differentiate between internal and external attention by using the frequency power spectra of the right parietal region of the brain as markers; in Putze et al. [152], it was shown that EEG data could be classified to discriminate between internal and external attention processes on a single-trial basis.

2.7.4 Attention in Augmented Reality Settings

A few studies have been conducted to determine the effects of augmented reality interfaces on the user. Although the rendering power of augmented reality devices has increased significantly in recent years, the user interface is still largely inspired by traditional desktop-based scenarios and makes inefficient use of AR's additional capabilities. While the fusion of real and virtual objects provides numerous opportunities to present additional information, it also introduces some undesirable side effects, such as split attention and increased visual complexity. These cognitive difficulties can result in mental fatigue or a high level of distraction [94]. Eyraud et al. [54] found that while AR enhances the user's ability to correctly allocate their attention when the displayed content supports the user's specific task, it distracts and degrades their attention allocation when the displayed content does not support the user's specific task. Dixon et al. [46] demonstrated in a practical example how distraction caused by augmented reality assistance during medical surgery resulted in inattentional blindness to unexpected findings. Taken together, this means that an augmented reality system can be designed to be unobtrusive when it's slow in its display and change of virtual content [42].

2.7.5 Attention-Adaptive Systems

The optimization of the user interfaces based on current attentional user states has only been a goal of a few of studies. The aforementioned differences in averaged neurophysiological data and gaze behavior suggest candidates for discriminative features that can be used in attention-adaptive systems that apply passive BCIs or eye tracking analysis. In addition to the suggested use case of HMD-AR systems, such interfaces may be beneficial in certain fields like air traffic or drone control, where operators are subjected to high levels of stress and cannot be distracted. The aim is to adapt the visible information to reduce the complexity of the task and thus the workload or distraction. Not all reported studies strictly work with attentional states as defined before; some also relate to similar or connected mental states.

Several available studies rely on psychological knowledge about visual attention mechanisms, such as cueing. Bonanni et al. [26] reduced users' cognitive load through the use of layered interfaces that adhered to attention theory's cueing and search principles, Lu et al. [116] evaluated a less distracting method of subtle cueing to aid visual search in augmented reality settings, and Biocca et al. [24] pioneered the use of an attention funnel as a three-dimensional cursor to direct the user's attention away from objects within the visual field. While these studies provide guidelines for developing attention-aware augmented reality interfaces, they do not incorporate a real-time assessment of the user's attentional state as suggested in this thesis. Current information about the gaze direction was used in McNamara and Kabeerdoss [123] and Sridharan and Bailey [171]. McNamara and Kabeerdoss [123] used gaze position to place labels in a mobile augmented reality environment in order to avoid visual clutter and masking. Sridharan and Bailey [171] directed the user's attention by combining gaze information, visual saliency, and key features in the image display visual cues in peripheral regions of the field of view.
As argued before, there is more to be considered about the attentional state than the gaze direction.

A few examples of systems that conduct research in similar directions have been published:
An early study by Heger et al. [77] in 2011 demonstrated a significant improvement in the effectiveness, efficiency, and user satisfaction of an EEG-based workload-adaptive interface when compared to a non-adaptive baseline in a user study on a robot-based information presentation system.
Mercier-Ganady et al. [127] used relaxation and concentration levels measured by an EEG system to adapt the level of camouflage ("the power of becoming invisible"). They augmented a Harry Potter world

by combining a computer monitor and an optical tracking system that would adapt the content based on the user state. The authors report that the mental state adaptation was more intuitive and motivating than button presses. The same aim of self-regulating sustained attention was followed by Karran et al. [96] who compared continuous neurofeedback to event-related feedback to no feedback and found that providing participants with the ability to self-regulate sustained attention has the potential to keep them engaged for extended periods of time and to increase on-task performance moderately while decreasing on-task error.

Aliakbaryhosseinabadi et al. [5] proposed an innovative online EEG-based BCI system that is capable of adapting to changes in the users' attention during real-time movement execution. Feedback on the user's attentional status reduced the amount of attentional diversion caused by the oddball task in both healthy controls and stroke patients. Their findings established that users' attention can be monitored and that real-time neurofeedback on the user's attentional state may be used to redirect the user's attention. The authors argue that monitoring users' attention level will have a significant impact on the future of BCI for neurorehabilitation.

2.8 Gaps to Fill

As outlined, the thesis's underlying question is composed of several thematic areas, the most significant of which are attention, eye-tracking, EEG-based BCIs, and augmented reality. While BCI and AR are relatively new fields of study, they are all supported by active communities and are constantly evolving.

For a very long time, researchers have been studying various cognitive states, such as human attention. While the precise definition is continually developing, there is already a wealth of information about the influences, effects, and neural activity associated with a variety of different attention states. Thus, based on current research and recent technological improvements, there is reason to believe that modeling and classification of attentional user states will be quite feasible using modern signal acquisition techniques and machine learning. To date, the literature does not cover detecting specific attentional states using EEG and eye tracking data in sufficient detail. This gap is filled in this doctoral thesis by systematically exploring different attentional state classification approaches for both modalities. Previous research has suggested different machine learning algorithms and feature extraction approaches for similar cognitive state classification. These will be tested and compared for the presented use case.
Given the intended application of classification results for BCIs, the current state of research on EEG-based BCIs can be used to inspire the work on this thesis. However, there are no previous works that focus on adapting system behavior by adding attention-awareness for several different attentional states. The suggested closed-loop system requires several other gaps to be filled: Significant emphasis should be placed on the system's reliability, and efforts should be made to optimize the setup and the data obtained. The aspects of data generation and combination, as well as real-time components of the system, must be explored in greater detail in this thesis than it has been in the existing literature.
Finally, the transfer of results to augmented reality applications is a critical research question that needs to be addressed. The results of previous works need to be pursued further, moving from fixed, screen-based experimental setup to dynamic applications that require user movements and setups that allow recordings outside of a laboratory environment. As previously stated, augmented reality is an exciting technology that could benefit significantly from the proposed interface's increased usability.

In summary, the state-of-the-art in all of the previously mentioned areas is advanced enough to allow for the connection and investigation of all of these topics in greater detail. Passive BCIs, or eye tracking-based cognitive state detection, offer a suitable interaction method between users and head-mounted AR devices. They do not require additional explicit user input and therefor do not disrupt the usage of the application. The attentional state, in particular, is an important aspect to classify because the persuasive and immersive nature of the AR content increases the chance for distraction and a higher workload. Consequently, the presentation of content needs to be carefully designed and adapted to the user state.
In this thesis, I suggest an end-to-end mobile interaction system for augmented reality that is capable of adapting it's behavior and the UI to the user's attentional state based on EEG and eye tracking data. It fills a gap in the scientific literature by expanding our knowledge on biosignal-based attention classification and combining the results with efficient and reliable BCI setups for augmented reality use cases.

3 Approach, Methodology and Results

"The whole is greater than the sum of its parts"

- attributed to Aristotle

The field of augmented reality is in its infancy. Nonetheless, there are many incentives to use AR and many scientists and industry executives are convinced that it will have a significant impact on our future. While we are now in a state of continuous information availability as a result of the information age's increasing digitalization, we are making significant strides toward ubiquitous information presentation. To accomplish this, the interaction between humans and machines must be constantly rethought and developed, for instance using Brain-Computer Interfaces. The experience of the user can be improved if the system is aware of the user's cognitive state and the system's behavior changes accordingly. The dissertation's overall objective is to enhance the usability by adapting AR systems to the attentional state of the user.
The fundamental issue is increased sensory input as a result of the information overload and the resulting increased need for attention, as shown in Section 1.2.1. Attentional mechanisms are always used to categorize stimuli as significant or insignificant. Adding attention-awareness is inspired by human-human communication and I hypothesize that it will lead to improved usability of the AR system. For the suggested system, suitable human-machine interaction techniques should not require the active report of the attentional state but rather seamlessly deduct information passively provided by the user. Biosignal data, such as EEG and eye tracking, can be utilized as the input for machine learning algorithms to predict the current attentional state of the user. Based on the attention label, the behavior or displayed content of an AR system can be adjusted to reduce the distraction and increase the usability.

Several studies were conducted for this doctoral thesis to systematically examine the various components of attention-aware interaction systems for augmented reality. The findings were published in 15 peer-reviewed articles and preprints. Table 3.1 summarizes all of these works, as well as the journals and conferences that published them. All studies are included in **Appendix A Accumulated Publications** as an appendix to this dissertation. An overview of how they relate to each other can be seen in Figure 3.1. They will be related to the dissertation's research question and objective in the following in order to provide a final evaluation of such systems (see Chapter 4).
The central steps that inspired the individual study designs are as follows:

1. determining the basic **discriminability** of the attentional states that ought to be distinguished,
2. optimizing the machine learning procedures and setup in terms of **reliability and efficiency**, and
3. examining **usability** improvements following attentional-state adaptations in attention-aware closed-loop AR systems.

The performed studies all required attentional state modeling but will be organized according to their main aim in the following. Several hypotheses and research questions required the collection of new data,

Table 3.1: List of publications and preprints for this dissertation including title, year of publication, and journal or conference. * not first author

	Full Title	Year	Journal or conference
A.1	EEG-based classification of internally-and externally-directed attention in an augmented reality paradigm	2019	Frontiers in human neuroscience
A.2	Using Brain Activity Patterns to Differentiate Real and Virtual Attended Targets during Augmented Reality Scenarios	2021	mdpi Information
A.3*	Model-Based Prediction of Exogeneous and Endogeneous Attention Shifts During an Everyday Activity	2020	ACM ICMI
A.4	Differentiating Endogenous and Exogenous Attention Shifts Based on Fixation-Related Potentials	2022	ACM IUI
A.5	Exploration of Person-Independent BCIs for Internal and External Attention-Detection in Augmented Reality	2021	ACM IMWUT
A.6	SSVEP-Aided Recognition of Internally and Externally Directed Attention from Brain Activity	2021	IEEE SMC
A.7	Usability Examination of EEG Electrodes in Proximity to the Ears for SSVEP Studies		Preprint
A.8	Machine Learning from Mistakes: Self-Improving Attention Classifier using Error-Related Potentials		under review
A.9	Imaging Time Series of Eye Tracking Data to classify Attentional States	2021	Frontiers in Neuroscience
A.10	Combining Implicit and Explicit Feature Extraction for Eye Tracking: Attention Classification Using a Heterogeneous Input	2021	mdpi Sensors
A.11	Multimodal EEG and Eye Tracking Feature Fusion Approaches for Attention Classification in Hybrid BCIs	2022	Frontiers in Computer Science
A.12	Real-time multimodal classification of internal and external attention	2019	ACM ICMI
A.13*	Augmented reality interface for smart home control using SSVEP-BCI and eye gaze	2019	IEEE SMC
A.14	Attention-aware brain computer interface to avoid distractions in Augmented Reality	2020	CHI
A.15	Attention-aware Translation Application in Augmented Reality for Mobile Phones		under review

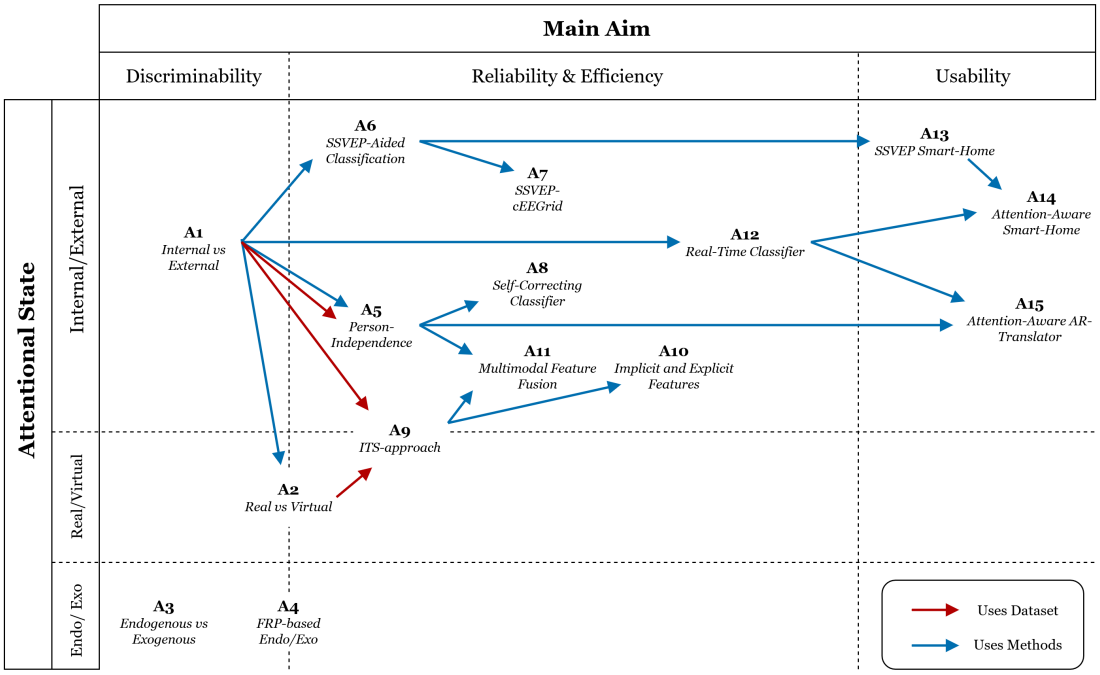


Figure 3.1: All studies performed for this thesis and how they relate to each other.

while other studies were based on existing datasets. For the usability examinations, end-to-end systems were implemented. Not all studies were performed in AR settings and some of the analyzed datasets were also not recorded using AR scenarios. The results achieved in the studies are attention and machine learning dependent and can be applied in many different scenarios; AR study results will likely hold for screen-based applications and vice versa. The main importance of the AR aspect is reflected in the end-to-end systems from the adaptational perspective.

3.1 Attentional State Discriminability

The dissertation's studies are all concerned with the detection of various attentional states. As discussed in Chapter 2.1, modeling numerous aspects of attention is possible. Due to the fact that head-mounted augmented reality displays primarily provide additional visual input and the system's adaptation is intended to be predominantly visual, the studies concentrated on visual attention. Three critical binary classification problems were identified for the adaptation goal discussed here.

The distinction between internally and externally directed attention is deemed to be the most useful categorization in this case. Internally directed attention is the deliberate suppression of sensory input in order to concentrate on thoughts and memories, for example. In this instance, external attention refers to a concentration on visual sensory input. The distinction between these two mutually exclusive states is substantial for this work because distraction can occur during periods of meaningful internally directed attention in response to salient visual sensory input. On the other hand, it may be necessary to intervene in cases of non-purposeful or goal-directed internal attention (e.g., mind wandering) and refocus the user's attention externally.
In the study **"Internal vs External" (A.1)**, we developed an AR paradigm that specifically requires the participant's internal and external attention. The typical AR elements and interactions have been included with great care. Participants are required to adjust their posture and head position in response to either a visual stimulus (external attention) or a pre-imagined stimulus during the experiment (internal attention). These two experiment conditions for the spatial alignment task serve as the ground truth, and thus the true labels for the EEG data that was recorded. The labeled EEG data was converted to continuous 13-second interval power spectral density features. Person-dependent LDA models were trained and tested on the features for the subsequent offline classification of the 14 participants' data. The average classification accuracy of approximately 85% demonstrated that discriminating between internally and externally directed attention is reliably possible in augmented reality scenarios. For three participants 100% accuracy was achieved.

Externally directed visual attention inspires the next step of determining the precise object in focus. While humans are capable of dividing their attention between multiple objects at the same time, the current focal point is most often identical to the center of gaze (overt attention). This can often be determined quite easily, particularly with the aid of eye tracking. However, specific conditions apply to augmented reality applications: Virtually generated content can appear or vanish at any time and for a brief period of time in specific locations. Additionally, the objects frequently exhibit a degree of transparency, which is a result of the projection technology. This could result in virtual objects obstructing the view of real objects. Thus, in addition to determining the current visual focus via eye tracking, it would be interesting to distinguish attention in augmented reality applications between real and virtual objects. This information could be used by the system to more deliberately reposition 3D content and possibly also to hide it in certain situations.
In the study **"Real vs Virtual" (A.2)**, attention was consciously directed toward either real or virtual playing cards in a PAIRS game. On a head-mounted display, participants were shown a game board and a few virtual accessories. In each trial, either all of the cards were real or all of the cards were replaced with virtual cards of identical design. Following that, the EEG and eye tracking data were analyzed using a neural network to test several hypotheses. Most importantly, it was discovered that the classification of "focus on real objects" and "focus on virtual objects" works with an accuracy of about 70% for person-specific EEG-based models. When both EEG and eye tracking data were used for classification, this accuracy could be increased to 77%. A cursory examination of person-independent trained models yielded simi-

larly optimistic results.

The third dimension of attention that is relevant for adaptation of AR systems is the nature of attentional shifts. More precisely, whether the new focus is the result of sensory or task-related attentional adaptation. If the readaptation of focus is motivated by the current task and thus reflects a top-down process (endogenous attention shift), the information in focus appears to be beneficial and interesting. On the other hand, if an attentional shift is "forced" (exogenous, bottom-up) by sensory input, it must be determined whether it is an unintentional distraction or a system-initiated attentional shift. This information about the motivations for attention shifts can assist the augmented reality system in learning from distracting inputs and adapting their display strategies to accommodate new information for the user.
The work published in the study **"Endogenous vs Exogenous" (A.3)** collected user data (video and audio) in a dynamic everyday environment. Participants were required to complete various table setting tasks. These data were classified according to exogenous and endogenous attentional shifts based on the current task and controlled environmental cues. On this basis, a model was developed that uses a combination of top-down and bottom-up models to predict and classify future focus shifts.
In a further study on this aspect of attention, the focus was more on the subsequent detection of endogenous and exogenous attentional shifts. Since visual attention, as described, is often represented by eye movements, fixation-related potentials in EEG data were used as classification input in the study **"FRP-based Endo/Exo" (A.4)**. In the experiment, participants had to solve tasks displayed on a computer screen that required them to direct visual attention to several areas in succession. In between, prominent stimuli were displayed repeatedly to divert the participants' attention. We recorded both EEG and eye tracking data. Labeled 0.7 s windows were cut out based on fixations on the task or a distractor. Again, using person-dependent classification with LDA models demonstrated that machine learning was capable of classifying data with greater accuracy than chance (approximately 60%). Notably, person-independent trained models performed only slightly worse but also achieved an accuracy of approximately 58.5%.

Overall, these studies largely confirmed the fundamental concept of classifying various aspects of attention. While all of these mentioned aspects can be considered useful for the adaptation of an interaction system in AR, the focus of follow-up studies was mainly on internally and externally directed attention. Specifically regarding distraction avoidance, I postulate that this distinction has the highest value for usability improvements. It can prevent distracting system behavior, while the distinction of real and virtual attention targets or endogenous and exogenous attention shifts can be used for the analysis of the current visual input and whether it needs to be adjusted. An additional advantage is that more related work has been performed on the subject of internally and externally directed attention yielding a larger quantity of available datasets for further analysis. The most significant challenge was designing the data collection studies, as attention is the result of the interaction of numerous factors; making it difficult to obtain labeled data with low label noise, which is required for the supervised learning procedures used here.
Further studies were conducted to optimize classification accuracy because there is still considerable room for improvement. Also, other necessary aspects of the proposed end-to-end system have not been of interest in these studies, but have been investigated in the following studies that used the aforementioned acquired datasets in part. Table 3.2 summarizes the major contributions of the four studies presented so far.

Table 3.2: Major contributions of the studies for attentional state discriminability

Study	Contributions
A.1	- internal/external attention in AR paradigm and dataset
	- successful internal/external attention classification in AR
A.2	- real/virtual attention in AR paradigm and dataset
	- success internal/external attention classification in AR
A.3	- model for endogenous and exogenous attention shifts
A.4	- endogenous/exogenous attention shifts paradigm and dataset
	- successful endogenous/exogenous attention shift classification

3.2 Reliability and Efficiency

When optimizing the proposed classifiers for the proposed overall system, it is necessary to consider more than classification accuracy. The reliability of the system is additionally determined by the temporal resolution and the robustness of the results for individual users, different settings, and manifold task variations. Not only does the system require high-quality data from appropriate biosignals, but it also requires a real-time component for a high temporal resolution. Due to the fact that the state of attention can and does change rapidly, it is critical to maintain a very regular update of the current prediction. This should be based on a brief time interval and made readily available.
On top of that, the system's configuration and operation are critical components for the final improvement. The efficiency of the required setup, and the efficiency with which the available data is used, need to be maximized. For instance, if a long calibration and training phase is necessary before every use to make the AR system attention-aware, an attention-unaware system might be more time-efficient. Ease of use is facilitated not only by the general setup, but also by the fact that no additional time is required in comparison to systems lacking attentional state information about the user. However, a more elaborate, time-consuming setup for data recordings frequently improves data quality, resulting in a higher reliability. Thus, when considering the optimization strategies discussed here, the trade-off between accuracy and simplicity must always be considered.
In summary, the objective of all optimizations is to create a classifier that is mobile, light-weight, uncomplicated, time-efficient, and reliable in prediction and labeling.

Numerous aspects of the problem that can be solved using machine learning were examined in greater detail in the study **"Person-Independence" (A.5)**. The study's analyses were conducted on the dataset recorded in **"Internal vs External" (A.1)**. The primary objective was to enhance person-independent classification in order to eliminate the need for recording individual training data in a possible end-to-end system. If the user must spend time calibrating the system each time before using it for its intended purpose, usability is significantly reduced because it is time-inefficient. Additionally, this study compared different machine learning algorithms and weighed different EEG data window lengths. As a result, we discovered that the highest classification accuracy on the data can be achieved without using individual-specific training when 4-second long data windows are classified using a shallow CNN (up to 88% accuracy).
In the study **"SSVEP-Aided Classification" (A.6)**, an alternative strategy for reducing or eliminating the classifier's training time prior to use was to incorporate SSVEP stimuli to discriminate between internal and external attention. The study's premise was that when attention is internally directed at the time of presentation, the neural response to the visual stimulus is predictably weaker. The classification accuracy was significantly improved by calculating SSVEP-specific metrics as additional features (in addition to general EEG-based features) for a classifier. However, the SSVEP metrics were insufficient on their own to produce comparable results. Even more so when the data was analyzed across participants. Neural SSVEP-correlates in the brain are strongest in the visual cortex and can thus be detected best using occipital EEG electrodes. Unless those are built-in into the head-mounted AR display, the two setups are likely to interfere reducing data quality and setup comfort. In the pilot study **"SSVEP-cEEGrid" (A.7)**, the reliability of SSVEP detection was examined for so-called cEEGrid electrode clusters. The EEG-electrodes are placed around each ear using an uncomplicated and light-weight setup that is not influenced by head-mounted displays. The systematic comparison of single electrodes and electrode clusters from a high density EEG recording showed that the SSVEP coefficients were significantly less reliable than for occipetal electrodes, but a classification was still possible above chance. One main advantage of SSVEP signals is that they are relatively stable across participants whereas a big disadvantage is the flicker-

ing nature of stimuli. The combination of cEEGrid electrodes, SSVEP-stimuli and head-mounted AR devices should be considered for some specific task setting, but are not suitable for general attention-aware AR systems.

Rather than being forced to choose between a person-dependent, training-intensive classifier and a person-independent classifier with low accuracy for the EEG data, the study **"Self-Correcting Classifier" (A.8)** investigated a middle ground. The study's objective was to gradually personalize a system-independent classifier during active use, thereby increasing its accuracy. This was accomplished by training an error-related potential classifier, which then detected errors in the actual classifier based on the participants' response to the attention-specific feedback. Thus, the overall system learned from its own errors by independently enabling and adapting the person-independent classifier. The results indicate that active use of the system significantly improves the detection of internal and external attention. As a result, the proposed system may provide an alternative to the trade-off between training time and classification accuracy.

In general, it is more complicated and time consuming to use an EEG system than it is to use an eye tracker. Therefore, if equivalent accuracy in determining attention can be achieved solely through eye movement classifications, the setup is significantly simplified. Particularly, given that new head-mounted augmented reality displays, such as the HoloLens 2, already include this functionality.
While eye tracking data was analyzed in the aforementioned studies, it was always inferior to EEG data. A limited number of EEG and eye tracking co-registration datasets are available for attentional state recognition. Thus, we tested an innovative feature generation method in the study **"ITS-approach" (A.9)**, focusing only on improving the classification performance of eye tracking data. Three distinct algorithms were used to convert time series eye tracking data from three distinct attention studies (including **A.1** and **A.2**) to images (Imaging Time Series; ITS). The attentional states were then classified using a CNN with the images as features. We demonstrated that this method outperforms conventional methods for eye tracking classification by a significant margin. This was true for both person-dependent and person-independent models. One main difference is that the ITS-features implicitly represent the time series and the network has to learn meaningful units itself, whereas the conventional methods extract explicit oculometric descriptions of the gaze behavior.
The classification accuracy was further improved in the work **"Implicit and Explicit Features" (A.10)** by combining the novel image features from **"ITS-approach" (A.9)** with classical statistical eye movement features in a heterogeneous feature set. Surprisingly, person-independent (and thus training-free) classification outperforms personalized models on these features. The dataset was created by another research group and included 154 participants. All analyses were conducted for 3 and 8 second windows and resulted in an average of 71.7% and 76.7% for leave-one-person-out training and testing procedures, respectively. Unfortunately, because no EEG data are available, no direct comparison is possible, but it suggests that eye tracking recordings alone may be sufficient to distinguish between underlying internal and external attention.
In the study **"Multimodal Feature Fusion" (A.11)**, a multimodal approach was used to incorporate EEG and eye tracking data from a co-registration study during the classification process. To be precise, the objective was to compare three distinct methods of feature fusion at different stages of the classification process: early fusion at the feature level, middle fusion at the decision level, and late fusion at the decision level. Previously published BCI studies that were based on these multimodal data used either early or late fusion and rarely compared the two. The middle fusion approach leverages the adaptability of neural network structures to combine the best characteristics of both approaches. The results indicate that while middle and late feature fusion have a slight advantage over early feature fusion, none of the

Table 3.3: Major contributions of the studies for more reliability and efficiency

Study	Contributions
A.5	- person-independent internal/external attention classification - systematic classification algorithm comparison - EEG window length analysis for real-time systems
A.6	- investigation of improved internal/external attention classification using SSVEP - internal/external attention with SSVEP stimuli dataset recording
A.7	- systematic analysis of EEG electrode positions for attention classification with focus on electrodes in proximity to the ears
A.8	- internal/external attention classifier with feedback paradigm and dataset - Error-related potential detection during different attentional states - online classifier retraining for personalization of person-independent classifier - online closed-loop feedback system
A.9	- novel method for eye tracking feature generation - classification performance evaluation for several attention datasets compared to state-of-the-art features
A.10	- systematic comparison of implicit and explicit feature sets for eye tracking based attention classification
A.11	- systematic comparison of feature fusion strategies for multimodal attention classification
A.12	- real-time multimodal internal/external attention classification system and dataset

three approaches can be ruled out completely. In general, it can be concluded that using multimodal datasets outperforms using single modalities.

In the study **"Real-Time Classifier" (A.12)**, a real-time classification system for internal and external attention was demonstrated and tested. The classification is based on 1.5-second time windows of EEG and eye tracking data, which are combined in a late fusion approach to determine individual classification.

Overall, these optimization studies indicate that numerous parameters can and must be adjusted to improve the system. Many of these changes, however, are not only user-dependent, but also strongly related to the application area for which the system is being used. Though not all data in these studies were collected using AR, it can be assumed that these broad statements about the classification of attentional states apply to tasks with a variety of attentional aspects and presentation media.
While multimodal datasets improve classification accuracy, they frequently complicate the setup. Typically, the best results are obtained using personal or adaptable neural networks that receive EEG data as input. However, eye tracking data performs better in terms of person-independence. The majority of the studies cited here considered the real-time aspect by keeping the analyzed time windows as brief as possible and comparing different lengths. In summary, a few seconds (approximately between 3 and 8) offer the optimal trade-off between accuracy and timeliness. Regarding the chosen classification algorithms, our results so far suggest that they only play a minor importance for the performance as long as they are well engineered and the data quality and preprocessing is appropriate.
The contribution of each study presented here is summarized in Table 3.3.

3.3 Usability

As previously stated, there are numerous unique aspects to the implementation of an attention-aware interaction system in augmented reality. Modalities, setup, and machine learning classifiers can all be adapted to fit the application area. To enable meaningful interface and behavior adaptations, classification of multiple attention aspects in the same system, in particular, would require a very specific use case. The focus of the end-to-end systems in this thesis was instead on mobility and ease of setup, on the one hand, and implementation strategies for an exemplary real-time system, on the other. The systems were evaluated primarily on the basis of their usability in comparison to a control system that did not include attention sensitivity.

To begin, in the study **"SSVEP Smart-Home" (A.13)**, an augmented reality-based smart-home system was developed in which certain controls are selected using one's attention and decoded based on SSVEP signals and eye tracking. The evaluation of this active attention-based BCI revealed that the system operated reliably and that users rated it as simple to use.
Following that, in the study **"Attention-Aware Smart-Home" (A.14)**, the results of **"Real-Time Classifier" (A.12)** and **"SSVEP Smart-Home" (A.13)** were combined to create an attention-aware smart-home system that incorporates both, an active and a passive BCI for attention detection. The attentional state insensitive system displays the smart-home's flickering buttons whenever they appear within the range of the head-mounted display's world camera. However, one can imagine numerous scenarios in which the head is moved without the intention of operating the system. In the worst-case scenario, one is even internally focused and simply allows their gaze to wander. It is extremely distracting when several flickering buttons appear in the center of the field of view at such a moment. Thus, the attention-sensitive system adapts its behavior in response to the current user state: When the user is internally focused, virtual buttons are not displayed.
Although classification accuracy was not as high as it was in the offline analyses described previously in this thesis, user ratings indicated that the attention-aware system was preferred. They gave a higher rating to usability and a lower rating to distractions. This effect may be enhanced with increased classification accuracy (for example, using the results from the aforementioned studies).
However, the most common criticism leveled at this system was its laborious calibration and uncomfortably awkward setup.

As a solution to that, the primary focus of a second end-to-end system was on a user-friendly setup; once again, with the goal of integrating internal and external attention-awareness into the system.
The system described in **Study "Attention-Aware AR-Translator" (A.15)** projects the augmented reality elements using a smartphone, and the EEG data is collected using a lightweight consumer-grade headset. All processing steps take place on smartphones, which means they can be used at any time and from any location. The EEG cap is self-contained and does not require the application of electrode gel. The system's objective was to enhance an augmented reality-based translator app that projects translated text as 3D elements onto the screen. While the setup was well-suited and the classification worked satisfactorily, the adaptation mechanisms chosen were found to vary in their comfort level depending on the individual. While some users appreciated the intended behavior, others desired the exact opposite. This effect demonstrates yet again how the implementation of an attention-aware interaction system is context-, user- and application-specific.
The main contributions of the three presented studies focusing on system usability are summarized in Table 3.4.

Table 3.4: Major contributions of the studies focusing on usability examinations

Study	Contributions
A.13	- end-to-end SSVEP based smart-home system implementation - assessment of system control based on attention shifts
A.14	- added real-time attention-awareness for the smart-home system - usability comparison of attention-aware and -unaware systems
A.15	- end-to-end smartphone-based AR-BCI system implementation - system behavior adaptation based on real-time attention classification - light-weight consumer grade BCI setup - usability comparison of attention-aware and -unaware systems

4 Conclusion

AR may be in its infancy but it is here to stay.
As argued in this thesis, head-mounted displays in particular rapidly increase the user's workload as a result of the abundance of additional information. The disadvantages and distractions that accompany them should be avoided to the greatest extent possible. The optimal way to accomplish this is for the system to adapt to the user's attentional state without the user actively communicating it. This paves the way for more natural human-machine interaction. In addition to previous considerations and related work, the dissertation's studies examined numerous different aspects and all necessary components for implementing attention-aware interaction systems for augmented reality. Most importantly, the studies have confirmed the two hypotheses: (1) classification of various important aspects of human attention can be accomplished using available biosignals; and (2) the adaptation of the system to the attentional state can result in an increase in the system's usability. Several significant aspects of this topic could be explored for the first time through the sequential work performed here. That includes the use of EEG and eye tracking for this purpose, possible approaches for training-free classifiers, real-time systems, and finally end-to-end systems with different setups.
However, the breadth of possible applications combined with the diversity of machine learning techniques and recording devices enables an infinite number of additional studies. This chapter will cover the follow-up studies and possible future directions, as well as shortcomings and ethical considerations.

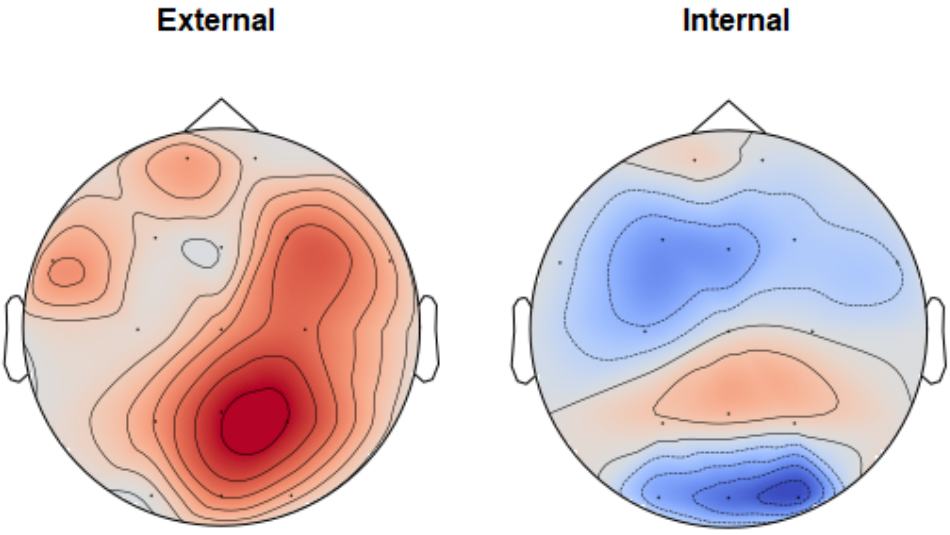


Figure 4.1: Relevances generated using the LRP averaged over time and all trials for each class. red = positive relevance, blue = negative relevance

4.1 Late-Breaking Work

The papers and preprints presented in this doctoral thesis already inspired follow-up studies that will extend the current results. During the time of the preparation of this thesis, they were in different stages of experiment design, data collection and analysis and have not yet been finalized but might lead to additional publications affiliated to this Ph.D. project.

A majority of research so far has concentrated on the distinction between internal and external attention. This information is critical for avoiding distracting virtual content display.
Shallow CNNs appear to be the most promising solution for this classification problem using EEG data. When neural networks are used, a central question is frequently which features of the data are ultimately necessary for classifiability. In contrast to approaches such as LDA, no prior features have been generated that can be used to infer neural processes. Due to the lack of transparency in the network's decision-making, it is difficult to verify model generalizability and thus prevents further insight into neurophysiological phenomena. Nevertheless, such information would aid in the further development of neural networks and their adaptation to this data. For instance, it may be beneficial to ensure that the classification is not too strongly influenced by task-related noise. To address this black box problem, an ongoing work is currently experimenting with Explainable AI techniques. The purpose of this study is to examine the models developed in **"Person-Independence" (A.5)** via layerwise-relevance propagation (LRP). With the aid of LRP, the network's decisions regarding electrode positions on the scalp are analyzed and visualized as an interpretable relevance heatmap. More precisely, the relevance of each neuron to the prediction is backpropagated through the network using appropriate propagation rules, layer by layer, until it reaches the input features, which are easily interpretable in the majority of applications. This corresponds to the propagation of the classification result back to the electrode positions and points in time in this use case. This is referred to as a "deep Taylor decomposition" [136]. Each layer's relevance imposes a conservation rule, such that the total sum of the layers' relevance is constant. So far, LRP has been shown to highlight neurophysiologically plausible activation patterns for attention tasks in the

Figure 4.2: AR-museum example. The virtual information dashboard covers the real painting in part.

participant with the best classification performance (see Figure 4.1). The study's final results will shine a light on which configurations of the neural networks are responsible for higher or lower classification performances and which aspects of the data can be deemed useful in the classification process. The additional understanding of the models can be used to improve and further engineer neural networks for EEG-based classification.

Another follow-up study, which is still in progress, is based on the idea of **"Virtual vs. Real." (A.2)**. When overlapping objects are involved, the utility of a classifier of attention to virtual or real objects is enhanced further. Consider a scenario in which a hologram obscures a real object that is of interest to the user at the time. The slight transparency of holograms, which is inherent in the technology, tempts users to simply "look through" the virtual objects in these instances. In such cases, the direction would be indistinguishable from pure gaze. This classification problem will be tackled in the study by combining EEG and eye tracking data. As a suitable use case, an augmented reality-enabled museum tour was selected. Real paintings are accompanied by virtual information panels (see Figure 4.2). The participants are then required to complete tasks that require them to either look at the information on the panels or at the entire image. Following that, the recorded data will be analyzed offline. If the classification can be made reliably, the next step would be to incorporate it into an online system. In this case, if conflicting states are detected, the virtual objects should be hidden or repositioned.

The presented study **"SSVEP-cEEGrid" (A.7)** was a preliminary analysis for the possible combination of cEEGrid electrodes for an AR-based SSVEP study, such as the attention-aware smart-home system from **"Attention-Aware Smart-Home" (A.14)**. In an ongoing work, these electrodes were used for screen-based recordings of SSVEP signals, and the compatibility with an HMD was tested (see Figure 4.3. If the classification accuracy using the simplified setup is reliable, the usability of the suggested smart-home system and other attention-aware SSVEP-based systems would highly increase. Future follow-up studies using the cEEGrid are planned to involve attentional state detection without SSVEP stimuli to determine their feasibility.

Figure 4.3: Setup of the cEEGrid electrodes and the Microsoft HoloLens 2. The face mask was mandatory due to Covid-19 regulations at the time of the study.

In a further study on "Webcam-Based Classification", the method described in **"ITS-approach" (A.9)** was applied to webcam-captured eye gaze data. It was discovered that the data quality plays a critical role in classification accuracy. Although the positive results of **"ITS-approach" (A.9)** could not be replicated, several additional factors affecting these values were already identified. For the purposes of this thesis's overall question, this means that both the quality of the eye tracker (i.e. sampling frequency, lighting conditions) and the camera positioning play a critical role in the system's design. Once all confounding factors for the classification accuracy are identified, attentional state detection using webcam recordings during online tasks or experiments can be evaluated, for instance for online learning scenarios.

4.2 Future Work

The presented publications of this thesis and ongoing follow-up research works have framed a good understanding of feasible future projects that would help developing attention-aware interaction systems for AR.

One aspect that has not yet been evaluated in a real-time system is discriminating between endogenous and exogenous attentional shifts. User-specific profiles could be created to describe the properties of virtual stimuli to which a user is particularly sensitive. Precisely, the attention-shifts and other attentional state predictions could be recorded for newly presented information in an AR application to calculate person-specific values and perception thresholds. The system could then adapt actively and continuously to how virtual information is presented, depending on the task.
By adding a classifier that distinguishes between real and virtual targets of attention, the analysis would become even more holistic and interesting lines of research could evolve. In general, additional paradigms for evaluating the usability of real-time adapting end-to-end systems could be investigated.

In following research works regarding all of these attentional states, not only the accuracy of classification per individual should be optimized. The greater the degree to which algorithms produce person-independent results, the better. Additionally, in the future, a greater emphasis should be placed on the algorithms' efficiency in terms of memory usage and computing time. Computing power should be provided by the augmented reality device itself, rather than by separate devices that perform the classification, as is currently the case in the reported studies. This compartmentalization of the system impairs its usability in a wide variety of mobile scenarios.

Similarly, future work should optimize the placement of the EEG electrodes. This dissertation has already examined alternative recording devices such as the cEEGrid **(A.7)** and the Muse headset **(A.15)**. Another possibility is to conduct the EEG directly over dry electrodes attached to the augmented reality headset. Microsoft, for example, relies on a design for its Hololens that allows for the placement of electrodes beneath the headband in both the frontal and occipital regions. According to previous neuropsychological findings, as discussed in Section 2.1, these positions would be ideal for the attentional aspects discussed here. Thus, in the long run, prototypes of the presented end-to-end system could be created that record, process, and playback entirely on a single device.
Only if it is possible to record EEG data without additional effort when using the AR system can an increase in general usability be expected.

Following the results for the chosen biosignal modalities, the classification process needs to be further optimized by combining EEG and eye tracking data as efficiently as possible. Thus far, the results have not indicated a clear preference for a particular approach for feature fusion but it has been shown that their combination outperforms unimodal approaches. Further research should consider advanced machine learning techniques, such as a neural network with an attention layer, to replace the decision rule. Ultimately, several attention classifiers could be combined to create a system that does not require user-specific training and makes predictions based on short data intervals. The system could track the attentional direction and analyse attention shifts to adapt the behavior and information presentation. The setup should be light-weight and allow for testing under real-world conditions, for instance in the use case scenario of a factory worker.

4.3 Critical Discussion

The studies presented here support the two hypotheses presented at the beginning: attentional classification is possible, and attentional adaptation improves usability of AR systems. Additionally, the feasibility of such an end-to-end system was demonstrated using either a smartphone or head-mounted display. However, as evidenced by the considerations for additional work, there are still some areas that require substantial improvements. On top of that, the advancement of such technologies always entails an ethical obligation, which needs be discussed. One of the main issues is data protection.

4.3.1 Current Shortcomings

In any case, the primary criticism leveled at previous studies is the setup's complexity in comparison to the mediocre classification accuracy. The combination of the AR device and an additional EEG cap is incompatible with user-oriented development, as the time required to prepare for use is simply too long. At the moment, calibrating the device requires the help of trained personnel. The use of alternative EEG headsets degrades data quality and availability, and thus degrades classification accuracy. When it comes to convenience in general, relying solely on eye tracking data for classification would undoubtedly be more user-friendly.
However, there are some drawbacks, including the effects of changing lighting conditions, the lower performance of people who wear glasses or contact lenses, the extreme reliance of eye movements on the current task, and the instability of the cameras pointed at the eyes during movements.

Motion artifacts, in general, are an issue that requires additional attention in studies conducted under increasingly realistic conditions (e.g. during less controlled movements, for longer use times). Multimodal data are the most promising for the intended use because they can compensate for each other's shortcomings. For those purposes, the technological implementation of data collection must be enhanced before considering a real-world application. If the data can be used to improve usability, but its collection reduces comfort and the amount of time one can work with the system without experiencing pain, there can be no generalized increase in usability. Thus, it is unclear which way the trade-off between system quality and compressibility will shift. Currently, the first companies are investing in AR and VR headsets with inbuilt EEG-electrodes (e.g. DSI-VR300 by Neurospec AG [48]) which would facilitate the setup immensely.
Although usability studies have demonstrated that improvements are possible even with mediocre classification accuracy, these evaluations did not take the complexity of the setup into account. To be considered useful, the benefit of attention sensitivity must significantly outweigh the disadvantage of additional biosignal data collection. The same holds true for the additional time required for individual calibration. The importance of training-free solutions has been emphasized previously, but should be reiterated here. Without the ability to develop reliable person-independent real-time classifiers based on a convenient setup, usability improvements are severely limited.

Another finding that raises concerns is the high difference in preference for interface adaptations between users. Apparently, as mainly study **"AR translate" (A.15)** has shown, users can have quite opposing views when it comes to the usefulness of attention-aware mechanisms. While some preferred that the system paused during internally directed attention, others preferred the behavior that was based on wrong classification, namely when it paused during externally directed attention. Such preferences could be due to a person's ability to inhibit external stimuli during internally directed attention or the speed with which people grasp their surroundings and are able to comprehend new stimuli. In this work, it

was already suggested that the attention classification also needs to be performed to recognize a person's attentional capabilities in retrospective to displayed content (by classifying attention shifts and attention on real and virtual objects). Within the scope of this thesis it was not yet possible to determine in how far flexible, personalized UI adaptations are feasible. A framework that allows momentary attentional-state adjustments based on user-dependent attentional capabilities and preferences (which themselves would have to be readjusted constantly based on the user's reaction to the input) could easily reach a level of complexity that makes the system's behavior increasingly unpredictable. Such unpredictability, however, requires thorough further considerations, ethically and legally.

4.3.2 Ethical Considerations

New technologies always raise ethical and legal concerns, as they come with a great deal of responsibility. This is true for both medical applications and mass market technologies, which are the issue in question. In this case, the immersiveness and general ability to use the system, in particular, require closer examination from a moral standpoint.

Due to the fact that the systems presented are intended to be as persuasive and ubiquitous as possible, the effect of frequent use on the user must be taken into account. The considerations in this thesis are intended to advance the cause of attention-responsive technology. This could result in users being more and more overly reliant on the system and ceasing to monitor their own attention state which becomes especially dangerous if classification errors occur.
For instance, when operating certain machines, if systems are designed to be turned off automatically whenever the operator's attention wanders - and the operator trusts that this will happen - accidents may occur. It is possible to lose one's intuition for one's own state of attention.
Another issue that needs to be considered is accessibility. If one assumes that augmented reality-based systems will become "normal" in many areas in the future, or even necessary for certain professions, for example, a great deal of emphasis needs to be placed on ensuring that their use does not exclude people. It has already been discussed how certain personal characteristics can affect the way eye movements are recorded (glasses, contact lenses, possibly eye position, etc).
When BCIs are used, as they are here with EEG, similar usage restrictions may apply, but these are not well researched yet. The literature frequently mentions what is referred to as BCI illiteracy. It is suggested that BCIs may not be effective for approximately 20% of the population [25, 45]. While delving into the details of this is beyond the scope of this thesis, the implications for using the described system are critical. If BCI-based augmented reality systems become widely adopted but are unable to be used effectively by some people, ethical implications must be considered [37].
For many years, neuroethicists have debated the ethical implications of BCIs [69, 73]. It is repeatedly stated that a distinction must be made between medical and research systems, as well as between industrially motivated systems. A key difference is, among other things, the purpose for which the user data is collected and whether this justifies the invasion of privacy.

Data Protection

The supposedly most serious ethical issue is data privacy and data misuse prevention [37, 71, 141, 153]. Even though we cannot yet speak directly of "mind reading," information about our brain activity is probably the most intimate and private information we can reveal about ourselves. These concerns about privacy and agency have long been a source of concern for developers in a variety of fields. Hackers may steal digitally stored neural data or it may be used inappropriately by companies to which users grant

access.

Thus, one must be concerned about the possibility of an attention detection system being abused to spy on the user. This level of monitoring and observation by companies, the employer or even the state can easily resemble a modern version of Jeremy Bentham's panopticon where no one ever knows if they are watched or not [122]. Bentham believed that through this seemingly constant surveillance, all groups of society could be altered. Morals would be reformed, health would be preserved, industry would be revitalized, and so forth. However, the power of constant surveillance comes with great responsibility [20]. Michel Foucault, a French philosopher and big critic of the panopticon contended that the panopticon's ultimate objective is to produce a state of conscious visibility in the inmates to ensure optimal behavior [122]. Constant surveillance – or the concept of constant surveillance – regulates even the smallest minutiae of everyday life. This is what Foucault refers to as a "discipline blockade." Shoshanna Zuboff, a philosopher and psychologist, emphasizes what she refers to as "surveillance capitalism" [201]. She argues that, in addition to the disciplinary purposes, it can also be employed for marketing purposes. She defined the modern personal computer's duty as a "information panopticon" capable of monitoring an individual's work output. This becomes more applicable with more immersive and ubiquitous technology such as AR. The data collection bears a strong resemblance to the panopticon, as it is a one-way information conduit.

Similar concerns are frequently raised when other biosignal data is collected, for instance by fitness trackers. With access to the data, insurance companies and employers may be able to negotiate their way out of legal disputes.
On the other hand, access to the system or personalized user analyses could be used to manipulate the user. Personalization of advertising is a common example of such manipulation based on user data. While one can debate the potential benefits in advertisement, manipulating the user's state of attention would have serious consequences and give hackers considerable control over the user.

Neuroethicists have compelled developers to prioritize device security, to safeguard consumer data more diligently, and to stop requiring access to social media profiles and other sources of personal information as a condition of device use [37]. Nonetheless, as consumer neurotechnology gains traction, ensuring acceptable privacy standards continues to be a challenge.

4.4 Summary

Attention detection, adaptation, and regulation are all critical concepts in interaction, not only between humans but also between humans and machines; particularly as the degree of immersion increases and the desire for intuitive interaction increases. Augmented reality is an extremely promising technology in which significant investment has already been made and which will continue to grow in importance over the next few years. Optimizing this technology's usability in order to maximize its benefits and avoid potential problems in human-machine interaction should begin as soon as possible so that the technology can grow from it and directly establish certain standards.

The research and related work presented in this thesis demonstrates that both EEG and eye tracking data collected from users can be used to adapt an augmented reality system to the user's state. Both have their benefits and drawbacks in terms of processing and recording. A combination of the two modalities appears reasonable, and neural networks, in particular, offer numerous opportunities for further engineering of the classification process. While classification has been demonstrated to be feasible, much work will be required to optimize the predictions' robustness and efficiency. Particular emphasis should be placed on the avoidance of explicit user-dependent training phases for classifiers and the real-time aspect. Apart from the algorithmic optimization, the technological aspect requires further development in a setup-centered manner.

Based on the presented findings, making broad statements about the actual adaptation procedures is exceedingly difficult, as each application and user appear to have highly unique requirements that must be re-evaluated for new use cases. The primary focus should always be on the system's error-free usability, even if the user state detection does not work reliably.

Apart from the system's incorrect behavior, the most serious consequence for the user would be a misuse of the data collected during the process. The proposed system's ethical and legal implications necessitate a particularly careful design, the complexity of which does not appear to be fully understood.

Although much work remains to be done before this technology can be widely deployed, the results presented here provide an optimistic outlook on its feasibility. I was able to fill a gap in the literature that combines the fields of human-machine interaction, cognitive science, machine learning, BCIs, eye tracking and AR and proposes an exciting technological advancement in the future.

Acronyms

AI	artificial intelligence
ANOVA	analysis of relevance
AR	augmented reality
AV	augmented virtuality
BCI	brain-computer interface
CNN	convolutional neural network
DAN	dorsal attention network
ECoG	electrocorticograph
EEG	electroencephalogram
EFRP	eye-fixation-related potential
EOG	electrooculogram
EP	evoked potential
ERP	event-related potential
FEF	frontal eye fields
fMRI	functional magnetic resonance imaging
fNIRS	functional near-infrared spectroscopy
FRP	fixation-related potential
HMD	head-mounted display
HUD	heads-up display
IFG	inferior frontal gyrus
IPL	inferior parietal lobe
IPS	inferior parietal sulcus
ITS	imaging time series
LDA	linear discriminant analysis
LRP	layerwise-relevance propagation
LSTM	long short-term memory
MFG	middle frontal gyrus
ML	machine learning
OMT	ocular mircotremor
OST-HMD	optical see-through head-mounted display
PCA	principle component analysis
ReLU	rectified linear units
SEP	sensory evoked potential
SPL	superior parietal lobe
SSEP	steady-state evoked potential
SSVEP	steady-state visually evoked potential
STG	superior temporal gyrus
TPJ	temporo-parietal junction
UI	user interface
VAN	ventral attention network
VR	virtual reality

Bibliography

[1] M. Abo-Zahhad, S. M. Ahmed, and S. N. Abbas. "A New EEG Acquisition Protocol for Biometric Identification Using Eye Blinking Signals". In: *International Journal of Intelligent Systems and Applications* 7.6 (2015), pp. 48–54.

[2] F. Aboitiz, T. Ossandón, F. Zamorano, B. Palma, and X. Carrasco. "Irrelevant stimulus processing in ADHD: catecholamine dynamics and attentional networks". In: *Frontiers in psychology* 5 (2014), p. 183.

[3] D. O. Adewole, M. D. Serruya, J. P. Harris, J. C. Burrell, D. Petrov, H. I. Chen, J. A. Wolf, and D. K. Cullen. "The evolution of neuroprosthetic interfaces". In: *Critical Reviews™ in Biomedical Engineering* 44.1-2 (2016), pp. 123–152.

[4] E. Ahissar, A. Arieli, M. Fried, and Y. Bonneh. "On the possible roles of microsaccades and drifts in visual perception". In: *Vision research* 118 (2016), pp. 25–30.

[5] S. Aliakbaryhosseinabadi, E. N. Kamavuako, N. Jiang, D. Farina, and N. Mrachacz-Kersting. "Online adaptive synchronous bci system with attention variations". In: *Brain-Computer Interface Research* (2019), pp. 31–41.

[6] E. B. Anderson, J. F. Mitchell, and J. H. Reynolds. "Attention-dependent reductions in burstiness and action-potential height in macaque area V4". In: *Nature neuroscience* 16.8 (2013), pp. 1125–1131.

[7] J. R. Anderson, D. Bothell, and S. Douglass. "Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis". In: *Psychological Science* 15.4 (2004), pp. 225–231.

[8] C. Andreu-Sánchez, M. Á. Martín-Pascual, A. Gruart, and J. M. Delgado-García. "Viewers Change Eye-Blink Rate by Predicting Narrative Content". In: *Brain Sciences* 11.4 (2021), pp. 1–10.

[9] S. Annerer-Walcher, S. M. Ceh, F. Putze, M. Kampen, C. Körner, and M. Benedek. "How Reliably Do Eye Parameters Indicate Internal Versus External Attentional Focus?" In: *Cognitive Science* 45.4 (2021), e12977.

[10] K. Anton-Erxleben and M. Carrasco. "Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence". In: *Nature Reviews Neuroscience* 14.3 (2013), pp. 188–200.

[11] P. Aricò, G. Borghini, G. Di Flumeri, N. Sciaraffa, and F. Babiloni. "Passive BCI beyond the lab: current trends and future directions". In: *Physiological measurement* 39.8 (2018), 08TR02.

[12] M. Arvaneh, I. H. Robertson, and T. E. Ward. "A P300-Based Brain-Computer Interface for Improving Attention". In: *Frontiers in Human Neuroscience* 12 (2019), p. 524.

[13] M. Azmandian, M. Hancock, H. Benko, E. Ofek, and A. D. Wilson. "Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences". In: *Proceedings of the CHI conference on human factors in computing systems*. San Jose, CA, USA: Association for Computing Machinery, 2016, pp. 1968–1979.

[14] S. Balakrishnama and A. Ganapathiraju. "Linear discriminant analysis-a brief tutorial". In: *Institute for Signal and information Processing* 18 (1998), pp. 1–8.

[15] F. Behrens, M. MacKeben, and W. Schröder-Preikschat. "An improved algorithm for automatic detection of saccades in eye movement data and for calculating saccade parameters". In: *Behavior research methods* 42.3 (2010), pp. 701–708.

[16] M. Benedek, E. Jauk, R. E. Beaty, A. Fink, K. Koschutnig, and A. C. Neubauer. "Brain mechanisms associated with internally directed attention and self-generated thought". In: *Scientific reports* 6.1 (2016), p. 22959.

[17] M. Benedek, R. J. Schickel, E. Jauk, A. Fink, and A. C. Neubauer. "Alpha power increases in right parietal cortex reflects focused internal attention". In: *Neuropsychologia* 56.1 (2014), pp. 393–400.

[18] M. Benedek, R. Stoiser, S. Walcher, and C. Körner. "Eye behavior associated with internally versus externally directed cognition". In: *Frontiers in Psychology* 8 (2017), pp. 1–9.

[19] A. Benitez-Andonegui, R. Burden, R. Benning, R. Möckel, M. Lührs, and B. Sorger. "An augmented-reality fNIRS-based brain-computer interface: a proof-of-concept study". In: *Frontiers in neuroscience* 14 (2020), p. 346.

[20] J. Bentham. *The panopticon writings*. Verso Books, 2011.

[21] H. Berger. "Über das Elektroenkephalogramm des Menschen". In: *Archiv für Psychiatrie und Nervenkrankheiten* 87.1 (1929), pp. 527–570.

[22] C. W. Berridge and B. D. Waterhouse. "The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes". In: *Brain research reviews* 42.1 (2003), pp. 33–84.

[23] F. Beverina, G. Palmas, S. Silvoni, F. Piccione, S. Giove, et al. "User adaptive BCIs: SSVEP and P300 based interfaces". In: *PsychNology J.* 1.4 (2003), pp. 331–354.

[24] F. Biocca, A. Tang, C. Owen, and F. Xiao. "Attention funnel: omnidirectional 3D cursor for mobile augmented reality platforms". In: *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. Montreal, Canada: ACM, 2006.

[25] B. Blankertz, C. Sanelli, S. Halder, E. Hammer, A. Kübler, K.-R. Müller, G. Curio, and T. Dickhaus. "Predicting BCI performance to study BCI illiteracy". In: *BMC Neurosci* 10.Suppl 1 (2009), P84.

[26] L. Bonanni, C.-H. Lee, and T. Selker. "Attention-based Design of Augmented Reality Interfaces". In: *Extended Abstracts on Human Factors in Computing Systems*. Portland, OR, USA: ACM, 2005.

[27] C. A. Bosman, J.-M. Schoffelen, N. Brunet, R. Oostenveld, A. M. Bastos, T. Womelsdorf, B. Rubehn, T. Stieglitz, P. De Weerd, and P. Fries. "Attentional stimulus selection through selective synchronization between monkey visual areas". In: *Neuron* 75.5 (2012), pp. 875–888.

[28] A.-M. Brouwer, M. A. Hogervorst, B. Oudejans, A. J. Ries, and J. Touryan. "EEG and Eye Tracking Signatures of Target Encoding during Structured Visual Search". In: *Frontiers in Human Neuroscience* 11 (2017), p. 264.

[29] A.-M. Brouwer, B. Reuderink, J. Vincent, M. A. J. van Gerven, and J. B. F. van Erp. "Distinguishing between target and nontarget fixations in a visual search task using fixation-related potentials". In: *Journal of Vision* 13.3 (July 2013), pp. 17–17.

[30] T. J. Buschman and E. K. Miller. "Serial, covert shifts of attention during visual search are reflected by the frontal eye fields and correlated with population oscillations". In: *Neuron* 63.3 (2009), pp. 386–396.

[31] J. Carmigniani and B. Furht. "Augmented reality: an overview". In: *Handbook of augmented reality* (2011), pp. 3–46.

[32] S. M. Ceh, S. Annerer-Walcher, C. Körner, C. Rominger, S. E. Kober, A. Fink, and M. Benedek. "Neurophysiological indicators of internal attention: An electroencephalography–eye-tracking coregistration study". In: *Brain and behavior* 10.10 (2020), e01790.

[33] Z. Chen. "Object-based attention: A tutorial review". In: *Attention, Perception, & Psychophysics* 74.5 (2012), pp. 784–802.

[34] M. M. Chun, J. D. Golomb, and N. B. Turk-Browne. "A Taxonomy of External and Internal Attention". In: *Annual Review of Psychology* 62.1 (2011), pp. 73–101.

[35] P. Cipresso, L. Carelli, F. Solca, D. Meazzi, P. Meriggi, B. Poletti, D. Lulé, A. C. Ludolph, V. Silani, and G. Riva. "The use of P300-based BCIs in amyotrophic lateral sclerosis: from augmentative and alternative communication to cognitive assessment". In: *Brain and behavior* 2.4 (2012), pp. 479–498.

[36] P. Cohen, S. G. West, and L. S. Aiken. *Applied multiple regression/correlation analysis for the behavioral sciences*. Psychology press, 2014.

[37] A. Coin, M. Mulder, and V. Dubljević. "Ethical aspects of BCI technology: what is the state of the art?" In: *Philosophies* 5.4 (2020), p. 31.

[38] H. Collewijn and E. Kowler. "The significance of microsaccades for vision and oculomotor control". In: *Journal of Vision* 8.14 (2008), pp. 20–20.

[39] B. W. Connors and W. G. Regehr. "Neuronal firing: Does function follow form?" In: *Current Biology* 6.12 (1996), pp. 1560–1562.

[40] N. R. Cooper, R. J. Croft, S. J. Dominey, A. P. Burgess, and J. H. Gruzelier. "Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses". In: *International Journal of Psychophysiology* 47.1 (2003), pp. 65–74.

[41] M. Corbetta and G. L. Shulman. "Control of goal-directed and stimulus-driven attention in the brain". In: *Nature reviews neuroscience* 3.3 (2002), pp. 201–215.

[42] I. Damian, C. S. (Tan, T. Baur, J. Schöning, K. Luyten, and E. André. "Augmenting Social Interactions: Realtime Behavioural Feedback Using Social Signal Processing Techniques". In: Seoul, South Korea: ACM, 2015.

[43] R. Desimone and J. Duncan. "Neural mechanisms of selective visual attention". In: *Annual review of neuroscience* 18.1 (1995), pp. 193–222.

[44] DHL. *Logistics trend radar. delivering insight Today. creating value Tomorrow.* [Online; accessed 05-January-2022], https://www.dhl.com.

[45] T. Dickhaus, C. Sannelli, K.-R. Müller, G. Curio, and B. Blankertz. "Predicting BCI performance to study BCI illiteracy". In: *BMC Neuroscience* 10.1 (2009), P84.

[46] B. J. Dixon, M. J. Daly, H. Chan, A. D. Vescan, I. J. Witterick, and J. C. Irish. "Surgeons blinded by enhanced navigation: The effect of augmented reality on attention". In: *Surgical Endoscopy* 27 (2013), pp. 454–461.

[47] M. Donk and W. Van Zoest. "Effects of salience are short-lived". In: *Psychological Science* 19.7 (2008), pp. 733–739.

[48] *DSI-VR300 by Neurospec AG*. [Online; accessed 22-January-2022], https://www.neurospec.com.

[49] *EEG - ECG - Biosensors*. [Online; accessed 05-January-2022], `http://neurosky.com/`.

[50] T. Elbert, B. Rockstroh, W. Lutzenberger, and N. Birbaumer. "Biofeedback of slow cortical potentials. I". In: *Electroencephalography and clinical neurophysiology* 48.3 (1980), pp. 293–301.

[51] C. Elmadjian and C. H. Morimoto. "GazeBar: Exploiting the Midas Touch in Gaze Interaction". In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. virtual: ACM, 2021.

[52] J. Enright. "Changes in vergence mediated by saccades." In: *The Journal of physiology* 350.1 (1984), pp. 9–31.

[53] C. Erkelens, R. Steinman, and H. Collewijn. "Ocular vergence under natural conditions. II. Gaze shifts between real targets differing in distance and direction". In: *Proceedings of the Royal Society of London. B. Biological Sciences* 236.1285 (1989), pp. 441–465.

[54] R. Eyraud, E. Zibetti, and T. Baccino. "Allocation of visual attention while driving with simulated augmented reality". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 32 (2015), pp. 46–55.

[55] J. Faller, B. Z. Allison, C. Brunner, R. Scherer, D. Schmalstieg, G. Pfurtscheller, and C. Neuper. "A feasibility study on SSVEP-based interaction with motivating and immersive virtual and augmented reality". In: *CoRR* abs/1701.03981 (2017), pp. 1–6.

[56] A. M. Feit, S. Williams, A. Toledo, A. Paradiso, H. Kulkarni, S. Kane, and M. R. Morris. "Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design". In: *Proceedings of the CHI conference on human factors in computing systems*. Denver, CO, USA: ACM, 2017.

[57] E. E. Fetz. "Operant conditioning of cortical unit activity". In: *Science* 163.3870 (1969), pp. 955–958.

[58] A. Finke, K. Essig, G. Marchioro, and H. Ritter. "Toward FRP-based brain-machine interfaces-single-trial classification of fixation-related potentials". In: *PLoS ONE* 11.1 (2016), e0146848.

[59] B. Fischer and E. Ramsperger. "Human express saccades: extremely short reaction times of goal directed eye movements". In: *Experimental brain research* 57.1 (1984), pp. 191–195.

[60] C. Flavián, S. Ibáñez-Sánchez, and C. Orús. "The influence of scent on virtual reality experiences: the role of aroma-content congruence". In: *Journal of Business Research* 123 (2021), pp. 289–301.

[61] J. W. de Fockert, G. Rees, C. D. Frith, and N. Lavie. "The role of working memory in visual selective attention". In: *Science* 291.5509 (2001), pp. 1803–1806.

[62] P. Fries, T. Womelsdorf, R. Oostenveld, and R. Desimone. "The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area V4". In: *Journal of Neuroscience* 28.18 (2008), pp. 4823–4835.

[63] A. Fuchs. "Saccadic and smooth pursuit eye movements in the monkey". In: *The Journal of Physiology* 191.3 (1967), pp. 609–631.

[64] N. Galley, D. Betz, and C. Biniossek. "Fixation durations - Why are they so highly variable?" In: *Das Ende von Rational Choice? Zur Leistungsfähigkeit der Rational-Choice-Theorie*. Vol. 93. Jan. 2015, pp. 83–106.

[65] R. Gavas, D. Chatterjee, and A. Sinha. "Estimation of cognitive load based on the pupil size dilation". In: *Proceedings of the International Conference on Systems, Man, and Cybernetics*. Banff, Canada: IEEE, 2017, pp. 1499–1504.

[66] A. D. Gavrilov, A. Jordache, M. Vasdani, and J. Deng. "Preventing model overfitting and underfitting in convolutional neural networks". In: *International Journal of Software Science and Computational Intelligence* 10.4 (2018), pp. 19–28.

[67] Grandviewresearch. "Augmented Reality Market Size & Share Report". In: *Next Generation Technologies* (2021).

[68] S. Grassini, A. Revonsuo, S. Castellotti, I. Petrizzo, V. Benedetti, and M. Koivisto. "Processing of natural scenery is associated with lower attentional and cognitive load compared with urban ones". In: *Journal of Environmental Psychology* 62 (2019), pp. 1–11.

[69] G. Grübler, A. Al-Khodairy, R. Leeb, I. Pisotta, A. Riccio, M. Rohm, and E. Hildt. "Psychosocial and ethical aspects in non-invasive EEG-based BCI research—a survey among BCI users and BCI professionals". In: *Neuroethics* 7.1 (2014), pp. 29–41.

[70] L. F. Haas. "Hans berger (1873–1941), richard caton (1842–1926), and electroencephalography". In: *Journal of Neurology, Neurosurgery & Psychiatry* 74.1 (2003), p. 9.

[71] D. Hallinan, P. Schütz, M. Friedewald, and P. De Hert. "Neurodata and neuroprivacy: Data protection outdated?" In: *Surveillance & Society* 12.1 (2014), pp. 55–72.

[72] D. W. Hansen and Q. Ji. "In the eye of the beholder: A survey of models for eyes and gaze". In: *IEEE transactions on pattern analysis and machine intelligence* 32.3 (2009), pp. 478–500.

[73] P. Haselager, R. Vlek, J. Hill, and F. Nijboer. "A note on ethical aspects of BCI". In: *Neural Networks* 22.9 (2009), pp. 1352–1357.

[74] B. Y. Hayden and J. L. Gallant. "Combined effects of spatial and feature-based attention on responses of V4 neurons". In: *Vision research* 49.10 (2009), pp. 1182–1187.

[75] M. Hayhoe and D. Ballard. "Eye movements in natural behavior". In: *Trends in cognitive sciences* 9.4 (2005), pp. 188–194.

[76] HBR. "Augmented Reality in the real world". In: *Harvard Business Review* (Nov-Dec 2017), p. 59.

[77] D. Heger, F. Putze, and T. Schultz. "An EEG adaptive information system for an empathic robot". In: *International Journal of Social Robotics* 3 (2011), pp. 415–425.

[78] M. L. Heilig. *Sensorama simulator*. Google Patents. 1962.

[79] E. H. Hess and J. M. Polt. "Pupil size in relation to mental activity during simple problem-solving". In: *Science* 143.3611 (1964), pp. 1190–1192.

[80] S. A. Hillyard and L. Anllo-Vento. "Event-related brain potentials in the study of visual selective attention". In: *Proceedings of the National Academy of Sciences* 95.3 (1998), pp. 781–787.

[81] J. Himes. "Top 5 use cases for augmented reality (AR) in 2021". In: *CGS blog* (July 2021).

[82] J. E. Hoffman and B. Subramaniam. "The role of visual attention in saccadic eye movements". In: *Perception & Psychophysics* 57 (1995), pp. 787–795.

[83] F. Hutmacher. "Why is there so much more research on vision than on any other sensory modality?" In: *Frontiers in psychology* 10 (2019), p. 2246.

[84] S. Hutt, J. Hardey, R. Bixler, A. Stewart, E. Risko, and S. K. D'Mello. "Gaze-Based Detection of Mind Wandering during Lecture Viewing." In: *International Conference on Educational Data Mining*. Wuhan, China: ERIC, 2017.

[85] H.-J. Hwang, S. Kim, S. Choi, and C.-H. Im. "EEG-based brain-computer interfaces: a thorough literature survey". In: *International Journal of Human-Computer Interaction* 29.12 (2013), pp. 814–826.

[86] *Ikea Place App*. [Online; accessed 30-October-2021], https://www.ikea.com.

[87] L. Itti and C. Koch. "Computational modelling of visual attention". In: *Nature reviews neuroscience* 2.3 (2001), pp. 194–203.

[88] L. Itti and C. Koch. "Feature combination strategies for saliency-based visual attention systems". In: *Journal of Electronic imaging* 10.1 (2001), pp. 161–169.

[89] H. Iwata, H. Yano, T. Uemura, and T. Moriya. "Food simulator: A haptic interface for biting". In: *Proceedings of the IEEE Virtual Reality*. Chicago, IL, USA: IEEE, 2004, pp. 51–57.

[90] A. K. Jain, J. Mao, and K. M. Mohiuddin. "Artificial neural networks: A tutorial". In: *Computer* 29.3 (1996), pp. 31–44.

[91] W. James. *The principles of psychology*. Vol. 1. Cosimo, Inc., 2007.

[92] A. Johnson and R. W. Proctor. *Attention: Theory and practice*. Sage, 2004.

[93] S. Jones and S. Dawkins. "The sensorama revisited: evaluating the application of multi-sensory input on the sense of presence in 360-degree immersive film in virtual reality". In: *Augmented reality and virtual reality* (2018), pp. 183–197.

[94] S. Julier, Y. Baillot, D. Brown, M. Lanzagorta, and M. Lanzagorta. "Information Filtering for Mobile Augmented Reality". In: *IEEE Comput. Graph. Appl.* 22.5 (2002), pp. 12–15.

[95] K. Kansaku, N. Hata, and K. Takano. "My thoughts through a robot's eyes: An augmented reality-brain–machine interface". In: *Neuroscience research* 66.2 (2010), pp. 219–222.

[96] A. J. Karran, T. Demazure, P.-M. Leger, E. Labonte-LeMoyne, S. Senecal, M. Fredette, and G. Babin. "Toward a hybrid passive bci for the modulation of sustained attention using EEG and fNIRS". In: *Frontiers in human neuroscience* 13 (2019), p. 393.

[97] I. Katidioti, J. P. Borst, D. J. Bierens de Haan, T. Pepping, M. K. van Vugt, and N. A. Taatgen. "Interrupted by Your Pupil: An Interruption Management System Based on Pupil Dilation". In: *International Journal of Human-Computer Interaction* 32.10 (2016), pp. 791–801.

[98] Y. Ke, P. Liu, X. An, X. Song, and D. Ming. "An online SSVEP-BCI system in an optical see-through augmented reality environment". In: *Journal of neural engineering* 17.1 (2020), p. 016066.

[99] S. Kim, S. Lee, H. Kang, S. Kim, and M. Ahn. "P300 Brain–Computer Interface-Based Drone Control in Virtual and Augmented Reality". In: *Sensors* 21.17 (2021), p. 5765.

[100] S. Kishore, M. González-Franco, C. Hintemüller, C. Kapeller, C. Guger, M. Slater, and K. J. Blom. "Comparison of SSVEP BCI and eye tracking for controlling a humanoid robot in a social environment". In: *Presence: Teleoperators and Virtual Environments* 23.3 (2014), pp. 242–252.

[101] M. L. Knapp, J. A. Hall, and T. G. Horgan. *Nonverbal communication in human interaction*. Cengage Learning, 2013.

[102] R. J. Krauzlis, L. P. Lovejoy, and A. Zénon. "Superior colliculus and visual spatial attention". In: *Annual review of neuroscience* 36 (2013), pp. 165–182.

[103] A. Kumar, L. Gao, E. Pirogova, and Q. Fang. "A review of error-related potential-based brain–computer interfaces for motor impaired people". In: *IEEE Access* 7 (2019), pp. 142451–142466.

[104] J. S. Kumar and P. Bhuvaneswari. "Analysis of Electroencephalography (EEG) signals and its categorization–a study". In: *Procedia engineering* 38 (2012), pp. 2525–2536.

[105] M. Kusunoki, J. Gottlieb, and M. E. Goldberg. "The lateral intraparietal area as a salience map: the representation of abrupt onset, stimulus motion, and task relevance". In: *Vision research* 40.10-12 (2000), pp. 1459–1468.

[106] V. A. Lamme and P. R. Roelfsema. "The distinct modes of vision offered by feedforward and recurrent processing". In: *Trends in neurosciences* 23.11 (2000), pp. 571–579.

[107] H. A. Lamti, P. Gorce, M. M. Ben Khelifa, and A. M. Alimi. "When mental fatigue maybe characterized by Event Related Potential (P300) during virtual wheelchair navigation". In: *Computer methods in biomechanics and biomedical engineering* 19.16 (2016), pp. 1749–1759.

[108] Y. LeCun. "Deep learning & convolutional networks." In: *Proceedings of the IEEE Hot Chips Symposium*. Cupertino, CA, USA, 2015, pp. 1–95.

[109] Y. LeCun, L. Jackel, L. Bottou, A. Brunot, C. Cortes, J. Denker, H. Drucker, I. Guyon, U. Muller, E. Sackinger, et al. "Comparison of learning algorithms for handwritten digit recognition". In: *Proceedings of the IEEE International conference on artificial neural networks*. Vol. 60. Perth, Australia, 1995, pp. 53–60.

[110] A. Legatt. "Evoked Potentials". In: *Encyclopedia of the Neurological Sciences*. Ed. by M. J. Aminoff and R. B. Daroff. Second Edition. Oxford: Academic Press, 2014, pp. 228–231.

[111] A. Lenhardt and H. Ritter. "An augmented-reality based brain-computer interface for robot control". In: *Proceedings of the International Conference on Neural Information Processing*. Springer. Vancouver, Canada, 2010, pp. 58–65.

[112] G. Li, S. Huang, W. Xu, W. Jiao, Y. Jiang, Z. Gao, and J. Zhang. "The impact of mental fatigue on brain activity: a comparative study both in resting state and task state using EEG". In: *BMC neuroscience* 21 (2020), pp. 1–9.

[113] G. W. Lindsay. "Attention in psychology, neuroscience, and machine learning". In: *Frontiers in computational neuroscience* 14 (2020), p. 29.

[114] P. Liu, Y. Ke, J. Du, W. Liu, L. Kong, N. Wang, X. An, and D. Ming. "An SSVEP-BCI in Augmented Reality". In: *Proceedings of the Engineering in Medicine and Biology Society*. IEEE. Berlin, Germany, 2019, pp. 5548–5551.

[115] S. Liu and D. Manocha. "Sound synthesis, propagation, and rendering: a survey". In: *arXiv 2011.05538* (2020).

[116] W. Lu, B. L. H. Duh, and S. Feiner. "Subtle cueing for visual search in augmented reality". In: *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality, Science and Technology Papers*. Atlanta, GA, USA, 2012, pp. 161–166.

[117] B. Mahesh. "Machine learning algorithms-a review". In: *International Journal of Science and Research* 9 (2020), pp. 381–386.

[118] *Marktstudie: Wo steht Deutschland bei augmented und virtual reality?* [Online; accessed 19-October-2021], https://www.ptc.com/.

[119] A. M. Martinez and A. C. Kak. "PCA versus LDA". In: *IEEE transactions on pattern analysis and machine intelligence* 23.2 (2001), pp. 228–233.

[120] K. Matsuda. *Hyper-Reality*. May 2016. URL: http://hyper-reality.co/.

[121] C. J. McAdams and J. H. Maunsell. "Effects of attention on the reliability of individual neurons in monkey visual cortex". In: *Neuron* 23.4 (1999), pp. 765–773.

[122] A. McKinlay and K. Starkey. *Foucault, management and organization theory: From panopticon to technologies of self*. Sage, 1998.

[123] A. McNamara and C. Kabeerdoss. "Mobile Augmented Reality: Placing Labels Based on Gaze Position". In: *Adjunct Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*. Merida, Mexico, 2016.

[124] C. Meng and X. Zhao. "Webcam-based eye movement analysis using CNN". In: *IEEE Access* 5 (2017), pp. 19581–19587.

[125] R. Menges, C. Kumar, and S. Staab. "Improving user experience of eye tracking-based interaction: Introspecting and adapting interfaces". In: *ACM Transactions on Computer-Human Interaction* 26.6 (2019), pp. 1–46.

[126] J. Mercier-Ganady, F. Lotte, E. Loup-Escande, M. Marchal, and A. Lecuyer. "The Mind-Mirror: See your brain in action in your head using EEG and augmented reality". In: *Proceedings of the IEEE Virtual Reality*. Los Angeles, CA, USA, 2014, pp. 33–38.

[127] J. Mercier-Ganady, M. Marchal, and A. Lécuyer. "BC-invisibility power: introducing optical camouflage based on mental activity in augmented reality". In: *Proceedings of the 6th Augmented Human International Conference*. Singapore, 2015, pp. 97–100.

[128] J. Mercier-Ganady, M. Marchal, and A. Lécuyer. "The mind-window: brain activity visualization using tablet-based AR and EEG for multiple users". In: *Proceedings of the 6th Augmented Human International Conference*. Singapore, 2015, pp. 197–198.

[129] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino. "Augmented reality: A class of displays on the reality-virtuality continuum". In: *Telemanipulator and telepresence technologies*. Vol. 2351. International Society for Optics and Photonics. 1995, pp. 282–292.

[130] E. K. Miller and T. J. Buschman. "Neural mechanisms for the executive control of attention". In: *The Oxford Handbook of Attention*. Ed. by A. C. (Nobre and S. Kastner. 2014.

[131] J. F. Mitchell, K. A. Sundberg, and J. H. Reynolds. "Differential attention-dependent response modulation across cell classes in macaque visual area V4". In: *Neuron* 55.1 (2007), pp. 131–141.

[132] H. Miyashita. "Norimaki synthesizer: taste display using ion electrophoresis in five gels". In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. Honululu, HI, USA: ACM, 2020, pp. 1–6.

[133] H. Si-Mohammed, F. Argelaguet Sanz, G. Casiez, N. Roussel, and A. Lécuyer. "Brain-Computer Interfaces and Augmented Reality: A State of the Art". In: *Graz Brain-Computer Interface Conference*. Graz, Austria, 2017.

[134] H. Si-Mohammed, J. Petit, C. Jeunet, F. Argelaguet, F. Spindler, A. Evain, N. Roussel, G. Casiez, and A. Lécuyer. "Towards BCI-based Interfaces for Augmented Reality: Feasibility, Design and Evaluation". In: *IEEE Transactions on Visualization and Computer Graphics* (2018), pp. 1–1.

[135] P. Mohan, W. B. Goh, C.-W. Fu, and S.-K. Yeung. "DualGaze: Addressing the midas touch problem in gaze mediated VR interaction". In: *Proceedings of the International Symposium on Mixed and Augmented Reality Adjunct*. IEEE. Munich, Germany, 2018, pp. 79–84.

[136] G. Montavon, A. Binder, S. Lapuschkin, W. Samek, and K.-R. Müller. "Layer-wise relevance propagation: an overview". In: *Explainable AI: interpreting, explaining and visualizing deep learning* (2019), pp. 193–209.

[137] E. Musk et al. "An integrated brain-machine interface platform with thousands of channels". In: *Journal of medical Internet research* 21.10 (2019), e16194.

[138] T. Nakano, M. Kato, Y. Morito, S. Itoi, and S. Kitazawa. "Blink-related momentary activation of the default mode network while viewing videos". In: *Proceedings of the National Academy of Sciences* 110.2 (2013), pp. 702–706.

[139] C. S. Nam, A. Nijholt, and F. Lotte. *Brain–computer interfaces handbook: technological and theoretical advances*. CRC Press, 2018.

[140] C. S. Nam, J. Woo, and S. Bahn. "Severe motor disability affects functional cortical integration in the context of brain–computer interface (BCI) use". In: *Ergonomics* 55.5 (2012), pp. 581–591.

[141] S. Naufel and E. Klein. "Brain–computer interface (BCI) researcher perspectives on neural data ownership and privacy". In: *Journal of neural engineering* 17.1 (2020), p. 016039.

[142] B. Noudoost, M. H. Chang, N. A. Steinmetz, and T. Moore. "Top-down control of visual attention". In: *Current opinion in neurobiology* 20.2 (2010), pp. 183–190.

[143] J. S. Oliveira, F. O. Franco, M. C. Revers, A. F. Silva, J. Portolese, H. Brentani, A. Machado-Lima, and F. L. Nunes. "Computer-aided autism diagnosis based on visual attention models using eye tracking". In: *Scientific reports* 11.1 (2021), pp. 1–11.

[144] J. Otero-Millan, S. L. Macknik, and S. Martinez-Conde. "Fixational eye movements and binocular vision". In: *Frontiers in integrative neuroscience* 8 (2014), p. 52.

[145] R. Parasuraman and P. M. Greenwood. "Selective attention in aging and dementia". In: *The attentive brain*. The MIT Press, 1998, pp. 461–487.

[146] S. Park, H.-S. Cha, J. Kwon, H. Kim, and C.-H. Im. "Development of an online home appliance control system using augmented reality and an ssvep-based brain-computer interface". In: *International Winter Conference on Brain-Computer Interface*. IEEE. Gangwon, South Korea, 2020, pp. 1–2.

[147] H. Pashler, J. C. Johnston, and E. Ruthruff. "Attention and performance". In: *Annual review of psychology* 52.1 (2001), pp. 629–651.

[148] R. Portillo-Lara, B. Tahirbegi, C. A. Chapman, J. A. Goding, and R. A. Green. "Mind the gap: State-of-the-art technologies and applications for EEG-based brain–computer interfaces". In: *APL bioengineering* 5.3 (2021), p. 031507.

[149] M. I. Posner. "Orienting of attention". In: *Quarterly journal of experimental psychology* 32.1 (1980), pp. 3–25.

[150] A. L. Proskovec, E. Heinrichs-Graham, A. I. Wiesman, T. J. McDermott, and T. W. Wilson. "Oscillatory dynamics in the dorsal and ventral attention networks during the reorienting of attention". In: *Human brain mapping* 39.5 (2018), pp. 2177–2190.

[151] F. Putze, J. Hild, R. Kärgel, C. Herff, A. Redmann, J. Beyerer, and T. Schultz. "Locating user attention using eye tracking and EEG for spatio-temporal event selection". In: *Proceedings of the International Conference on Intelligent User Interfaces*. Santa Monica, CA, USA, Mar. 2013, p. 129.

[152] F. Putze, M. Scherer, and T. Schultz. "Starring into the void?: Classifying Internal vs. External Attention from EEG". In: *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. ACM. Gothenburg, Sweden: Association for Computing Machinery, 2016, p. 47.

[153] S. Rainey, K. McGillivray, S. Akintoye, T. Fothergill, C. Bublitz, and B. Stahl. "Is the European Data Protection Regulation sufficient to deal with emerging data concerns relating to neurotechnology?" In: *Journal of Law and the Biosciences* 7.1 (2020), lsaa051.

[154] M. Rashid, N. Sulaiman, A. PP Abdul Majeed, R. M. Musa, B. S. Bari, S. Khatun, et al. "Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review". In: *Frontiers in neurorobotics* 14 (2020), p. 25.

[155] W. Rawat and Z. Wang. "Deep convolutional neural networks for image classification: A comprehensive review". In: *Neural computation* 29.9 (2017), pp. 2352–2449.

[156] K. Rayner and A. D. Well. "Effects of contextual constraint on eye movements in reading: A further examination". In: *Psychonomic Bulletin & Review* 3.4 (1996), pp. 504–509.

[157] G. L. Read and I. J. Innis. "Electroencephalography (EEG)". In: *The international encyclopedia of communication research methods* (2017), pp. 1–18.

[158] G. Rees and N. Lavie. "What can functional imaging reveal about the role of attention in visual awareness?" In: *Neuropsychologia* 39.12 (2001), pp. 1343–1353.

[159] D. Regan. "Some characteristics of average steady-state and transient responses evoked by modulated light". In: *Electroencephalography and clinical neurophysiology* 20.3 (1966), pp. 238–248.

[160] R. Rivu, Y. Abdrabou, K. Pfeuffer, A. Esteves, S. Meitner, and F. Alt. "StARe: Gaze-Assisted Face-to-Face Communication in Augmented Reality". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. Stuttgart, Germany: ACM, 2020, pp. 1–5.

[161] G. Rizzolatti, L. Riggio, I. Dascola, and C. Umiltá. "Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention". In: *Neuropsychologia* 25.1 (1987), pp. 31–40.

[162] D. A. Robinson. "A method of measuring eye movemnent using a scieral search coil in a magnetic field". In: *IEEE Transactions on bio-medical electronics* 10.4 (1963), pp. 137–145.

[163] M. Rodrigue, J. Son, B. Giesbrecht, M. Turk, and T. Höllerer. "Spatio-Temporal Detection of Divided Attention in Reading Applications Using EEG and Eye Tracking". In: *Proceedings of the International Conference on Intelligent User Interfaces*. Atlanta, Georgia, USA: ACM, 2015, pp. 121–125.

[164] J. K. Salisbury and M. A. Srinivasan. "Phantom-based haptic interaction with virtual objects". In: *IEEE Computer Graphics and Applications* 17.5 (1997), pp. 6–10.

[165] L. Savioja and U. P. Svensson. "Overview of geometrical room acoustic modeling techniques". In: *The Journal of the Acoustical Society of America* 138.2 (2015), pp. 708–730.

[166] R. M. Shiffrin and M. P. Czerwinski. "A model of automatic attention attraction when mapping is partially consistent." In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14.3 (1988), p. 562.

[167] J. J. Shih, D. J. Krusienski, and J. R. Wolpaw. "Brain-computer interfaces in medicine". In: 87.3 (2012), pp. 268–279.

[168] D. J. Simons and D. T. Levin. "Change blindness". In: *Trends in cognitive sciences* 1.7 (1997), pp. 261–267.

[169] R. Skarbez, M. Smith, and M. C. Whitton. "Revisiting Milgram and Kishino's Reality-Virtuality Continuum". In: *Frontiers in Virtual Reality* 2 (2021), p. 27.

[170] R. H. C. e. Souza and E. L. M. Naves. "Attention Detection in Virtual Environments Using EEG Signals: A Scoping Review". In: *Frontiers in Physiology* 12 (2021), p. 2051.

[171] S. Sridharan and R. Bailey. "Automatic Target Prediction and Subtle Gaze Guidance for Improved Spatial Information Recall". In: *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*. Tübingen, Germany: ACM, 2015, pp. 99–106.

[172] E. Styles. *The psychology of attention*. Psychology Press, 2006.

[173] E. F. Tait. "A report on the results of the experimental variation of the stimulus conditions in the responses of the accommodation convergence reflex". In: *Optometry and Vision Science* 10.12 (1933), pp. 428–435.

[174] K. Takano, N. Hata, and K. Kansaku. "Towards intelligent environments: An augmented reality-brain-machine interface operated with a see-through head-mount display". In: *Frontiers in Neuroscience* 5.60 (2011), pp. 1–5.

[175] *An Ancient Greek Tale*. [Online; accessed 19-October-2021], `https://www.storyarts.org/library/nutshell/stories/golden.html`.

[176] S. Treue and J. C. M. Trujillo. "Feature-based attention influences motion processing gain in macaque visual cortex". In: *Nature* 399.6736 (1999), pp. 575–579.

[177] X. G. Troncoso, S. L. Macknik, and S. Martinez-Conde. "Microsaccades counteract perceptual filling-in". In: *Journal of vision* 8.14 (2008), pp. 15–15.

[178] A. Ulahannan, P. Jennings, L. Oliveira, and S. Birrell. "Designing an adaptive interface: Using eye tracking to classify how information usage changes over time in partially automated vehicles". In: *IEEE Access* 8 (2020), pp. 16865–16875.

[179] *Upskill and Boeing Augmented Reality*. [Online; accessed 19-October-2021], http://go.upskill.io.

[180] L. Van Run and A. Van den Berg. "Binocular eye orientation during fixations: Listing's law extended to include eye vergence". In: *Vision research* 33.5-6 (1993), pp. 691–708.

[181] B. B. Velichkovsky, M. A. Rumyantsev, and M. A. Morozov. "New solution to the midas touch problem: Identification of visual commands via extraction of focal fixations". In: *procedia computer science* 39 (2014), pp. 75–82.

[182] J. J. Vidal. "Toward direct brain-computer communication". In: *Annual review of Biophysics and Bioengineering* 2.1 (1973), pp. 157–180.

[183] *Vision*. [Online; accessed 22-January-2022]. OpenStax CNX, Sept. 2020.

[184] S. Vossel, J. J. Geng, and G. R. Fink. "Dorsal and ventral attention systems: distinct neural circuits but collaborative roles". In: *The Neuroscientist* 20.2 (2014), pp. 150–159.

[185] S. Walcher, C. Körner, and M. Benedek. "Looking for ideas: Eye behavior during goal-directed internally focused cognition". In: *Consciousness and cognition* 53 (2017), pp. 165–175.

[186] M. Wang, R. Li, R. Zhang, G. Li, and D. Zhang. "A Wearable SSVEP-Based BCI System for Quadcopter Control Using Head-Mounted Device". In: *IEEE Access* 6 (2018), pp. 26789–26798.

[187] Y. Wang, K. Li, X. Zhang, J. Wang, and R. Wei. "Research on the Application of Augmented Reality in SSVEP-BCI". In: *Proceedings of the 6th International Conference on Computing and Artificial Intelligence*. Tianjin, China, 2020, pp. 505–509.

[188] Q. Wei, S. Zhu, Y. Wang, X. Gao, H. Guo, and X. Wu. "Maximum signal fraction analysis for enhancing signal-to-noise ratio of EEG signals in SSVEP-based BCIs". In: *IEEE Access* 7 (2019), pp. 85452–85461.

[189] A. Williams. "Reality check: How ar can improve efficiency in logistics". In: *Automotive Logistics* (2019).

[190] D. Wobrock, A. Finke, T. Schack, and H. Ritter. "Using Fixation-Related Potentials for Inspecting Natural Interactions". In: *Frontiers in Human Neuroscience* 14 (2020), p. 447.

[191] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. "Brain–computer interfaces for communication and control". In: *Clinical neurophysiology* 113.6 (2002), pp. 767–791.

[192] M. Xuelin Huang, J. Li, G. Ngai, H. V. Leong, and A. Bulling. "Moment-to-moment detection of internal thought from eye vergence behaviour". In: *arXiv 1901.06572* (2019).

[193] Y. Yanagida. "A survey of olfactory displays: Making and delivering scents". In: *Proceedings of the IEEE sensors*. Taipei, Taiwan, 2012, pp. 1–4.

[194] R. Yousefi, A. Rezazadeh Sereshkeh, and T. Chau. "Online detection of error-related potentials in multi-class cognitive task-based BCIs". In: *Brain-Computer Interfaces* 6.1-2 (2019), pp. 1–12.

[195] T. O. Zander, C. Kothe, S. Jatzev, and M. Gaertner. "Enhancing human-computer interaction with input from active and passive brain-computer interfaces". In: *Brain-computer interfaces*. Springer, 2010, pp. 181–199.

[196] X. Zhang and D. Wu. "On the vulnerability of CNN classifiers in EEG-based BCIs". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27.5 (2019), pp. 814–825.

[197] X. Zhao, Y. Du, and R. Zhang. "A CNN-based multi-target fast classification method for AR-SSVEP". In: *Computers in biology and medicine* (2021), p. 105042.

[198] X. Zhao, C. Liu, Z. Xu, L. Zhang, and R. Zhang. "SSVEP Stimulus Layout Effect on Accuracy of Brain-Computer Interfaces in Augmented Reality Glasses". In: *IEEE Access* 8 (2020), pp. 5990–5998.

[199] H. Zhou, R. J. Schafer, and R. Desimone. "Pulvinar-cortex interactions in vision and attention". In: *Neuron* 89.1 (2016), pp. 209–220.

[200] W. van Zoest and M. Donk. "The effects of salience on saccadic target selection". In: *Visual Cognition* 12.2 (2005), pp. 353–375.

[201] S. Zuboff. "Big other: surveillance capitalism and the prospects of an information civilization". In: *Journal of information technology* 30.1 (2015), pp. 75–89.

A

Accumulated Publications

Content

EEG-based Classification of Internally- and Externally-directed Attention in an Augmented Reality Paradigm

by

Lisa-Marie Vortmann, Felix Kroll, and Felix Putze

published in: Frontiers in Human Neuroscience
2019, Volume 13, Article 348
DOI: 10.3389/fnhum.2019.00348

Abstract

One problem faced in the design of Augmented Reality (AR) applications is the interference of virtually displayed objects in the user's visual field, with the current attentional focus of the user. Newly generated content can disrupt internal thought processes. If we can detect such internally-directed attention periods, the interruption could either be avoided or even used intentionally. In this work, we designed a special alignment task in AR with two conditions: one with externally-directed attention and one with internally-directed attention. Apart from the direction of attention, the two tasks were identical. During the experiment, we performed a 16-channel EEG recording, which was then used for a binary classification task. Based on selected band power features, we trained a Linear Discriminant Analysis classifier to predict the label for a 13-second window of each trial. Parameter selection, as well as the training of the classifier, were done in a person-dependent manner in a 5-fold cross-validation on the training data. We achieved an average score of approximately 85.37% accuracy on the test data ($\pm$ 11.27%, range = [66.7%, 100%], 6 participants $>$ 90%, 3 participants = 100%). Our results show that it is possible to discriminate the two states with simple machine learning mechanisms. The analysis of additionally collected data dispels doubts that we classified the difference in movement speed or task load. We conclude that a real-time assessment of internal and external attention in an AR setting in general will be possible.

Keywords: Internal Attention, External Attention, EEG, Augmented Reality, Classification, Brain-Computer Interface

Contribution Statement: Conceptualization of the project including the formulation of research aims and goals; Development or design of methodology; Supervision of the implementation; Data recording; Implementation of data analysis; Discussion of the results; Writing and reviewing of the manuscript.

Using Brain Activity Patterns to Differentiate Real and Virtual Attended Targets during Augmented Reality Scenarios

by

Lisa-Marie Vortmann, Leonid Schwenke, and Felix Putze

published in: MDPI Information
2021, Volume 12, Issue 6, page 226
DOI: 10.3390/info12060226

Abstract

Augmented reality is the fusion of virtual components and our real surroundings. The simultaneous visibility of generated and natural objects often requires users to direct their selective attention to a specific target that is either real or virtual. In this study, we investigated whether this target is real or virtual by using machine learning techniques to classify electroencephalographic (EEG) and eye tracking data collected in augmented reality scenarios. A shallow convolutional neural net classified 3 second EEG data windows from 20 participants in a person-dependent manner with an average accuracy above 70% if the testing data and training data came from different trials. This accuracy could be significantly increased to 77% using a multimodal late fusion approach that included the recorded eye tracking data. Person-independent EEG classification was possible above chance level for 6 out of 20 participants. Thus, the reliability of such a brain–computer interface is high enough for it to be treated as a useful input mechanism for augmented reality applications.

Keywords: augmented reality; neural networks; eye tracking; classification; attention; EEG

Contribution Statement: Conceptualization of the project including the formulation of research aims and goals; Development and design of methodology; Supervision of the implementation, data recording and analysis; Discussion of the results; Writing and reviewing of the manuscript.

Model-based Prediction of Exogeneous and Endogeneous Attention Shifts During an Everyday Activity

by

Felix Putze, Merlin Burri, Lisa-Marie Vortmann, and Tanja Schultz

published in: ICMI '20 Companion
Companion Publication of the 2020 International Conference on Multimodal Interaction, October 25–29, 2020, Virtual event, Netherlands
DOI: 10.1145/3395035.3425206

Abstract

Human attention determines to a large degree how users interact with technical devices and how technical artifacts can support them optimally during their tasks. Attention shifts between different targets, triggered through changing requirements of an ongoing task or through salient distractions in the environment. Such shifts mark important transition points which an intelligent system needs to predict and attribute to an endogenous or exogenous cause for an appropriate reaction. In this paper, we describe a model which performs this task through a combination of bottom-up and top-down modeling components. We evaluate the model in a scenario with a dynamic task in a rich environment and show that the model is able to predict attention future switches with a robust classification performance.

Keywords: attention shifts, top-down and bottom-up modeling, exogenous and endogenous attention

Contribution Statement: Involved in the development and design of the experiment, the discussion of the results and reviewing of the manuscript

Differentiating Endogenous and Exogenous Attention Shifts Based on Fixation-Related Potentials

by

Lisa-Marie Vortmann, Moritz Schult, and Felix Putze

published in: IUI '22
27th International Conference on Intelligent User Interfaces,
March 22–25, 2022, Helsinki, Finland
DOI: 10.1145/3490099.3511149

Abstract

Attentional shifts can occur voluntarily (endogenous control) or reflexively (exogenous control). Previous studies have shown that the neural mechanisms underlying these shifts produce different activity patterns in the brain. Changes in visual-spatial attention are usually accompanied by eye movements and a fixation on the new center of attention. In this study, we analyze the fixation-related potentials in electroencephalographic recordings of 10 participants during computer screen-based viewing tasks. During task performance, we presented salient visual distractors to evoke reflexive attention shifts. Surrounding each fixation, 0.7-second data windows were extracted and labeled as "endogenous" or "exogenous". Averaged over all participants, the balanced classification accuracy using a person-dependent Linear Discriminant Analysis reached 59.84%. In a leave-one-participant-out approach, the average classification accuracy reached 58.48%. Differentiating attention shifts, based on fixation-related potentials, could be used to deepen the understanding of human viewing behavior or as a Brain-Computer Interface for attention-aware user interface adaptations.

Keywords: fixation related potential, EEG, Attention, endogenous, exogenous, gaze detection, linear discriminant analysis

Contribution Statement: Conceptualization of the project including the formulation of research aims and goals; Development and design of methodology; Supervision of the implementation, data recording and analysis; Discussion of the results; Writing and reviewing of the manuscript.

Exploration of Person-Independent BCIs for Internal and External Attention-Detection in Augmented Reality

by

Lisa-Marie Vortmann, and Felix Putze

published in: IMWUT Journal
Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2021, Volume 5, Issue 2, pages 1-27
DOI: 10.1145/3463507

Abstract

Adding attention-awareness to an Augmented Reality setting by using a Brain-Computer Interface promises many interesting new applications and improved usability. The possibly complicated setup and relatively long training period of EEG-based BCIs however, reduce this positive effect immensely. In this study, we aim at finding solutions for person-independent, training-free BCI integration into AR to classify internally and externally directed attention. We assessed several different classifier settings on a dataset of 14 participants consisting of simultaneously recorded EEG and eye tracking data. For this, we compared the classification accuracies of a linear algorithm, a non-linear algorithm, and a neural net that were trained on a specifically generated feature set, as well as a shallow neural net for raw EEG data. With a real-time system in mind, we also tested different window lengths of the data aiming at the best payoff between short window length and high classification accuracy. Our results showed that the shallow neural net based on 4-second raw EEG data windows was best suited for real-time person-independent classification. The accuracy for the binary classification of internal and external attention periods reached up to 88% accuracy with a model that was trained on a set of selected participants. On average, the person-independent classification rate reached 60%. Overall, the high individual differences could be seen in the results. In the future, further datasets are necessary to compare these results before optimizing a real-time person-independent attention classifier for AR.

Keywords: EEG, attention, person-independence, eye tracking, augmented reality

Contribution Statement: Conceptualization of the project including the formulation of research aims and goals; Development and design of methodology; Implementation of the analysis; Discussion of the results; Writing and reviewing of the manuscript.

SSVEP-Aided Recognition of Internally and Externally Directed Attention from Brain Activity

by

Lisa-Marie Vortmann, Jonas Klaff, Timoo Urban, and Felix Putze

published in: IEEE SMC '21
Proceedings of the 2021 IEEE International Conference on Systems, Man, and Cybernetics, October 17–21, 2021, Melbourne, Australia
DOI: 10.1109/SMC52423.2021.9659098

Abstract

Steady-state visually evoked potentials (SSVEP) are a widely used paradigm for the detection of attended objects. However, their aid in the recognition of other attentional states has not yet been studied in detail. In this study (n=21), we assessed the benefits of including SSVEP stimuli as probes in a screen-based task to classify internal and external attention based on 16-channel EEG data offline. Previous studies have shown that the distinction between these two attentional states based on brain activity is possible. We compared several SSVEP-stimulus settings with a baseline where no SSVEP stimulus was present. Different flickering frequencies and stimulus placements were evaluated for the possibilities of different experimental setups. We found that the influence of the stimulus on the classification accuracy is highly dependent on the settings. The Linear Discriminant Analysis (LDA) performance increased significantly when an SSVEP-evoking stimulus with a low flickering frequency was present in the center of fixation. As well as when a Canonical Correlation Analysis (CCA)-coefficient was added as the SSVEP-specific feature to a generic band-power feature set. A simple training-free, person-independent threshold approach for internal and external attention detection resulted in accuracies significantly higher than chance based on SSVEP-features that were calculated only on three occipital electrodes. These results show that such stimuli can aid the recognition of internal and external attention. Thus, they can be used in experiments or applications for a more robust detection rate. Specifically, they could improve SSVEP-based BCI paradigms by adding another level of attention-awareness.

Keywords: EEG, SSVEP, internal attention, external attention, Linear Discriminant Analysis, Canonical Correlation Analysis

Contribution Statement: Conceptualization of the project including the formulation of research aims and goals; Development and design of methodology; Supervision of the implementation, data recording and analysis; Discussion of the results; Writing and reviewing of the manuscript.

Usability Examination of EEG Electrodes in Proximity to the Ears for SSVEP Studies

by

Lisa-Marie Vortmann

Internal Preprint

Abstract

Steady-State Visually Evoked Potentials are frequently used in Brain-Computer Interfaces for their robustness and person-independence. The most reliable EEG signals to detect them are usually obtained with occipetal electrodes due to their placement above the visual cortex. However, this placement may not always be possible. In this preliminary study, we examined the usability of other electrode positions in temporal locations. A high-density EEG dataset of 11 participants was used to compare different electrode clusters to detect visual attention on one of five flickering frequencies. The results suggest that visual attention can be reliably classified based on EEG signals around the ear. In the future, Ear-EEG (cEEGrid) could be combined with Steady-State Visually Evoked Potential Studies for easy, suitable setups.

Keywords: SSVEP, cEEGrid, electrode placement, EEG

Contribution Statement: Conceptualization of the project including the formulation of research aims and goals; Development or design of methodology; Implementation of data analysis; Discussion of the results; Writing and reviewing of the manuscript.

Usability Examination of EEG Electrodes in Proximity to the Ears for SSVEP Studies

Lisa-Marie Vortmann
Cognitive Systems Lab
University of Bremen
Bremen, Germany
vortmann@uni-bremen.de

Abstract

Steady-State Visually Evoked Potentials are frequently used in Brain-Computer Interfaces for their robustness and person-independence. The most reliable EEG signals to detect them are usually obtained with occipetal electrodes due to their placement above the visual cortex. However, this placement may not always be possible. In this preliminary study, we examined the usability of other electrode positions in temporal locations. A high-density EEG dataset of 11 participants was used to compare different electrode clusters to detect visual attention on one of five flickering frequencies. The results suggest that visual attention can be reliably classified based on EEG signals around the ear. In the future, Ear-EEG (cEEGrid) could be combined with Steady-State Visually Evoked Potential Studies for easy, suitable setups.

Index Terms

SSVEP, cEEGrid, electrode placement, EEG

Introduction and Related Work

Brain-Computer Interfaces (BCI) that aim at detecting visual attention have a long tradition of using Steady-State Visually Evoked Potentials (SSVEP) [8]. A visual stimulus that flickers with a steady frequency evokes potential changes in the EEG. The neural response follows the frequency of the stimulus and can be detected with electroencephalography (EEG). SSVEPs are frequently used because of their good Signal-to-Noise-Rate (SNR) [6] and robustness [2]. Visual stimulation with a steady frequency elicits the strongest neural correlate in the visual cortex. Thus, the strongest amplitude effects can be recorded with occipetal and parietal electrodes [2]. Lately, Canonical Correlation Analysis (CCA) has proven to be efficient for the classification of SSVEPs [4].

Augmented Reality (AR) is a recent technology that combines the display of virtual content in real-world surroundings. One possibility of displaying such content is via a see-through screen that is placed in front of the eyes. Such devices lack control elements that can be found in other user technologies (i.e keyboards or controllers). Instead, the interaction is performed through voice or gesture control. The combination of BCIs and AR offers an interesting possibility to add information exchange between user and device. This could for example be done by displaying visual flickering objects in AR and detect the attention with EEG.

Crucial for a good performance of the SSVEP detection is the quality of the EEG recording. With the experimental setup of Head-Mounted Display AR studies in mind, the electrode placement on the back of the head can be unsuitable. Hence, a different robust and uncomplicated setup could be beneficial for AR Applications that make use of SSVEPs for attention detection.

In this preliminary study, we considered and tested the usability of other electrode locations. Specifically, electrodes placed in temporal locations around the ear were compared to occipetal positions. cEEGrid[1] developed a technology for unobtrusive EEG acquisition around the ear. The adhesive, multi-channel, lightweight sensor arrays that are placed around the ear decrease the time needed for setup and cleaning and make EEG more compatible for real-life settings. They offer an interesting opportunity for SSVEP-based AR BCIs. They have been successfully used in several studies with other neural activity patterns ([1], [5], [7]). Recently, Leel and Lee (2020) [3] used cEEGrid systems to decode visual responses in non-stationary setups. They also used SSVEPs as input stimuli. In their results, they report that cEEGrid data from different walking conditions can be used to classify three distinct SSVEP signals above chance level.

DATASET

The EEG SSVEP Dataset I of the MAMEM (Multimedia Authoring & Management using your Eyes & Mind) project were used for this study. It contains EEG signals with 256 channels captured from 11 participants (3 female, 1 left-handed, 25 - 39 years) executing an SSVEP-based visual task. Five flickering frequencies (6.66, 7.50, 8.57, 10.00 and 12.00 Hz) were presented in isolation.

The visual stimulus was presented on a 22 inch screen in front of the participant. A pink box in the middle of a black screen flickered for 5 seconds with one of the mentioned frequencies while the participant was instructed to keep a steady gaze at the screen. Intervals between stimulations were at least 5 seconds. In the first phase (called ”adaptation” by the original author), the 5 stimuli were presented randomly for 80 seconds with 5 seconds stimulation and 5 seconds of rest (see Figure 1). After a 30 second rest, the participant was presented with the stimuli in a block-wise design. Every block consisted of the presentation of one frequency 3 times for 5 seconds with a 5 second rest in between. Blocks were separated by 30 second breaks (see Figure 2).

The recordings were performed using the EGI 300 Geodesic EEG System with a 256-channel HydroCel Geodesic Sensor Net (HCGSN). A sampling rate of 250 Hz was used for capturing the signals. The refresh rate of the LCD screen was 60 Hz.
To synchronize the stimulus with the EEG recordings, markers were added to the signal. These markers (DINS) were produced with the help of the Stim Tracker model ST - 100 by Cedrus and a light sensor.
In this study, three recordings of every participant were used. For further details on the dataset and the performed recordings, see `http://www.mamem.eu/results/datasets/`.

[1]http://ceegrid.com/home/

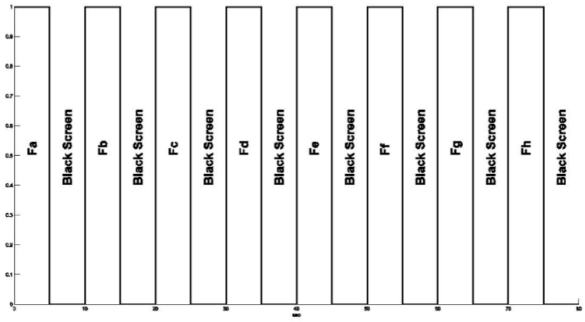


Fig. 1. Adaptation phase of 80 seconds with random stimulus order; Source: `http://www.mamem.eu/`

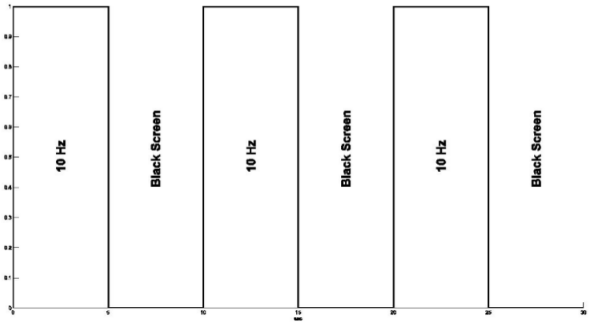


Fig. 2. Presentation design for a particular frequency. This was performed for every frequency with 30 rest in between; Source: `http://www.mamem.eu/`

METHODS

Preliminary to the classification of each trial, the EEG data was imported, examined for individual settings and preprocessed. The MATLAB-files that were offered by the original authors were transferred to Python. The analysis was performed in PyCharm using the MNE and Sklearn toolboxes.

Based on the DIN triggers that were recorded along with the EEG data, the trial starts were determined. Additionally, the DIN markers were used to detect the frequency that was used for the visual stimulation. Because of the slight differences in stimulation frequencies between participants, the frequencies that were used in the classification were detected individually in this first step. Each frequency for each trial was measured and each trial was assigned to one of the five original flickering frequencies. The frequencies for the classification were calculated by taking the average real frequencies in one category. This was done for each run of each participant.

During the cleaning of the EEG data, the signal was band-pass filtered between 1 Hz and 30 Hz with an infinite impulse response filter (IIR). All channels were re-referenced to average reference.

Based on the detected start of each trial, the data was cut into 3.5 second windows. The start of a window was 0.5 seconds after the onset of a trial to avoid the effects of neural responses that are evoked by the sudden stimulus onset.

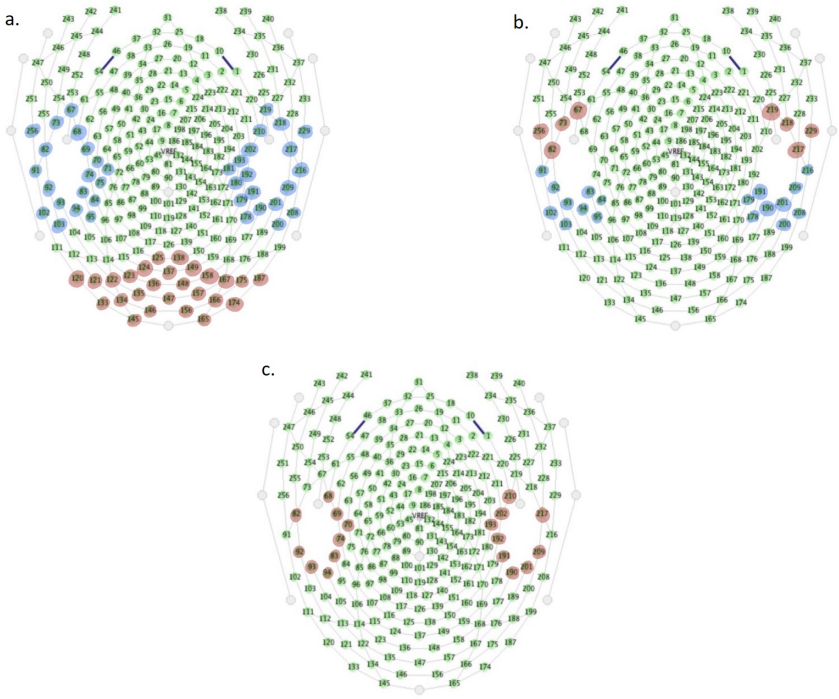


Fig. 3. Electrodes used for the calculation of subset classification accuracies; A. Blue = Ear-cluster, Red = Occipetal-Cluster; B. Blue = Behind the ear, Red = in front of the ear; C. imitating cEEGrid electrode placements

The described steps resulted in 33 datasets with 23 windows each. For each of the 11 participants, 3 separate recordings were analyzed. For the classification, a CCA was performed. The generated signals used in the process were generated individually for each dataset based on the computed average real flickering frequency of each category. The class with the highest correlation was used as a prediction for the window. These predictions were compared to the original labels to calculate the classification accuracy for each dataset.
Several different electrode subsets were examined for their accuracy, as well as individual electrodes and the complete set of electrodes. Figure 3 shows the electrode placements of the selected subsets.

RESULTS

For the reported results, only the classification accuracies achieved in individual runs was computed. The confidence of the classification decision was not regarded. Due to the equal display of five different flickering frequencies, the chance level for correct classification is 0.2 for all trials.

The classification accuracy using a single electrode cluster containing all 256 electrodes resulted in accuracies in the range of 0.13 to 0.7 (for 23 trials) with an average of $0.38(\pm 0.13)$. The generously chosen electrode clusters around the ears and occipetal cortex were chosen as described in Figure 3.a. The cluster covering the visual cortex achieved the best accuracies with a mean of 0.81 ± 0.26 (range = $[0.17, 1]$). The cluster around the ears performed significantly better than a cluster of all electrodes (mean = 0.57 ± 0.23, range = $[0.22, 0.97]$, $p < 0.001$). The exact

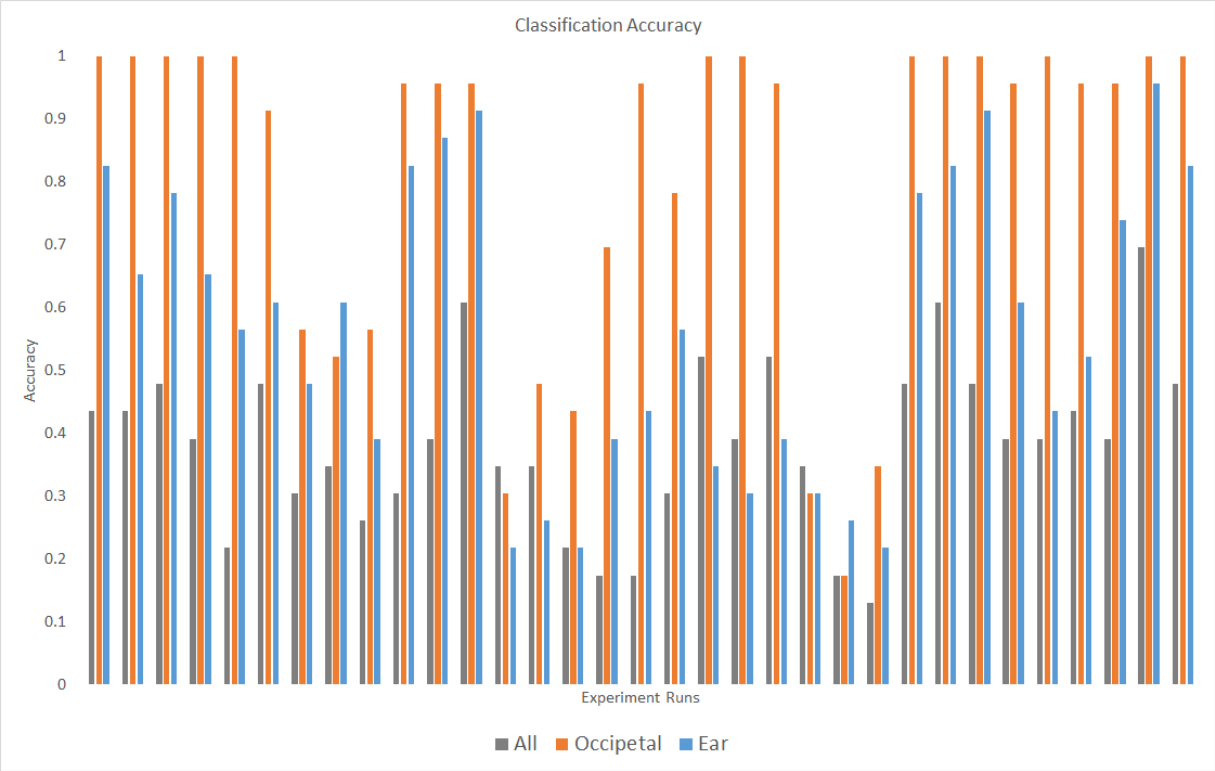


Fig. 4. Run-wise classification accuracies for all electrodes combined, occipetal electrodes and ear electrodes. Three runs always belong to the same participant.

classification results for every run are shown in Figure 4. Three participants had noticeably lower results in every run compared to the other participants (3, 5 & 8).

A direct comparison of the runs between the occipetal and the ear cluster shows that in 9.1% of the runs, the ear cluster achieves higher or equal results. In 18.2% the results are less than 10% below the results of the occipetal cluster (see Figure 5). A *Pearson's r* of 0.69 suggests a strong positive correlation between the classification accuracies.

In the next step, all electrodes were tested individually for their classification accuracy. The electrodes that performed best were all located in the occipetal area with accuracies up to 76%. For 14 of the 33 runs, a single electrode was sufficient to classify all 23 trials with an accuracy of 100%. A categorized visualization of the classification result from every single electrode can be seen in Figure 6. The distribution is almost symmetric with the worst results in fronto-temporal areas.

Based on the results from the single-electrode analysis, four new symmetric electrode clusters around the ears were chosen (see Figure 3.b). These allowed a comparison between the left and the right ear and electrodes in front of the ear or behind the ear. As seen in Figure 7 a dominance for one ear could not be established over all

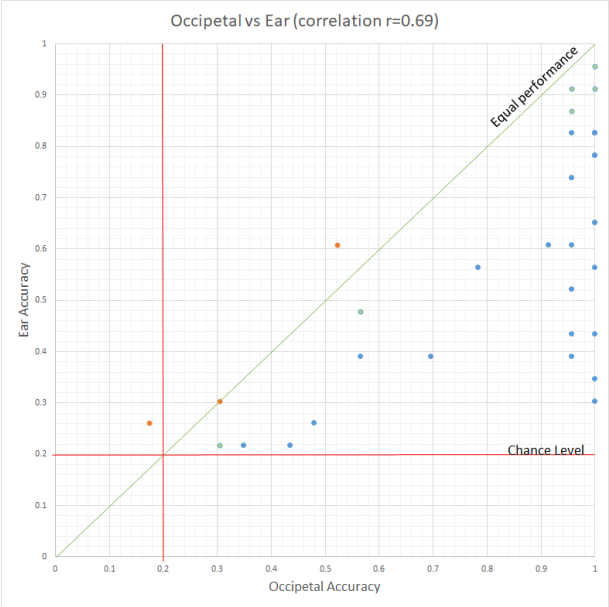


Fig. 5. Run-wise comparison of the classification accuracies for the occipetal electrode cluster and the clusters around the ears. Orange dots = Accuracy around the ear equal or better; Green dots = Accuracy around the ear less than 0.1 worse; Blue dots = Accuracy around the ear more than 0.1 worse

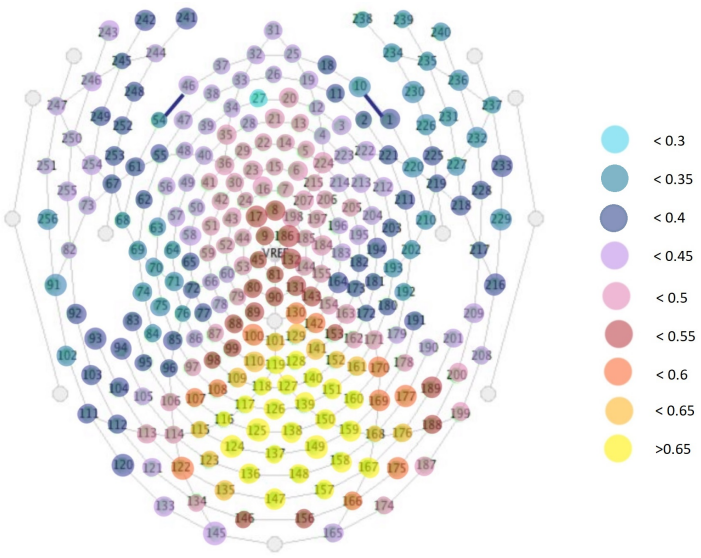


Fig. 6. Color-coded classification accuracies that were computed based on single electrodes.

participants but on average, the electrode clusters behind the ears performed better than in front of the ears (Right ear: Front = 0.4, Back = 0.55; Left ear: Front = 0.42, Back = 0.5). Additionally, a combination of both clusters behind the ears was tested, which lead to the highest results of 60% on average.

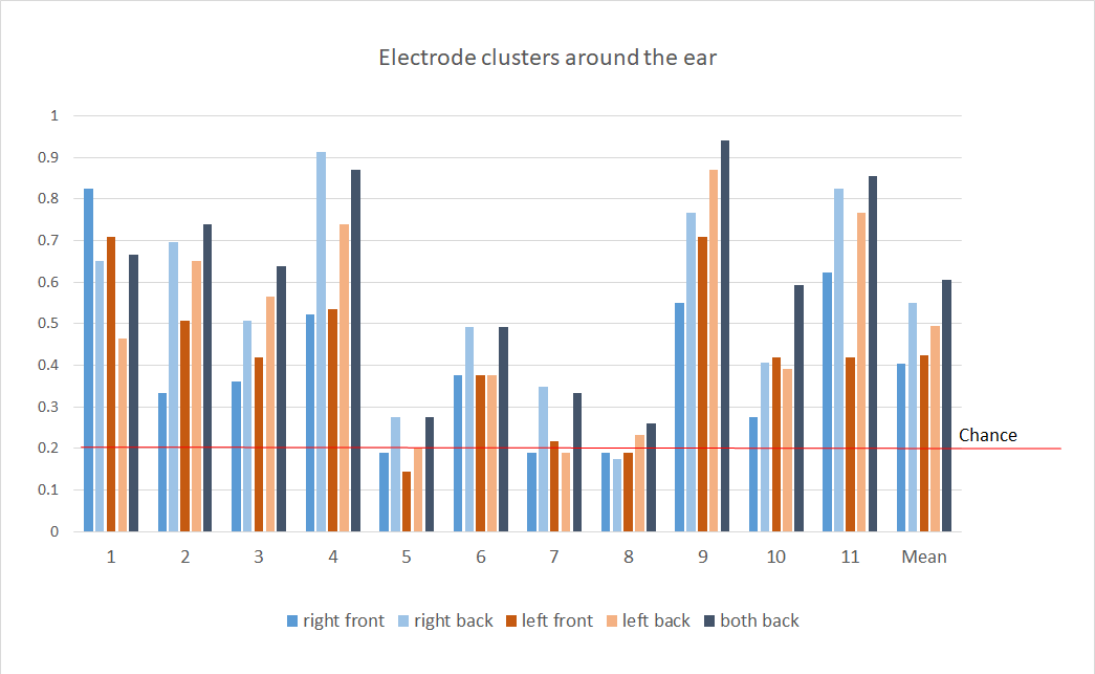


Fig. 7. participant-wise (combined runs) classification accuracies for both ears separated between electrodes in front of the ear and behind the ear

Following the goal of using cEEGrid systems for SSVEP studies in the future, an analysis was run with an electrode cluster that imitates the position of the electrodes in a cEEGrid system (see Figure 3.c). We assume that the electrodes closest to the ears of the 256 electrode cap would be in similar positions as a cEEGrid system. Figure 8 shows the classification results for each ear individually as well as combined for every experiment run. Overall, the combination of both ears improved the classification results. In one run, the accuracy even reached 100%. On average, the cluster performed with a 56% accuracy. However, this has more meaning once it is compared to other classification results of the same run because the dataset could not allow for much better classification.
Thus, we compared the cEEGrid imitation cluster results with the results of the occipetal cluster. The results are summarized in Table I. For one participant, the cEEGrid imitation cluster performed better than the occipetal cluster and for one participant it performed equally. On average the ear-EEG achieved 24% less accuracy. Relative to the performance of the occipetal cluster, the cEEGrid cluster had an average performance loss of 30%.

In a final analysis, we attempted an even closer cEEGrid imitation by leaving out a Common Average Rereferencing (CAR). Instead, we calculated bipolar channels of the two ears. The CAR was left out because in a cEEGrid setup, the information from the other electrode positions would not be available. The nine difference channels of homologous electrodes were used for the CCA. On average, the classification accuracy was 0.38. The accuracies for every run and combined for every participant can be seen in Figure 9.

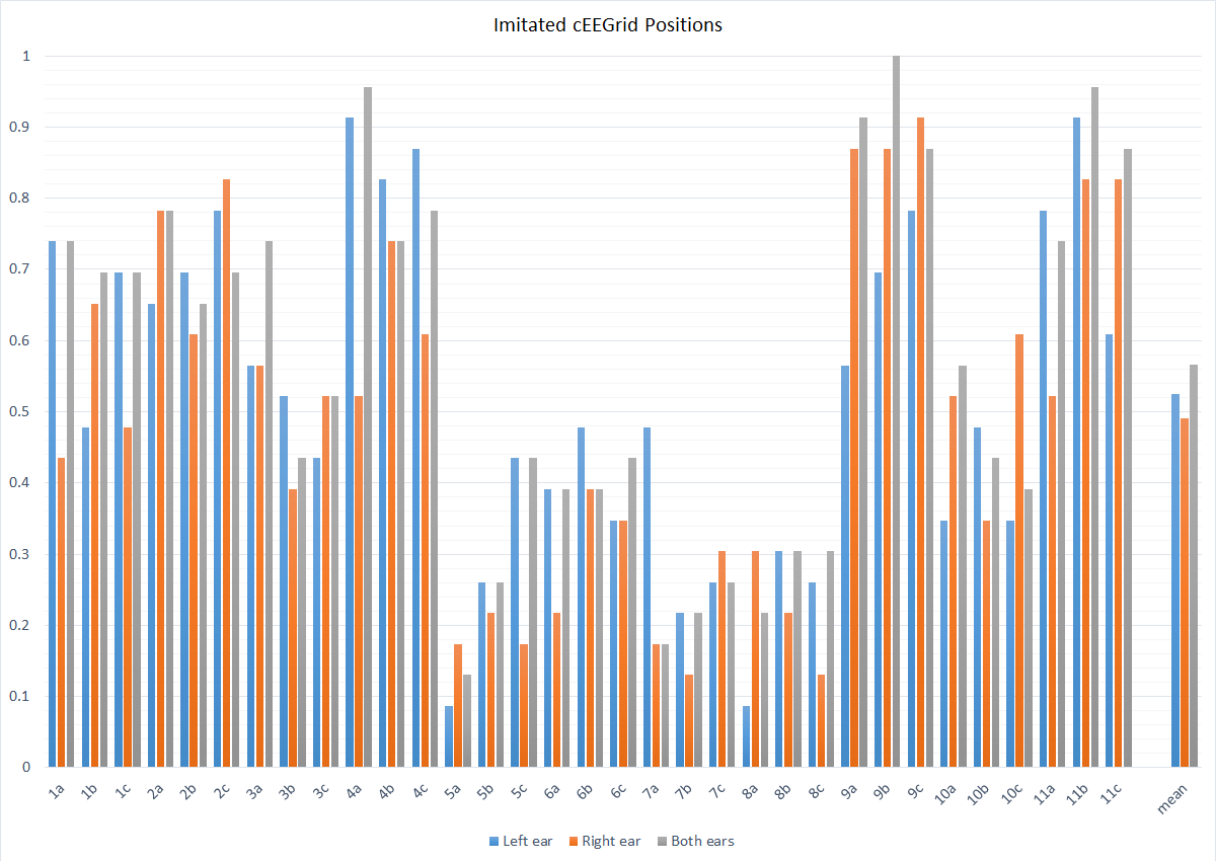


Fig. 8. Run-wise comparison of the classification accuracy achieved by the electrode clusters imitating cEEGrid systems. Seperated between left and right ear only or a combined cluster.

CONCLUSION

As a motivation for this study, we had SSVEP studies with Head-Mounted Augmented Reality Displays (HMD-AR) in mind. The typical design of such devices impedes electrode placements in occipetal locations. Additionally, wet electrodes that are placed in the hair of a participant decrease the usability of any AR-application. Ear-EEG (cEEGrid) offers a more comfortable setup. By replacing the traditional EEG caps with a cEEGrid System, AR SSVEP studies that aim at detecting visual attention would profit from shorter preparation times, less discomfort, and a more suitable real-life setting.

The results of this study and of Leel and Lee (2020) [3] show that EEG electrodes that are placed around the ear contain enough neural correlates of the stimulus flickering to decode the center of attention better than chance. The accuracy of electrodes on the occipetal cortex could not be reached with this dataset. However, the results offer an optimistic outlook on experiments that combine cEEGrid and Augmented Reality for attention detection.

In the future, we would like to perform such an experiment to test the accuracy and usability of the setup. The presumably lower classification accuracy than in traditional SSVEP setups could be made up for by the comfort of the setup.

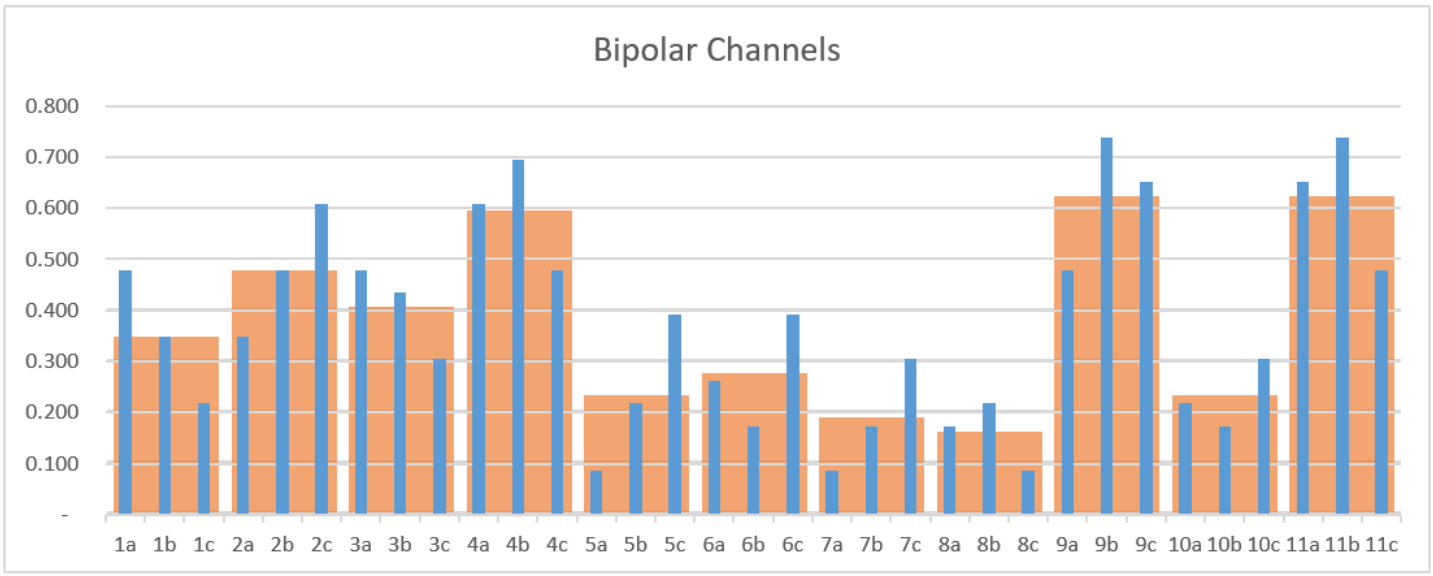


Fig. 9. Classification accuracies over all trials of one run, calculated without CAR, based on difference channels between both ears (blue). The average was calculated per participant (orange).

TABLE I

RESULTS OF THE IMITATED CEEGRIG CLUSTERS AND THE OCCIPETAL CLUSTER COMPARED. THE AVERAGE OF THE THREE RUNS PER PARTICIPANT WAS COMPUTED. TOTAL DIFFERENCE: THE ABSOLUTE DIFFERENCE BETWEEN THE CLASSIFICATION ACCURACIES. CALCULATED BY SUBSTRACTING THE AVERAGE ACCURACY OF THE CEEGRID IMITATION CLUSTER FROM THE AVERAGE ACCURACY OF THE OCCIPETAL CLUSTER. PERFORMANCE LOSS/GAIN: ASSUMING THAT THE OCCIPETAL CLUSTER ARCHIVES THE OPTIMAL PERFORMANCE POSSIBLE ON THE DATASET, HOW MUCH IS LOST BY USING THE CEEGRID IMITATION CLUSTER INSTEAD? CALCULATED IN PERCENT WHEN THE OCCIPETAL ACCURACY IS SET AS 100%.

No.	Both ears cEEGrid	Occipetal cluster	Total Diff. (Occ.- Ear)	Performance Change (Occ. as 100%)
1	0.71	1.00	0.29	-28.99 %
2	0.71	0.97	0.26	-26.87 %
3	0.57	0.55	-0.2	+2.63 %
4	0.83	0.96	0.13	-13.64 %
5	0.28	0.41	0.13	-32.14 %
6	0.41	0.81	0.41	-50.00 %
7	0.22	0.99	0.77	-77.94 %
8	0.28	0.28	-	±0.00 %
9	0.93	1.00	0.07	- 7.25 %
10	0.46	0.97	0.51	-52.24 %
11	0.86	0.99	0.13	-13.24 %
mean	0.57	0.81	0.24	-30.08 %

REFERENCES

[1] Martin G Bleichner, Bojana Mirkovic, and Stefan Debener. Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison. *Journal of Neural Engineering*, 13(6):066004, oct 2016.

[2] O. Friman, I. Volosyak, and A. Graser. Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces. *IEEE Transactions on Biomedical Engineering*, 54(4):742–750, 2007.

[3] Y. Leel and M. Lee. Decoding visual responses based on deep neural networks with ear-eeg signals. In *2020 8th International Winter Conference on Brain-Computer Interface (BCI)*, pages 1–6, 2020.

[4] Masaki Nakanishi, Yijun Wang, Yu-Te Wang, and Tzyy-Ping Jung. A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials. *PLOS ONE*, 10(10):1–18, 10 2015.

[5] Marlene Pacharra, Stefan Debener, and Edmund Wascher. Concealed around-the-ear eeg captures cognitive processing in a visual simon task. *Frontiers in Human Neuroscience*, 11:290, 2017.

[6] Ramesh Srinivasan, Fathallah Alouani-Bibi, and Paul Nunez. Steady-state visual evoked potentials: Distributed local sources and wave-like dynamics are sensitive to flicker frequency. *Brain topography*, 18:167–87, 03 2006.

[7] Annette Sterr, James K. Ebajemito, Kaare B. Mikkelsen, Maria A. Bonmati-Carrion, Nayantara Santhi, Ciro della Monica, Lucinda Grainger, Giuseppe Atzori, Victoria Revell, Stefan Debener, Derk-Jan Dijk, and Maarten DeVos. Sleep eeg derived from behind-the-ear electrodes (ceegrid) compared to standard polysomnography: A proof of concept study. *Frontiers in Human Neuroscience*, 12:452, 2018.

[8] Zafer İşcan and Vadim V. Nikulin. Steady state visual evoked potential (ssvep) based brain-computer interface (bci) performance under different perturbations. *PLOS ONE*, 13(1):1–17, 01 2018.

Machine Learning from Mistakes: Self-Improving Attention Classifier using Error-Related Potentials

by

Lisa-Marie Vortmann, Timo Urban, and Felix Putze

Under Review

Abstract

The detection of a person's attentional state via a Brain-Computer Interface opens up numerous possibilities, such as improving application usability or effectively triggering warnings in dangerous situations. Because EEG data varies significantly between individuals and changes during a recording, practical use of a Brain-Computer Interface is challenging. Often, prior to usage, each person's training data is collected, and an individual model is then trained for detection. This method of calibration will be replaced in this work by a self-improving online learning system. This involves personalizing a person-independent model for attentional state detection during runtime. The labels needed for adaptation are generated using automatically detected error-related potentials. We present the system that was developed based on pre-trained models of the two classifiers. This system was used to evaluate different strategies of adaptation and label generation. A statistically significant adaptation rate of 0.088 was achieved across all available subjects, based on simulations with pre-recorded data. These results suggest that person-dependent models for attentional state detection could in the future be substituted by self-improving classifiers that do not require a dedicated training data collection.

Keywords: online learning, EEG, ERP, error-related potential, attention, classification, CNN

Contribution Statement: Conceptualization of the project including the formulation of research aims and goals; Development and design of methodology; Supervision of the implementation, data recording and analysis; Discussion of the results; Writing and reviewing of the manuscript.

Machine Learning from Mistakes: Self-Improving Attention Classifier using Error-Related Potentials

LISA-MARIE VORTMANN, Cognitive Systems Lab, University of Bremen, Germany
TIMO URBAN, Cognitive Systems Lab, University of Bremen, Germany
FELIX PUTZE, Cognitive Systems Lab, University of Bremen, Germany

The detection of a person's attentional state via a Brain-Computer Interface opens up numerous possibilities, such as improving application usability or effectively triggering warnings in dangerous situations. Because EEG data varies significantly between individuals and changes during a recording, practical use of a Brain-Computer Interface is challenging. Often, prior to usage, each person's training data is collected, and an individual model is then trained for detection. This method of calibration will be replaced in this work by a self-improving online learning system. This involves personalizing a person-independent model for attentional state detection during runtime. The labels needed for adaptation are generated using automatically detected error-related potentials. We present the system that was developed based on pre-trained models of the two classifiers. This system was used to evaluate different strategies of adaptation and label generation. A statistically significant adaptation rate of 0.088 was achieved across all available subjects, based on simulations with pre-recorded data. These results suggest that person-dependent models for attentional state detection could in the future be substituted by self-improving classifiers that do not require a dedicated training data collection.

CCS Concepts: • **Human-centered computing** → **Interactive systems and tools**; Ubiquitous and mobile computing systems and tools; • **Computing methodologies** → **Online learning settings**; *Modeling methodologies.*

Additional Key Words and Phrases: online learning, EEG, ERP, error-related potential, attention, classification, CNN

1 INTRODUCTION

A brain-computer interface (BCI) enables direct communication between the human brain and a computer. Explicit user input via conventional interfaces (keyboard, voice commands, etc.) becomes obsolete [3]. Thus, BCIs are especially beneficial for individuals who have nerve palsies but they are also beneficial in areas where individuals are unable to perform user intputs due to other activities [8]. Apart from explicit communication via conscious interaction with a device, more implicit possibilities for adapting the behavior of machines to the corresponding user are thus provided.

Electroencephalography (EEG) is one technique for recording neural activity with a low latency. The collected data is then classified by machine learning and used as user input in a BCI. EEG signals recorded at the scalp can contain information about the user's mental state, including attention, fatigue, motivation, and error recognition [3].
While general patterns for certain cognitive states can be found across individuals, EEG signals differ in their detail and people respond differently to external stimuli. This complicates the task of developing generalizable models who are valid for everyone. The primary disadvantage of person-dependent models is that they must be trained for each new person before being used. As a result, a calibration phase is required for the collection of initial training data. Such training data typically requires labeled trials that indicate the current cognitive state of the user to train a person-dependent model. Such a lengthy calibration phase has an adverse effect on the practical use of such systems. On top of the inter-individual differences, during a recording, a user's EEG data may change, for example, due to a slight shift of the electrodes on the head [36]. This would also necessitate additional calibration phases during operation, preventing

Authors' addresses: Lisa-Marie Vortmann, vortmann@uni-bremen.de, Cognitive Systems Lab, University of Bremen, Germany; Timo Urban, Cognitive Systems Lab, University of Bremen, Germany; Felix Putze, Cognitive Systems Lab, University of Bremen, Germany.

the system from being used continuously.

As mentioned, labeled personal data can be used to train a model specifically for the individual but it can also be used to personalize an existing model. This can be performed as an initial adaptation (if individual data is available) or the person-independent model can be re-trained during use. This process is referred to as online learning and is already in use in a variety of applications [21]. Online learning may be used to circumvent the need for initial data collection and calibration, allowing for end-to-end adaptation of the general model while being used. One big challenge for this technique is determining the correct label of the data that is required to improve the classifier successfully.
We will demonstrate that an EEG-based BCI can learn from its own errors and thus improve a general model during use for each individual user, without the need for dedicated training phases.

In this work, we focus on the detection of external and internal attention. External attention is a term that refers to the concentration of attention on information perceived through the senses. Internal attention is focused on information that is already stored in memory or is generated through thought [13]. Such attention detectors can be installed in vehicles to improve safety, for example. Another application is to create attention-sensitive interaction systems for increasingly pervasive technologies (e.g., head-mounted augmented reality displays). By determining the user's current state of attention, the displayed information can be reduced to only task-relevant input [32]. On top of that, attention-adaptive BCIs can be used in the therapeutic field to treat psychological deficits associated with attention in a neurofeedback therapy [20].

Such EEG-based attentional state classification yields the best results using person-dependent machine learning models [32, 36]. However, as mentioned previously, these require the systematic recording of labeled training data, which takes a significant amount of time. The training data collection reduces the usability of such a system and, consequently, the benefits associated with such an attention-adaptive system.
To address this issue, we will personalize and thus enhance a generic, person-independent model for recognizing attention during runtime without the explicit knowledge of the user. The data for online learning, and thus for retraining the general classifier, will be generated based on the detection of error-related potentials (ErrP). Perceiving errors results in quantifiable electrical signals in the medial frontal and central brain regions [36]. These signals can be identified in EEG data by a distinct characteristic: a negative deflection occurs approximately 50–200 milliseconds after the error occurs, followed by a positive deflection after 200–500 milliseconds. The detection of ErrPs using machine learning can be used to perform error-based learning, correct erroneous instructions, or prevent them from being executed. ErrPs-based error correction has already been tested in a variety of application domains. For instance, in human-robot interaction or in brain-controlled prostheses [36].

While the user is interacting with the system, it displays attentional state-specific feedback. If, as a result of the misclassification, the feedback does not correspond to the user's actual attentional state, displaying it results in an error-related potential. When the correct feedback is displayed no error-related potential is triggered. On the basis of these potentials, newly recorded EEG data can be correctly labeled and used to improve the person-independent model for each individual.
With the proposed approach, an initial training data collection becomes obsolete and the usability of EEG-based attentional state classification systems improves significantly. Immediately after the setup, the system can be used

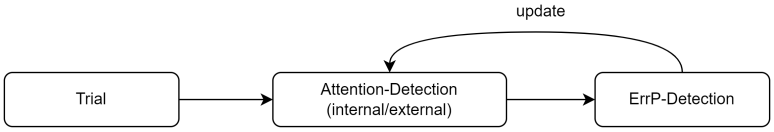


Fig. 1. The general idea of the implemented online learning system that updates the attentional state classifier based on the results of an error-related potential classifier that detects wrong system behavior.

out-of-the-box and it improves during runtime. This improvement is performed without explicit user feedback or interaction and thus happens without user-awareness.

1.1 Objective and Hypotheses

The objective of this study is to develop a system that adapts an EEG-based model for detecting external and internal attention at runtime using ErrP as a basis for online learning (see Figure 1). In other words, the system learns from its errors, which are communicated implicitly by the user. The system is expected to significantly improve recognition rates compared to other training-free person-independent models, making BCIs more practical for future attention recognition.

We pose the following underlying hypotheses: (1) Machine learning can detect error-related potentials that result from erroneous feedback, (2) retraining a person-independent classifier with person-dependent data personalizes the classifier and thus improves classification accuracy, and (3) real-time implementation of the proposed classifier that learns from error-related potentials is possible.

The primary work steps and challenges in this work were divided into four distinct stages. First, a suitable experiment paradigm had to be identified that elicits both internal and external attention and is capable of displaying specific feedback.
Following that, it was necessary to develop the appropriate classifiers. One classifier is required to determine the current attentional state, while another is required to determine the ErrP that will be used to label the new data. Training data were collected from EEG recordings in order to develop the two classifiers and train one person-independent model for each. These model-independent individuals form the foundation of the online learning system.
A third step evaluated various strategies for adapting the attentional state recognition model. This includes identifying the specific procedure for retraining the person-independent model and the most suitable data samples.
Finally, the acquired knowledge needed to be integrated into a functioning real-time system.

1.2 Related Work

The three fields of related work that are combined in this study are attentional state classification from EEG data, error-related potential detection in EEG data and online learning systems for BCIs. To the best of our knowledge, no previous work has combined these topics to implement and evaluate a real-time self-improving attentional state classifier that detects ErrPs.
The following works provide insight into appropriate data preprocessing, feature extraction, and classification methods, as well as an effective visual feedback and study design.

1.2.1 EEG-based Attentional State Classification. Numerous papers on the detection of attentional states have been published in recent years. Different types of attention have been considered, and various classification methods have been used. For instance, there is a distinction between internal and external attention [30–33], the differentiation of the attentional focus between two mental tasks [35], the recognition of attentional focus on real or virtual objects in Augmented Reality [34], or the classification of the attentional state into different levels [8, 17].

Wang et al. [35] developed a system for determining an individual's attentional state while driving a car. Participants were assigned at random to either drive a car or complete a math task in a virtual environment. Individuals' EEG data were segmented into 400ms overlapping windows and Power Spectral Density (PSD) features were generated for the delta, theta, alpha, and beta frequency bands. A Support Vector Machine (SVM) was used to classify each trial into one of the two tasks, achieving a recognition rate of 84.6%.
Gaume et al. [8] present an EEG-based BCI for monitoring various states of attention in their work. They identified three distinct levels of attention that correlate with the difficulty of a visual task. To perform the classification, an LDA was used with PSD features extracted from the EEG data as input. The authors discovered that reliably distinguishing between all three states was difficult and thus tested a pairwise differentiation of the three states. The runs were classified into 5s and 30s segments for this purpose, reaching recognition rates of 85% and 75% respectively.
Similarly, Mohammadpour and Mozaffari [17] distinguish between four distinct levels of attention. Four mental tasks are used to generate varying degrees of attention for this purpose: Closing the eyes and relaxing, silent reading, solving a math problem, and a Continuous Performance Task (CPT) in which the participant perceived alternating visual stimuli and was required to press a key on the keyboard in response to certain stimuli. A shallow Artificial Neural Network (ANN) with only one hidden layer was used for classification and Discrete Wavelet Transform (DWT) features from the delta-, theta-, and alpha frequency bands were computed. DWT features combine frequency and time domain information and are frequently used in place of PSD features when the temporal perspective is of particular interest. Their results confirm that reliably detecting more than two attentional states is difficult. All analysis were performed in a person-dependent fashion and only the resting state was reliably detected, with an average detection rate of 79.75%.

As mentioned, this work focuses on the detection of internally and externally directed attention. Chun et al. [5] developed a complete taxonomy covering both classic and contemporary disputes on the subject. They acknowledge the field's complexity and ubiquity and provide an ordered framework, covering related studies on detectable neural differences. The effects include increased alpha activity in times of internally directed attention [1, 6, 25], as well as increased delta band activity [10]. Building up on the work of Putze et al. [24] who reliably classified internal and external attention on a single-trial basis in a person-dependent fashion, Vortmann et al. [33] developed a BCI for real-time internal and external attention classification based on 32-electrode EEG and binocular eye tracking data. During the EEG analysis, pre-processing steps that are not feasible in an online scenario were avoided. PSD features were computed for all channels using the alpha, beta, and theta frequency bands and used for classification with an LDA. In a person-dependent offline evaluation of the ten recordings, very different results were obtained, with individual recognition rates ranging from 56% to 81%. On average, a recognition rate of 70.58 percent was obtained, demonstrating the feasibility of EEG-based person-dependent recognition of internal and external attention. Real-time classification was also demonstrated successfully by training on a person's first 40 trials and then classifying the remaining trials (person-dependent classification). A recognition rate of 60.67% was achieved.

Vortmann et al. [30] and Vortmann and Putze [32] conducted additional research into the person-independent recognition of internal and external attention in the context of augmented reality (AR). In this study, EEG data and eye movements were recorded from 14 participants using an eye tracker, and different sections were tested for the classification of individual passages ranging in length from 1 to 13 seconds. Additionally, both linear and non-linear algorithms were evaluated, as well as various neural networks. They found that the best result of 60% accuracy for person-independent recognition can be obtained using a shallow Filter Bank Common Spatial Pattern Neural Network (sFBCSP-NN) on 4-second data windows.

Thus, person-independent BCIs for detecting attentional states are possible to develop. However, achieving high detection rates in this manner is challenging. Numerous commonalities could be extracted from the related work regarding the preprocessing steps for EEG data preparation and specific techniques for attention detection. For instance, the focus of activity in the frontal lobe and the use of frequency filtering to narrow the frequency range.

1.2.2 Error-Potential Detection. Applications for EEG-based detection of error-related potentials have already been developed in a variety of fields, including human-robot interaction and error detection in virtual reality (VR) applications.

Si-Mohammed et al. [27] examined the occurrence of ErrPs in a VR environment using various types of feedback presentations. They defined three distinct types of feedback presentations that are prone to errors: erroneous motion when interacting with a virtual reality object, erroneous visual feedback, and erroneous background appearances. These three types of feedback were provoked randomly in the experiment and the EEG data was analyzed for the presence of ErrPs. It was discovered that ErrPs occurred significantly more frequently when the error was task-related. Additionally, the authors noted that the individual's focus should be on the error's source.

Kumar et al. [15] summarize significant research on EEG-based BCIs for ErrPs detection. For instance, the frontocentral and parietal regions of the brain have been identified as the primary sources of the ErrP signal. The associated activity is most intense in the theta and alpha frequency bands. The ErrP is reported to appear 50-500ms after the erroneous feedback is presented in EEG data. The autors state that the most accurate classification results were obtained when temporal and spectral domain features were combined. These claims are supported by Usama et al. [29] and Ehrlich and Cheng [7].
Both works demonstrated successfully the detection of ErrPs while imagining hand and foot movements, as well as the transfer of the imagined movements to a robot. The recognition rates of participants varied significantly. Ehrlich and Cheng [7] demonstrate a strong correlation between stimulus selection and ErrP recognition, as evidenced by highly divergent classification results across experiments. We conclude that the manner of feedback presentation is critical for achieving a high rate of ErrP recognition.

Similarly to the approach in this work, Yousefi et al. [36] presented an algorithm for detecting ErrPs in real time. In their study, participants were asked to complete four distinct cognitive tasks multiple times and in a random order. After each task, a feedback was computed in real time and reflects the BCI's assessment of which task the participant had just completed. If an ErrP was identified, the system displayed an alternative task guess. The ErrPs were classified using Regularized Linear Discriminant Analysis (rLDA), and the EEG data were cropped to a 1.4-second window following the display of feedback. A bandpass filter with a passband of 1–12 Hz was used to remove artifacts from the data, followed

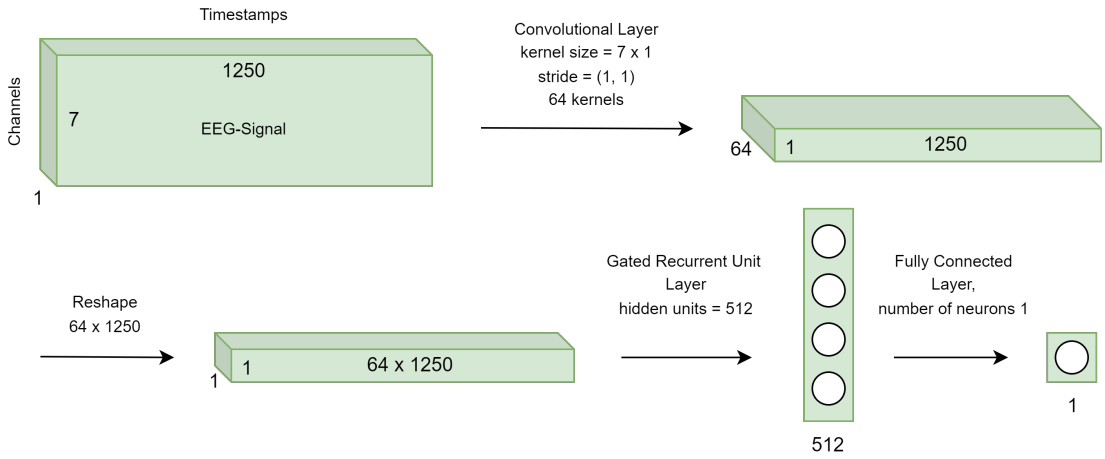


Fig. 2. Illustration of the architecture of the CNN-RNNet adapted from Tuleuov and Abibullaev [28].

by Independent Component Analysis (ICA). To perform ICA in real time, a three-minute baseline of EEG data was collected at the start of each recording and used to calculate ICA weights. On average, 0.74 accuracy can be obtained across all recordings and individuals.

An alternative deep learning approach for event-related potential (ERP) detection was suggested by Tuleuov and Abibullaev [28]. They compare a Shallow Convolutionas Neural Network (CNN), a Gated Recurrent Neural Network (GRU), and a CNN-RNNet hybrid solution comprised of a one-dimensional CNN and a GRU. The CNN-RNNet concept is to first learn a channel weighting to give EEG data from specific channels a greater influence on the classification in order to capture spatial features in the ERP data. Following that, the GRU is to extract temporal features from the data and finally to determine whether or not an ERP signal exists. ERP data were recorded with 16 EEG channels and filtered with a band-pass filter in the range of 0.5 to 12 Hz in an experiment involving eleven participants. The three architectures were compared to a standard method for classifying ERPs, namely an SVM approach. The three presented models outperformed SVM in this case and were able to learn temporal and spectral features from the data, reaching a recognition rate of between 83% and 84%. The authors recommend the CNN-RNNet as a best practice (see Figure 2).

The presented related work on this topic will inspire the design of the ErrP classifier. Moreover, they showed that visual feedback should be closely related to the task and presented in a way that is present to the participants so that attention is drawn to the feedback. Just like for the attentional state classifier, training person-independent models for ErrP detection will be a challenge. The results of the related work for person-independent classifiers were not as good as those obtained using person-dependent models. The ErrP recognizer was always calibrated before the use of the BCI to ensure that it is tailored to the individual and produces the best results.

1.2.3 Online Learning BCI Systems. We will implement an online learning system that steadily adapts the current model. According to Buttfield and Millan [3], dynamic adaptation of a BCI is dangerous and should be approached cautiously. Rapid and unpredictable changes may confuse the user or cause the classifier's performance to degrade.

The authors do, however, believe that dynamic adaptation during use is necessary for a BCI to be an effective tool. As discussed before, the EEG signals will change over time, both within and between recordings. They developed an online learning system that enabled them to slightly improve their classifier by detecting ErrPs. This was accomplished by first training the classifiers with each participant and then applying the online learning to another recording. Since the publication in 2006, the use of online learning systems increased due to more advanced machine learning techniques.

Kim et al. [12] presents an reinforcement learning based online learning system for the gesture control of robots. The robot is provided feedback about the performed action following the detection of an ErrP. Each participant completed five data sets, the first four of which were used to train the ErrP recognizer and the fifth of which was used to validate the online learning. A balanced accuracy of 91% was achieved in the online recognition using the pre-trained person-dependent ErrP models, which was sufficient for learning a mapping between human gestures and robot actions.
Luo et al. [16] presented another BCI based on reinforcement learning and ErrP recognition. Twelve participants participated in offline experiments in which they were asked to mentally choose between two possible robot movements. Following that, a simulated robot performed one of the two movements. If the system's choice was not consistent with the individual's decision, ErrPs should occur. These recordings were used to train the ErrP recognizer, and then online recordings with additional participants were conducted for the purpose of evaluating the BCI. The ErrP recognizer achieved an average performance of 67.49% and the average improvement following the reinforcement learning was 15.21%.

Very recently, Chiang et al. [4] created a system for detecting ErrPs in a Steady State Visually Evoked Potential (SSVEP) classifier. The purpose of this study was to demonstrate that a calibration phase prior to BCI use can be omitted in favor of adaptation via online learning. To accomplish this, a screen-based experiment with flickering left, right, and up arrows was used. The participants were asked to focus on an error and after a brief pause, a feedback shows which arrow was detected by the classifier, eliciting an ErrP for wrong feedback. For the online learning, the autors only used trials in which no ErrP was detected with a confidence level of at least 0.75. Multiple recordings of participants were made in order to pre-train the ErrP and SSVEP recognizers. Afterwards, individual runs of further recordings were divided into blocks, with all stored data being used to adapt the pre-trained model after each completed block. The autors found that the model can be adapted in a few runs and has a higher recognition rate than the model that is not adapted. The result is comparable to that of a calibration phase, suggesting that online learning may eventually eliminate the need for a calibration phase.

These previous works promised a successful implementation of the suggested self-improving system but also show the gap in the literature about combining the suggested attentional state classification with online learning.
As can be seen, the emphasis has recently shifted toward deep learning methods. These can be applied directly to raw EEG data, obviating the need for additional feature extraction, as feature extraction can be learned independently by the model architecture. Additionally, deep learning methods have demonstrated promising results for person-independent classification, which is why this work focuses on these methods in particular.

2 METHODS

The development of the proposed real-time online learning system required several recording and analysis iterations. The foundation of all of the recordings was a suitable experimental design that was inspired by previous works on

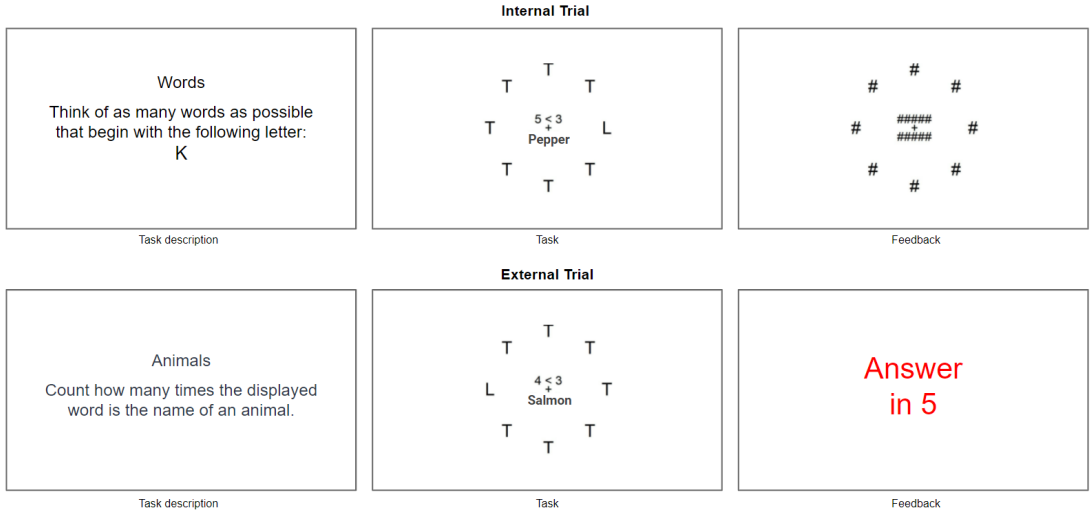


Fig. 3. Example for an internal and external trial with correct feedback.

internally and externally directed attention and error-related potentials. This experiment was then used without an online adaptation to collect a set of recordings for the preliminary offline analysis and person-independent model training. In a next step, the same offline available datasets were used to evaluate a pseudo-online learning system in a simulation. Finally, the implemented system was tested during real-time recordings.

2.1 Experimental Design

For the recordings, a screen-based experiment was developed that presents participants with a variety of tasks and feedbacks.

2.1.1 Attention Tasks. The experiment was based on the task design shown in Vortmann et al. [33]. The tasks, which require both internal and external attention from the user, are classified into three internal and three external task categories. Please refer to Vortmann et al. [33] for details on each task type. PsychoPy [23] was used to develop and execute the experiment.

Each run begins with a brief description of the upcoming task. Small images are alternately displayed on the screen at 0.8s intervals regardless of the task. Each trial lasts between 8 and 12 seconds. External tasks require these visual stimuli, but internal tasks do not. The Figure 3 shows two example trials.

The end of a run is accompanied by an acoustic signal, which enables participants to quickly identify the end of a task and to shift their focus away from it. They are asked to provide an answer to the external trials. Internal trials did not require an answer.

2.1.2 Feedback. Two distinct feedback mechanisms are used depending on the attentional state elicited by the task. For internal trials, the screen freezes after approximately 5 seconds to prevent the visual stimuli from changing until the end of the trial. The purpose of the freezing screen is to allow participants to focus more intently on the internal task without being disturbed by the changing content on the screen.

In the case of external tasks, a red text is displayed after 5 seconds to inform participants of the remaining time until the answer is requested (see Figure 3).

Both types of feedback are task-specific and are displayed centrally on the screen. In case a wrong visual feedback is presented, it has a direct effect on the task's processing. For external trials with an internal trial feedback, counting changes is no longer possible and for internal trials, participants will be surprised by being prompted for an answer. The participants should quickly become aware of this wrong feedback, which should result in an ErrP.

2.1.3 Experimental Procedure. The experiment includes 186 runs to ensure that sufficient data is collected for training the machine learning models. Each task type is used only once during the first six runs, and no feedback is provided. The remaining 180 runs include feedback and each task type is presented 30 times in a random order. For the collection of the offline available dataset, the displayed feedback was not based on the attentional state classifier. Instead correct feedback is displayed in two-thirds of runs for each task type, while incorrect feedback is displayed in one-third of runs. The runs with ErrPs are distributed evenly between external and internal tasks, thus totaling 60 runs with ErrPs.

After 100 runs, there is a self-paced rest period to allow the individual to recover. To increase motivation, a score was added to the tasks as an additional incentive for participants to complete them correctly. This is visible prior to each task and can be increased through successful completion of external tasks.

2.2 Data Recording

The development and evaluation of the proposed self-improving attentional state classifier required the recording of a designated EEG dataset. The previously described experiment was conducted with twelve healthy participants to generate an initial data set for offline use (mean age 32.3 years ± 16.63, age range [22,62], 3 male, 9 female, 0 diverse). All participants had normal vision or corrected-to-normal vision. Three datasets were excluded from further analysis due to technical problems and the resulting unusable data. The local ethics committee approved the experiment and all stored data was completely anonymized.

2.2.1 Procedure. Prior to each experiment, participants were informed about hazards and risks associated with EEG recordings and signed the consent form. Following an introduction video explaining the tasks, participants were given the opportunity to clarify any questions with the experiment leader and go over each task once in a tutorial. During the aforementioned break after 100 trials, participants completed an experiment-specific questionnaire, which was repeated following the conclusion of the experiment. Additionally, all participants completed the NASA Task Load Index [11] and the Mind-Wandering Questionnaire [19]. The experiment lasted an average of 58 minutes, with an additional preparation time of approximately one hour.

2.2.2 Apparatus. The participants' EEG was recorded using the actiCHamp system from Brain Products GmbH [2]. On the cap, the 32 gel-based electrodes were arranged in a standard 10-20 system with an additional ground electrode on the forehead. EEG data were collected at a sampling rate of 500Hz and impedance was kept below 20Hz. All recordings were conducted in a shielded room that minimizes electrical radiation from the environment, resulting in a lower level

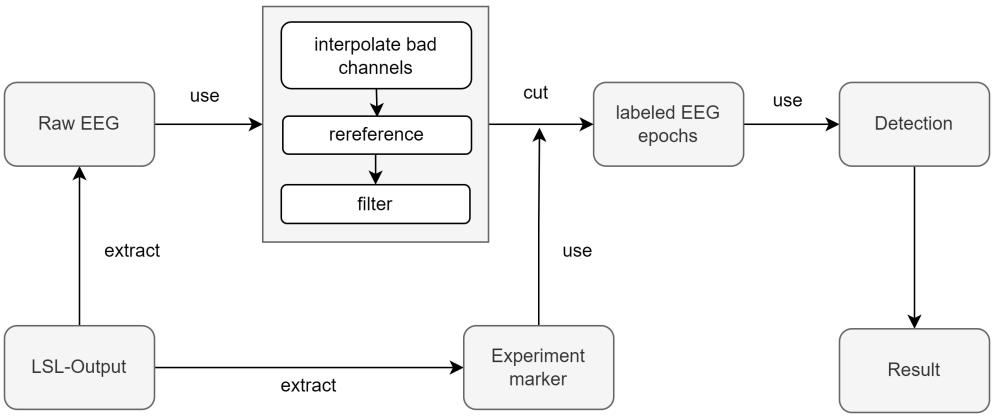


Fig. 4. EEG preprocessing pipeline

of interference effects in the recorded EEG signal.
The EEG and experimental data were recorded and combined using LabRecorder from the LSL framework [14].

2.2.3 EEG Preprocessing. The EEG preprocessing was performed using the MNE framework [9] and consisted of the basic steps of cleaning, trimming and labeling the data (see Figure 4).
The LSL output is used to extract all raw EEG data and experiment labels. Prior to further processing, extremely noisy channels were rejected following visual inspection and interpolated. On average, 0.33 electrodes were rejected. Next, channels T7 and T8 were set as reference channels before rereferencing the cleaned data, and the frequency range was bandpass-filtered using an MNE Finite Impulse Response (FIR) filter, following the discussed related work. These frequency ranges contain very little information about a person's mental state. Instead, they contain numerous interfering effects, such as the 50 Hz noise caused by electrical appliances. The exact filter frequencies for this work will be mentioned below.
The EEG data was segmented into appropriate windows using the experiment's markers, which include the start and end times of a run and the associated task, as well as the feedback onset. This epoching and further preprocessing steps were classifier-specific and will be explained in more detail below.

2.3 Classifier

Prior to developing the online learning system, the system's two most critical components were designed and evaluated. This section describes the machine learning models used to classify the attentional state and error-related potentials. These two classifiers constitute the backbone of the self-improving classification system. Without adequate recognition of error-related potentials, it is impossible to adapt the attentional state classifier successfully. With this in mind, the attentional state model must be able to produces accurate person-independent classification results while also being adaptable. The chosen model must utilize as little individual-specific EEG data as possible while still achieving an increase in recognition rate. The machine learning models were implemented using PyTorch [22].

2.3.1 Offline Training and Testing. Both classifiers were trained and tested offline on the described dataset prior to the evaluation of the online learning model.
Initially, we performed a person-dependent analysis to have baseline values for later comparison and to test the accuracy of the chosen classifiers in general. For this approach, each individual participants data was used for a 10-fold cross validation, with an 80:20 split into training and validation data. The splits were random but stratified to account for balanced classes. The reported results are the average achieved accuracy on the test sets over all 10 folds.

The person-independent models that will later be used in the online learning system were also trained and tested for each person based on the offline dataset. We chose a leave-one-out approach, meaning that the training of the ErrP and attention classifiers was performed on all but one datasets and the testing was performed on the remaining participants dataset. This was repeated for each participant. Again, the available training data was randomly split with a quote of 80:20 into training and validation data. It has to be noted that this approach means, that the person-independent model is different for each participant.
For the person-independent attention classification models, we additionally used the 10 dataset collected in Vortmann et al. [33]. The participants performed the same attention task but no feedback was given, thus, no data for an ErrP classifier was available. A preliminary analysis showed that the larger combined dataset increased the accuracy of the person-independent attention classifier.

2.3.2 Attention Classifier. The EEG data was bandpass-filtered between 1-40 Hz, because this frequency range is said to contain attention-relevant information [8]. The aforementioned epoching was used to cut the data into 5-second windows, starting 0.5 seconds are task onset.

For the classification of internally and externally directed attention, we used the Deep4Net from the Braindecode framework by Schirrmeister et al. [26] because it proved to achieve the highest accuracies in preliminary analysis. It is a deep convolutional neural network that works with raw EEG data as the input. For details on the network architecture please refer to the original paper.
During the network training we applied their suggested cropping strategy that cut the 2500 sample long EEG windows into 1979 overlapping crops of 522 samples. Each crop was classified independently and the final prediction was the result of a majority vote. The network was trained with a batch size of 64, a learning rate of 0.01, the Adam-optimizer, and for 100 epochs. These parameters were previously optimized using a gridsearch.

Offline Training and Testing Results. The chance level to correctly classify the attentional state with the given dataset is 0.5. On average, the person-dependent classification accuracy was 0.78 and the person-independent classification accuracy was 0.63. The individual classifier performances can be seen in Figure 5.
Generally speaking, the person-dependent model achieves a higher level of accuracy for all individuals. There is the highest difference between the two models for participants 004, 007, and 011, whereas the smallest difference was found for participant 008 with an accuracy of 0.59 for the person-dependent model and 0.55 for the person-independent model. The classifier performance of the individuals varies greatly. For example, in a person-dependent model, person 005 achieves the highest value with an accuracy of 0.91 and person 008 achieves the lowest value with an accuracy of 0.59. These findings are supported previous research that showed that in EEG-based classification, performance varies significantly between individuals, and a person-dependent model produces superior results [7].

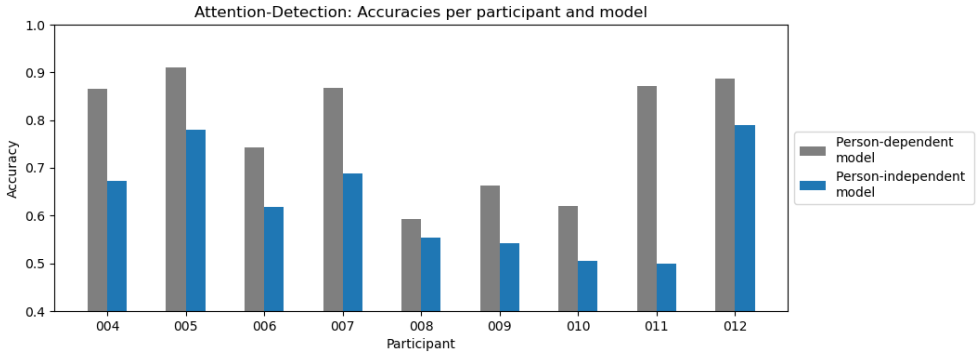


Fig. 5. Accuracies of the person-dependent and person-independent internal and external attention classification per participant

The chance level is not exceeded by the person-independent models of participants 010 and 011. However, because the results of all person-dependent models are significantly above the chance level, it is possible to generate a model for detecting attentional states above the random level for each person in the dataset and the goal of personalizing the person-independent models seems reasonable.

2.3.3 Error-Related Potential Classifier. The EEG data was preprocessed by applying a bandpass filter to the frequency range 1 - 12 Hz, as demonstrated in previous work [15, 36]. Moreover, not all 32 channels are utilized, but only a subset of seven. This subset includes channels Fp1, Fp2, F3, F4, Fz, FC1, and FC2, which are located in the frontal lobe of the brain and have been identified as being relevant by Kumar et al. [15]. Each trial is cut into a 2.5-second window, beginning with the onset of the feedback. Typically, an ErrP signal occurs between 200 and 800 milliseconds after the erroneous feedback [15]. The time window was chosen to be slightly larger here, as it is possible that the individual will need a moment to process the feedback. As a result, the likelihood of missing an ErrP signal is reduced using a 2.5-second time range, as it would otherwise be outside the cropped window.

Along with spectral features, the temporal perspective of the EEG data is critical for ErrP detection, as mentioned previously. Thus, we chose a combination of CNN and Recurrent Neural Network (RNN) to identify the occurrence of ErrP signals in the data. We apadted the approach by Tuleuov and Abibullaev [28] that was presented in Figure 2. The architecture is designed in such a way that individual CNN channels are prioritized. As a result, the CNN should discover which channels contain the most information for detecting ErrPs. The EEG data is then fed into an RNN, which is supposed to identify ErrP signals while taking the temporal dimension into account.
The CNN-RNN network is trained over a period of 20 epochs. The learning rate is set to 0.001 and all other parameters were the same as for the attention classifier.

Data from one participant (008) had to be excluded from training the ErrP classifier, as the experiment-specific questionnaires revealed a misunderstanding of the task, which resulted in misinterpreted feedback. There were 1440 samples remaining from eight recordings, and the chance of correctly classifying them is 0.67 due to the feedbacks' uneven distribution.

Per participant, two distinct person-dependent and two distinct person-independent models were trained, one for incorrect feedback during internal attention and one for incorrect feedback during external attention. These are then used in the online system to determine the classification of the user's attention state. Previously, such a split had outperformed a joint classifier in tests.

Offline Training and Testing Results. The average accuracy of person-dependent models is 0.75, while the average accuracy of person-independent models is 0.71.

Comparing the internal and external models for six participants per analysis method reveals a significant difference between the internal and external models (see Figure 6). This distinction is evident in both person-dependent and person-independent analyses. Additionally, the difference is consistent across analyses for all participants except for participants 011 and 012. Because the relationship between feedback types is balanced, ErrPs are not significantly more classifiable in either feedback on average.

In the following, the two person-dependent and two person-independent models are considered jointly for each participant. The average accuracy of the person-independent model exceeds the chance level for all participants but participant 005. For six of the eight datasets, the person-dependent model performs better than the person-independent model. In general, the difference between the two methods of analysis is smaller than the difference between the two models for the attention classification. This could be due to the fact that the person-dependent models have a limited amount of training data. After separating the external and internal models, there are only 90 samples remaining for each model, which must be divided into training, validation, and test sets.
There is also some variation between individuals, but to a lesser extent than in attention recognition. Person 004 has the highest accuracy in the person-dependent analysis, at 0.82, compared to person 007, who has an accuracy of 0.71.

Due to the unbalanced label distribution in this data, we also evaluated other performance metrics. The ErrP classifer's person-dependent models have an average precision of 0.62 and an F1 score of 0.53. For the person-independent models, a precision of 0.6 and an F1 score of 0.52 were achieved. Because all models have a precision greater than 0.5, the number of correctly generated labels prevailed. But because the F1 score is lower, it shows that the models miss a significant number of ErrP signals. The results indicate that additional research should be conducted to determine whether all runs should be used for adaptation or only those in which an ErrP signal was detected.

2.4 Self-Improving Online Learning System

The results of the offline data analysis showed that person-dependent attentional state machine learning models outperform person-independent models. This supports the hypothesis that retraining a person-independent classifier with person-dependent data improves the classification accuracy (H2). Due to the aforementioned disadvantages of person-dependent models and initial training data collection, we proposed an online learning system that uses error-related potentials for self-improvement. Our findings for the ErrP classifier support the hypothesis that machine learning can detect error-related potentials that result from erroneous feedback (H1), which suggests that our idea is reasonable and the implementation of a classifier that learns from its mistakes is doable.

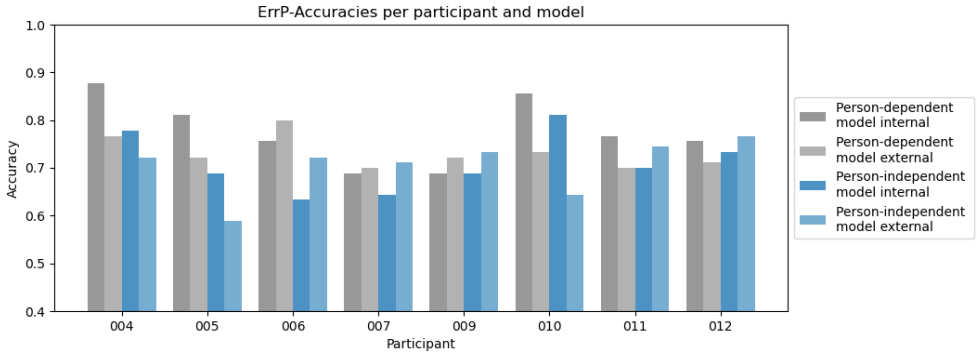


Fig. 6. Accuracies of the external and internal models from the person-dependent and person-independent analysis for the detection of error-related potentials per person and model.

After developing and validating appropriate methods and models for detecting attentional state and error-related potentials, they will be integrated into an online learning system. The overall goal of this system is to adapt the general person-independent model for attention recognition to personal data of each user during the runtime. To accomplish this, various strategies for retraining the CNN were compared. In the following, the system's requirements and design are described in detail. The system is capable of using both, new live recordings in real-time and offline existing recordings for a pseudo-online learning simulation.

2.4.1 Implementation. The implementation of the online learning system is highly personalized for the described experiment and the same steps are repeated for every trial. Four critical components of the online learning system can be used to visualize the information flow during each trial (see Figure 7). The experiment serves as the starting point and notifies the attention classifier that a task has started. Next, the pre-trained, person-independent attention classifier collects EEG data for 5.5 seconds and then determines which type of attention is required for the task. Following the prediction, the experiment displays the appropriate feedback in response to the decision. For pseudo-online learning simulations, the feedback is not adapted to the classified attentional state but was predefined. On top of that, the ErrP classifier requires the prediction of the attention classifier to chose the appropriate model for the ErrP detection, as well as the EEG data to classify the ErrP. If the pre-trained person-independent ErrP classifier recognizes an ErrP, the results are translated into labeled data and stored until enough new training data is available to retrain the attention classifier. Finally, the system starts retraining the model for attention classification in a background process. After each trial, the attention classifier determines whether a new model is available through online learning and replaces the existing one accordingly.

2.4.2 Retraining Strategies. Different strategies for retraining the attention classifier, as well as the label generation were implemented and compared. Methods from transfer learning were used to select the strategies [18].

For the training, the following three strategies were defined:
The *Continuous Training* strategy refers to training using the current model without making any additional adjustments. During the training process, it is possible to adjust the entire model or just the weights and biases.
The *Last Layer Training* method is limited to modifying the parameters of the final layer, which is responsible for

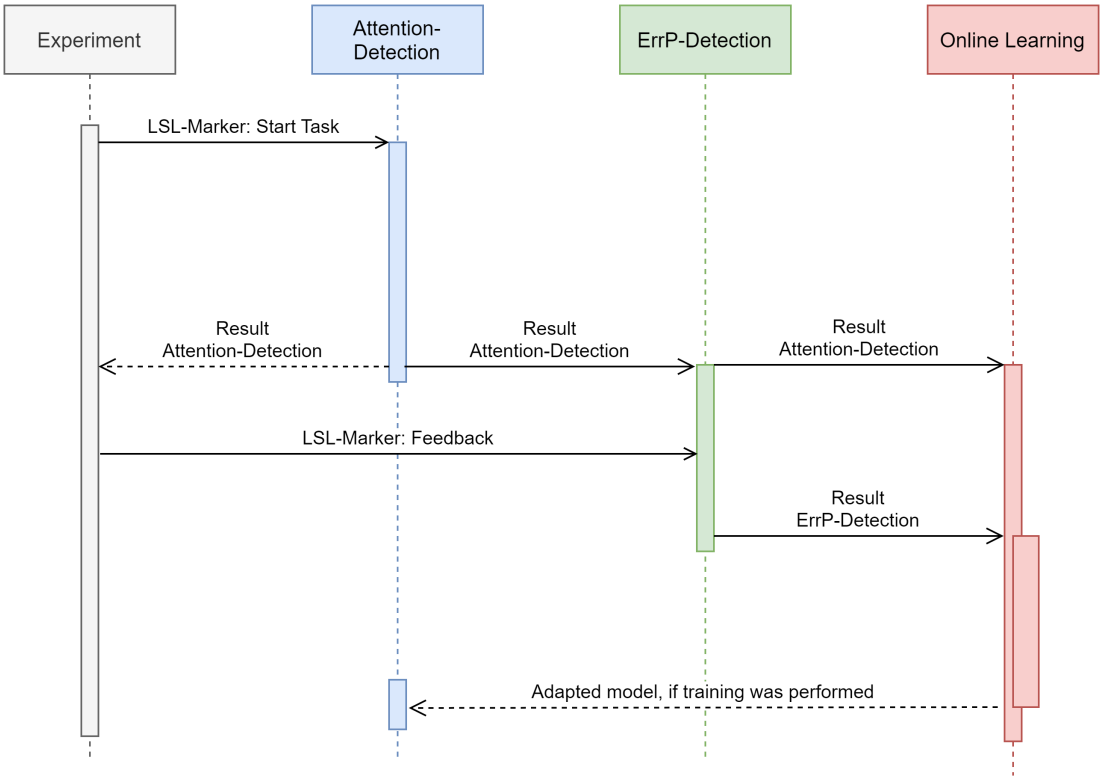


Fig. 7. The sequence model of a single trial in the online learning system.

decision-making. This strategy presupposes that the existing model is capable of extracting and weighting significant features from previous layers.
The strategy for *Last Layer Retraining* is similar. Only the final decision in the final layer, whether it is internal or external attention, is adjusted. However, prior to training, all weights and biases in the final layer are reinitialized. As a result, the decision process can be learned entirely from the new person-related data.

Generating labels for new samples is a critical component of online learning and is largely responsible for a successful adaptation. For this analysis, in addition to the ErrP-based strategy, the system also includes a label-based and a self-classifying strategy that are displayed in Figure 8.
The *ErrP-based strategy* was already presented as the main idea of this work. The newly generated personal data samples are labeled based on the predictions of an error-related potential classifier. We tested taking all new data samples for the retraining or only using the trials where an ErrP was detected.
The *label-based strategy* emulates a configurable-accuracy ErrP classifier. For instance, with a configured accuracy of 1.0, the label-based strategy correctly determines all labels. This is accomplished by utilizing the label provided by the experiment, regardless of whether the task is internal or external. As a result, this strategy is merely a comparison of theoretically achievable improvements. In a real-world scenario, it is not a substitute for the ErrP classifier.

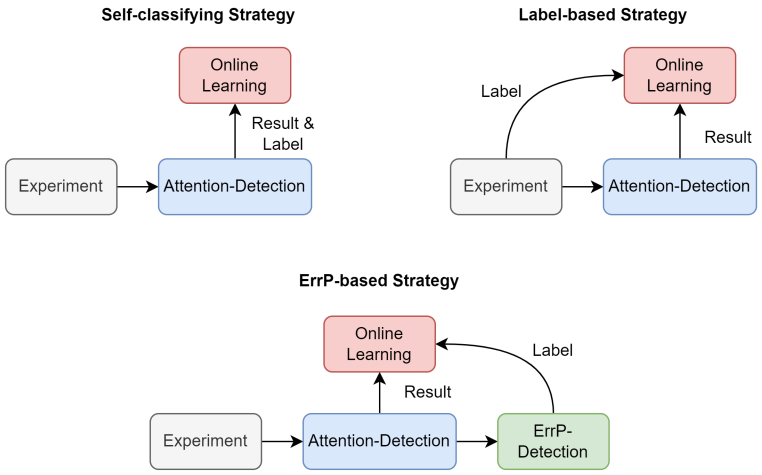


Fig. 8. The three labeling strategies that were implemented for reference and a discussion of the results.

On the other hand, the *self-classifying strategy* is an alternative to the ErrP classifier. The strategy simply uses the attention classifier's output as a label. Additionally, a confidence value can be specified to indicate the degree of certainty associated with the model's prediction. Thus, the model's prediction must be greater than this value in order for the label to be used for training. Where the model is too uncertain, samples are discarded. This strategy is used to determine whether the Deep4Net can improve its predictions on its own. This strategy is extremely straightforward and is intended to add a reference value to the ErrP classifier results.

In the live online learning system, only the ErrP-based labeling is used and the other two strategies are intended solely for the purpose of simulating pseudo-online learning in order to generate reference results. For the retraining approaches, the number of training epochs was reduced to 20 for a faster processing time.

2.4.3 Evaluation. The online learning system was evaluated using the previously recorded offline data for a real-time playback. It is necessary to demonstrate that adaptation is possible and which of the described adaptation strategies proves most successful. To accomplish this, an pseudo-online simulation using existing recordings was conducted. The biggest disadvantage of the simulation is that the shown feedback can obviously not be adapted in retrospective. However, the overall adaptation effects can be compared efficiently.
On the other hand, the system's real-time capability was to be demonstrated. For this purpose, the online system was tested in a pilot run.
For the pseudo-online learning based on the offline datasets, the previously described person-independent models were used for both the attentional state classifier and the ErrP classifier. Participant 008 was again excluded due to missing error-related potentials.

The following 19 configurations were tested in the pseudo-online learning simulation:

- Baseline - No adaptation (1 configuration)

- ErrP-based with Continuous Training, Last Layer Training and Last Layer Retraining. Each strategy is performed with adaptation based on all trials and based on only the trials with detected ErrP. (6 configurations)
- Label-based with Continuous Training, Last Layer Training and Last Layer Retraining. Each strategy is executed with the configured Accuracies of 0.6, 0.8 and 1.0. (9 configuration)
- Self-classifying with continuous training, last layer training and last layer retraining and a confidence value of 0.8. (3 configurations)

3 RESULTS

We will present the results of the questionnaires, pseudo-online learning simulation and real-time recordings, with a focus on the pseudo-online learning as it compromises all the suggested configuration comparisons. For all statistical analysis, a significance level of $\alpha = 0.05$ is assumed.

3.1 Questionnaires

The evaluation of the questionnaires provides insight into how participants perceived the experiment and assigned tasks.
The average of the participants' detected erroneous feedbacks is 58.33, which is very close to the actual value of 60 erroneous trials.
Four participants indicated that they were more capable of perceiving the image's erroneous freezing as a feedback type. On the other hand, two participants were able to detect the incorrect indication of an upcoming answer prompt more quickly. Three participants were unable to discern any difference in perception speed between the two types of feedback.
The participants were also asked to rank the difficulty and ease of various task types. Because responses varied significantly between participants, we conclude that no single task type was perceived as particularly difficult or easy.

The NASA Task Load Index identifies the demands placed on experiment participants. Along with the mental strain, participants were expected to exhibit a high level of performance and effort. Physical demand and level of frustration were deemed to be quite low. The participants' perceptions of mental and temporal demand were rather different.

The Mind Wandering Questionnaire results indicate that participants are frequently distracted in daily life. As a result, participants' minds frequently wander during lectures and presentations. Rather frequently, participants do not pay complete attention to the speaker and are thinking about something else at the same time. The fact that participants struggle to maintain focus during simple, repetitive tasks is critical for the experiment. This ability is precisely what is required during the experiment in order to perform the tasks correctly and detect erroneous feedback.

The results of the questionnaires are compared to the classification results in the discussion.

3.2 Pseudo-Online Learning Simulation

The pseudo-online learning simulation was performed on eight of the datasets recorded for this study, as described before.
To evaluate the improvement of the attention classifier, the last 45 trials will usually be compared with and without the personalization strategy. At this point, enough data should have been collected to adapt the person-independent model.

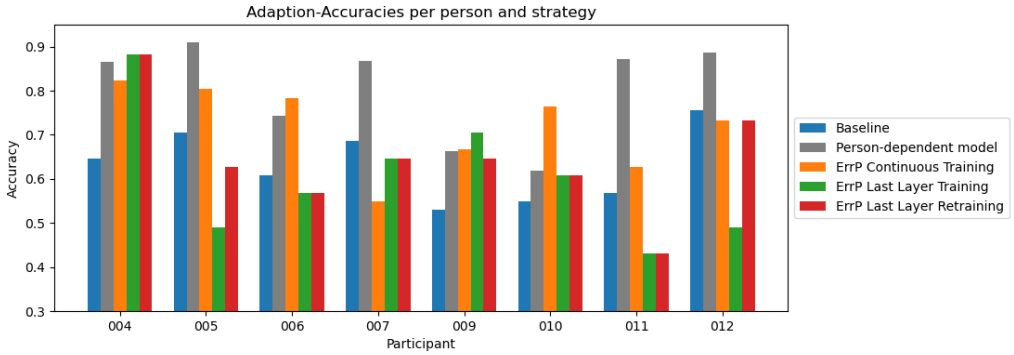


Fig. 9. The achieved accuracies for the last 45 trials using the ErrP-based labeling and each training strategy, compared to the person-dependent model and the baseline.

During the 135 trials before, the model was retrained 7 times depending on the chosen retraining strategy. We will compare the person-dependent model, the person-independent model (baseline) and all mentioned configurations.

3.2.1 Retraining Strategy Comparison. Figure 9 shows the achieved accuracies per participant during the last 45 trials for each training strategy using the ErrP-based labeling using all trials (with and without detected ErrPs). On average, these strategies achieved a higher accuracy than training with only classified ErrP signals in the trials.
The worst results were achieved following the Last Layer Training strategy (M = 0.6; SD = 0.13), and the Last Layer Retraining strategy (M = 0.64; SD = 0.12). However, the Last Layer Retraining was slightly better than the Baseline (M = 0.63; SD = 0.076). The highest accuracy on average, using an adaptation strategy was the Continuous Training (M = 0.72; SD = 0.09) but this was still significantly worse than the person-dependent model's average of 0.8 ± 0.1.

At least one strategy achieved a higher accuracy than the baseline for six of the eight participants. For participants 007 and 012, the baseline accuracy of 0.69 and 0.76 already achieved a high level of accuracy and the adaptation did not improve this. The results per participant vary significantly in terms of their degree of adaptation. On average, the Continuous Training strategy produced the best results across all individuals. As a result, this strategy will be used for the following comparisons.

We tested whether the ErrP-based continuous training strategy resulted in a significant improvement over the baseline using a one-sided paired t-test. The improvement in average accuracy of ErrP-based Continuous Training (M = 0.72; SD = 0.09) over the average accuracy of the baseline model (M = 0.63; SD = 0.076) was significant ($t(8) = 2.099$, $p = 0.036$).
Correspondingly, the significance of improvement within individual participants was assessed for the last 45 trials using the McNemar test (see Table 1). The improvement was significant for participants 004 and 006, with $p = 0.035$ and $p = 0.022$, respectively. All other participants showed no significant improvement. The McNemar test requires a very large difference from baseline for an improvement to be considered significant. Thus, only two participants improve significantly, despite the fact that six out of eight participants improve in accuracy over the last 45 runs. With a larger number of trials, it is expected that the improvement will be statistically significant.

Participant	004	005	006	007	009	010	011	012
p-value	**0.035**	0.267	**0.022**	0.324	0.265	0.080	0.664	1.000

Table 1. McNemar-Test for ErrP-based Continuous Training Strategy and the baseline model per participant

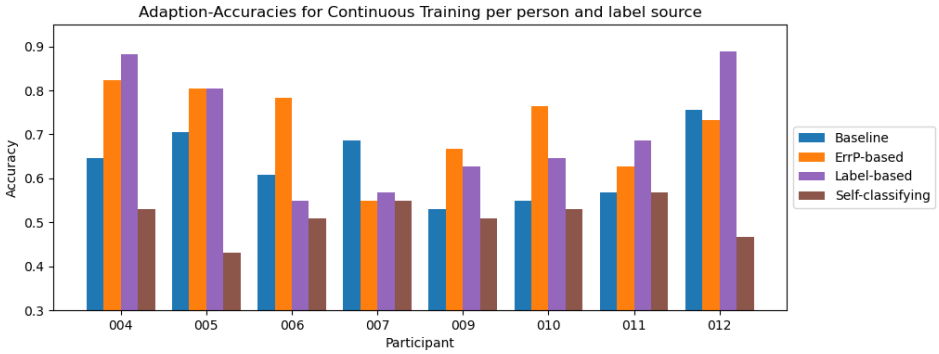


Fig. 10. The achieved accuracies for the last 45 trials per person using the Continuous Training Strategy and the different labeling strategies.

3.2.2 Labeling Strategy Comparison. Figure 10 shows the ErrP-based, label-based and self-classifying strategy compared against the baseline using the Continuous Training strategy. For the label-based strategy, the configuration with a precision of 1.0 was used. Thus, means that all runs with feedback displayed incorrectly were used for online learning. For the ErrP-based strategy, all runs were used for online learning as discussed before. The results show that the self-classifying strategy (M = 0.51; SD = 0.04) did not outperform the baseline (M = 0.63; SD = 0.076) for any individual during the last 45 runs. The self-classifying strategy resulted in a degradation of the original model for all participants. It is not suitable for personalizing the model. The lack of information about the correctness of the prediction is essential and cannot be compensated by the certainty of the prediction.

The ErrP-based (M = 0.72; SD = 0.09) and label-based (M = 0.71; SD = 0.13) strategies both show successful adaptations. Both were able to outperform the baseline in six of eight individuals. When looking at individuals, differences between the strategies can be seen. For example, the ErrP-based strategy achieved significantly higher accuracy in participants 006 and 008, while the label-based strategy achieved significantly higher accuracy in participant 012.

3.2.3 Accuracy Over Time. We examine the accuracy of selected adaptation strategies and the baseline over the course of the experiment for each participant. Three ErrP-based strategies are compared, with all runs being used for online instruction. Additionally, the label-based strategy with a precision of 1.0 is used, as is the self-classifying strategy.
The average accuracies over time for all participants and strategies are depicted in Figure 11. The plot begins after 16 runs, as all previous predictions were made using the same model and no adaptation may have occurred. With only a few personal trials (<60), the fluctuations in Accuracy are very strong. Accuracy no longer fluctuates significantly in the later stages, as the influence of individual predictions on accuracy decreases. As a result, adaptations become more difficult to detect near the end of the experiment.

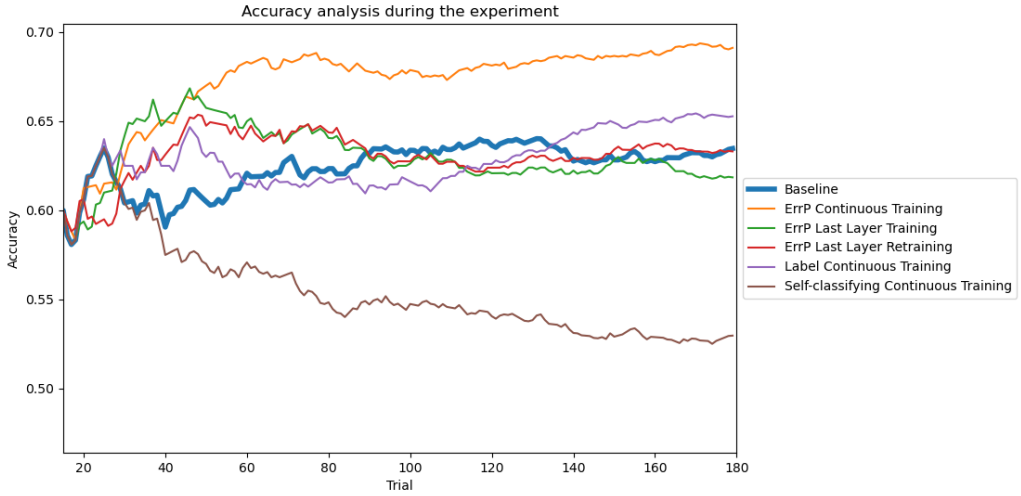


Fig. 11. The accuracy over the time of the experiment for different strategies, averaged for all participants

It demonstrates that the strategy based on ErrP-based Continuous Training achieves the greatest improvement. It is equally clear that the self-classifying strategy is ineffective and has a detrimental effect on the model.
Another observation is that the label-based strategy's accuracy increases as the experiment progresses. Due to the fact that this strategy uses only one-third of the runs for online learning, the model is retrained more slowly and infrequently.
Additionally, it is more evident here that the Continuous Training strategy quickly outperforms the baseline. Two iterations of retraining with a batch size of 16 is sufficient on average. The difference to the baseline increases almost continuously throughout the experiment and is greatest at the end.
On a more individual level, we observed that an improvement or deterioration in accuracy relative to the baseline occurs very early for all participants. In the majority of cases, this effect remains constant. Individuals' courses vary quite significantly. Expected progressions are seen in participants 004, 009, 010 and 011. For these individuals, a majority of the strategies perform better than the baseline. Participants 005, 007 and 012 already show a high accuracy by the baseline. For those participants, it was observed that adaptation strategies are frequently incapable of exceeding or maintaining this accuracy. Within the last 40 runs, both the Label-based and ErrP-based Continuous Training strategies have improved. Surprisingly, for participants 010 and 011, the baseline significantly improves after the first 90 trials.

3.2.4 Labeling Precision Comparison. Another interesting analysis is the effect of the precision chosen for the label-based strategies on the model's adaptation (see Figure 12). This analysis allows for conclusions about the performance requirements for a future ErrP classifier when online learning is limited to trials with recognized ErrP signals. The strategy with a label precision of 1.0 achieves the highest average accuracy (M = 0.71; SD = 0.13), and is capable of achieving a high degree of adaptation, particularly for participants 004 and 012. The average accuracy of 0.69 ± 0.09 is only slightly less than the label precision of 0.8. With an average precision of 0.57 ± 0.07, the strategy with a label precision of 0.6 is significantly less accurate and can only improve participants 006 and 009.

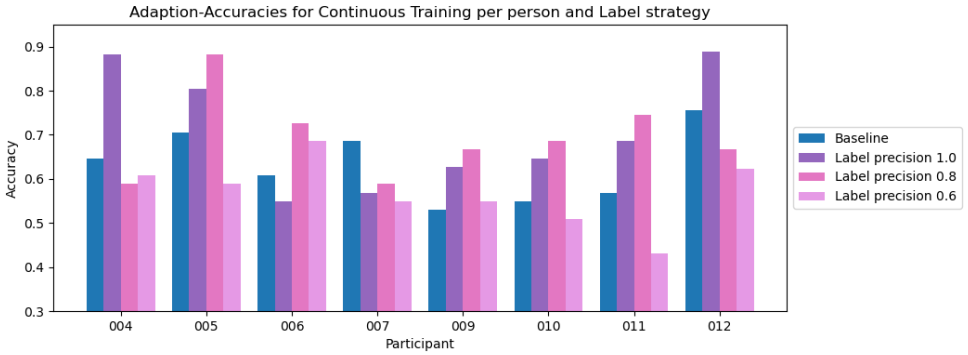


Fig. 12. The achieved classification accuracies for the last 45 trials per person using the Continuous Training strategy and different precision for the labeling.

3.2.5 Correlation Analysis. A Pearson correlation analysis is useful for explaining the factors affecting the adaptation. The rate of adaptation is calculated per participant by comparing the accuracy using the Continuous Training strategy to the accuracy of the Baseline using all trials. Correlations between the adaptation rate and the ErrP accuracies, the accuracies of the person-independent model, and the difference between the accuracies of the person-dependent and -independent models were calculated. All three factors are believed to have an effect on the rate of adaptation. The ratio of correctly generated to incorrectly generated labels for adaptation is determined by ErrP detection. The performance of the person-independent and person-dependent models provides insight into the adaptability potential. As a result, a correlation with the rate of adaptation is self-evident.

There is a strong correlation between the adaptation rate and the ErrP classifier's accuracy, with $r = 0.62$, $p = 0.1$. We also found a strong negative correlation between the adaptation rate and the baseline accuracy, indicating that the better the person-independent model is, the lower the adaptation rate is ($r = -0.62$, $p = 0.1$). Finally, a moderate correlation was found between the adaptation rate and the difference between the person-independent and -dependent models, with $r = 0.32$ and $p = 0.44$. Despite the strong correlation for the first two analyses, none of them were significant due to the small number of participants. This is expected to change as the population grows larger.
Consequently, we expect that a higher ErrP classifier accuracy and a lower baseline accuracy have the strongest effects on the adaptation rate, followed by the difference between the achievable accuracies using the person-dependent and person-independent models per person.
For the presented correlation analysis participant 012 had the most interesting results. Although the ErrP recognizer achieved a high accuracy of 0.75, no improvement in person 012 was observed as a result of the adaptation. This may be explained by the high accuracy of the person-independent model (0.79), as well as the small difference between the person-independent and -dependent models for this individual. The results of participant 004 were also interesting: Despite the person-independent model's accuracy of 0.67, a very high adaptation rate was achieved in this case. This could be attributed to the ErrP classifier's high accuracy of 0.75 and the person-dependent model's extremely high accuracy of 0.87.

3.3 Real-time Pilot Run

The real-time self-improving attention classifier was tested in in pilot runs to show the real-time mode of the system. The online system determines how the feedback is presented via the LSL stream. As a result, the internal and external feedback distributions are no longer identical, and incorrect feedback can be displayed depending on the performance of the attention detection model on an individual basis. In a preliminary test run, the online learning system was able to perform the classifications and training in real time for all recordings. There were no errors, delays in displaying feedback, or participant wait times.

4 DISCUSSION

The purpose of this work was to develop a self-improving attentional state classifier that recognizes its own errors via error-related potentials in response to attentional state-specific feedback. This goal was motivated by the requirement for training-free classifiers in order to improve their usability.
We hypothesized that error-related potentials could be detected for wrong feedback, personalizing a person-independent model improves the classification accuracy, and a real-time system is possible. We designed an appropriate experiment, collected user data, and classified the data using state-of-the-art machine learning algorithms. Internal and external attention were detected using a CNN, and ErrP signals were classified using a CNN-RNN network, which focuses more on the temporal dimension of the EEG data. When compared to a person-independent analysis, person-dependent analysis resulted in superior performance for all subjects in the study. This exemplified the potential of an online learning system that personalizes the person-independent model. Our models' performances are comparable to those found in related works. We developed an online learning system based on the trained person-independent models and demonstrated successful adaptation of attentional state classification. Several adaptation strategies were evaluated for this purpose, with the Continuous Training of the entire CNN achieving the best results. The results demonstrated a significant improvement over the baseline models.

The factors identified as influencing the success of personalization include the ErrP classifier's performance, the performance of the particular person's person-independent model, and the difference between the person-independent and person-dependent models. The aforementioned analyses were performed using pseudo-online learning simulations with existing recordings. Additionally, live adaptations were tested to demonstrate the system's real-time capability in a pilot study. Thus, all three hypotheses were supported by our results.

4.1 Experimental Design

The experimental design was adapted from previous research works and can be assumed suitable for internal and external attention classification. The main challenge was the design of an appropriate feedback that is obvious but also intuitive. The feedback was not understood by one participant which means it could be improved in the future.
The number of 180 runs and the duration of approximately one hour may have had a detrimental effect on participants' concentration, increasing the likelihood of errors. This conclusion is supported by the results of the NASA Task Load Index questionnaire, which revealed a high level of mental demand among experiment participants.

4.2 Classifier Results

The person-dependent models for attention recognition were trained and evaluated using data from individual participants. Over nine individuals, an average accuracy of 0.78 could be achieved. This value is comparable to the accuracy in Gaume et al. [8] of 0.75 for five-second runs with three mental tasks. Mohammadpour and Mozaffari [17] report an accuracy of 0.798 for detecting a resting state when compared to three other mental states in their work. The result in this work is most comparable to the accuracy of 0.706 obtained in the work of Vortmann et al. [33], as it was obtained using a similar experiment with the same tasks. However, they reported a higher classification accuracy using multimodal EEG and eye tracking data.

The average accuracy of 0.63 was achieved in the person-independent attentional state classification. This result can be compared to the result of Vortmann and Putze [32], where a shallow convolutional neural network was used to classify internal and external attention with an accuracy of approximately 0.6 for 4-second windows.
This shows that the results of the person-dependent and -independent analyses are comparable to those of previous work. The result for the person-dependent classification was even better than that of a similar work. The CNN used in this study appears to be suitable for classification of the attentional state in both person-dependent and person-independent ways.

An exciting finding is that the results of the person-dependent ErrP classifiers are insignificantly higher than those of the person-independent classifiers. However, the scarcity of person-specific data presents a challenge, as only 90 runs per person are available due to the model's division into an external and internal model. Accuracy can be increased by using more training data or by performing the classification with a single model.
Over eight subjects, the person-independent models achieved an average accuracy of 0.71. A person-independent accuracy of 0.84 was obtained in the work of Tuleuov and Abibullaev [28] who classified ERP signals and not ErrPs, which reduces the comparability.
This suggests that the ErrP classifier can be improved in the future. Nonetheless, it is advantageous that with the person-independent model, an online learning system's labels can be determined to be significantly more correct than incorrect on average.

The architecture of two ErrP models complicates their use and might not result in the creation of a general model for classifying ErrPs. It is difficult to determine whether the models are correctly detecting an ErrP signal in the EEG data as it could also detect a response to the visual feedback.

4.3 Online Learning System Performance

The online learning system's requirements were defined and a modular architecture was designed. On the basis of previously recorded data, the system was evaluated using various adaptation strategies and configurations. This approach was considered advantageous due to the high cost of acquiring new EEG data and the fact that only one online learning strategy can be used concurrently during an online recording. By simulating existing recordings, it is possible to evaluate and compare different strategies on the same dataset and under the same conditions.
One disadvantage of this method of evaluation is that the results of offline recordings are not entirely comparable to

those of online recordings. While the simulation attempts to replicate the online learning system in a realistic manner, the conditions are not identical. The ratio of runs with and without ErrP is always fixed and evenly distributed between external and internal tasks in offline recordings. Besides that, all offline recordings used have high-quality data, which is not always the case with online recordings. There, artifacts in the EEG data can have a significant impact on the performance of the online learning system. As a result, additional online recordings should be conducted to ascertain the system's efficiency. The analysis of simulated recordings was appropriate for evaluating various strategies and demonstrating the functionality of online learning.

The best results were obtained using the ErrP-based Continuous Training strategy, which had an average adaptation rate of 0.088 in comparison to the original model's accuracy. Luo et al. [16] report an average adaptation rate of 0.152. This is significantly higher, but the work records 750 trials per person which is significantly longer than in this work. A higher adaptation rate for the Continuous Training strategies over the Last Layer training strategies could suggest that the important features are person-dependent and need to be learned by the network.
The label-based adaptation strategy generates data for future work and may demonstrate the possibility of developing an ErrP recognizer with a precision of 1.0, 0.8, and 0.6. It is worth noting that the adaptation rates for 1.0 and 0.8 are quite similar. This means that achieving high adaptation rates does not require perfect precision of the ErrP classifier. On the other hand, 0.6 precision is insufficient for improvement.

As a result of the self-classifying strategy's performance, it is clear that this method is ineffective for adapting a person-independent model.

4.4 Future Work

In a future work, the live online learning system will be tested with the proposed configurations and several participants. In addition, it would be interesting to run multiple recordings with the same participants to explore adaptation behavior. It could be that the model needs to be readapted with each new recording or, on the other hand, that the model improves with each recording until a certain upper limit is reached. Where this upper limit lies and whether there are large differences between individuals would also be an interesting finding.
Further research should be done on the development of the ErrP recognizer, as there is still potential for improvement. Attempts should be made to develop a single model for classifying the two types of feedback. For example, a network from the Braindecode framework could be used, as an alternative to the CNN-RNN network. Possibly a more intuitive feedback could be found that participants can process more easily and does not require extensive explanation.
In general, it would be interesting to transfer the online learning system and the strategies used to a different problem.

Using techniques from Deep Reinforcement Learning would be another possibility for future work. These techniques would allow the model to be adapted in a different way and could conceivably allow for steady adaptation after each run, which is difficult with the current approach. Previously, training was done with batches and enough data for a batch had to be collected first. This step might no longer be necessary with reinforcement learning. The results from the related work presented look promising. Future work should therefore consider the use of deep reinforcement learning.

ACKNOWLEDGMENTS

REFERENCES

[1] Mathias Benedek, Rainer J Schickel, Emanuel Jauk, Andreas Fink, and Aljoscha C Neubauer. 2014. Alpha power increases in right parietal cortex reflects focused internal attention. *Neuropsychologia* 56 (2014), 393–400.

[2] Brain Products GmbH. 2021. actiCHamp. https://www.brainproducts.com/productdetails.php?id=74. Zugegriffen am: 13.11.2021.

[3] Anna Buttfield and Jose del R. Millan. 2006. Towards a Robust BCI: Error Potentials and Online Learning. *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society* 14 (07 2006), 164–8. https://doi.org/10.1109/TNSRE.2006.875555

[4] Kuan-Jung Chiang, Dimitra Emmanouilidou, Hannes Gamper, David Johnston, Mihai Jalobeanu, Edward Cutrell, Andrew Wilson, Winko W. An, and Ivan Tashev. 2021. A Closed-loop Adaptive Brain-computer Interface Framework: Improving the Classifier with the Use of Error-related Potentials. In *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*. 487–490. https://doi.org/10.1109/NER49283.2021.9441133

[5] Marvin M Chun, Julie D Golomb, and Nicholas B Turk-Browne. 2011. A taxonomy of external and internal attention. *Annual review of psychology* 62 (2011), 73–101.

[6] Nicholas R Cooper, Rodney J Croft, Samuel JJ Dominey, Adrian P Burgess, and John H Gruzelier. 2003. Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses. *International journal of psychophysiology* 47, 1 (2003), 65–74.

[7] Stefan K. Ehrlich and Gordon Cheng. 2019. A Feasibility Study for Validating Robot Actions Using EEG-Based Error-Related Potentials. *International Journal of Social Robotics* 11, 2 (01 Apr 2019), 271–283. https://doi.org/10.1007/s12369-018-0501-8

[8] Antoine Gaume, Gérard Dreyfus, and François-Benoît Vialatte. 2019. A cognitive brain–computer interface monitoring sustained attentional variations during a continuous task. *Cognitive Neurodynamics* 13, 3 (01 Jun 2019), 257–269. https://doi.org/10.1007/s11571-019-09521-4

[9] Alexandre Gramfort, Martin Luessi, Eric Larson, Denis Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, and Matti Hämäläinen. 2013. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* 7 (2013), 267. https://doi.org/10.3389/fnins.2013.00267

[10] Thalía Harmony, Thalía Fernández, Juan Silva, Jorge Bernal, Lourdes Díaz-Comas, Alfonso Reyes, Erzsébet Marosi, Mario Rodríguez, and Miguel Rodríguez. 1996. EEG delta activity: an indicator of attention to internal processing during performance of mental tasks. *International journal of psychophysiology* 24, 1-2 (1996), 161–171.

[11] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[12] Su-Kyoung Kim, Elsa Kirchner, Arne Stefes, and Frank Kirchner. 2017. Intrinsic interactive reinforcement learning – Using error-related potentials for real world human-robot interaction. *Scientific Reports* 7 (12 2017). https://doi.org/10.1038/s41598-017-17682-7

[13] Anastasia Kiyonaga and Tobias Egner. 2013. Working memory as internal attention: Toward an integrative account of internal and external selection processes. *Psychonomic Bulletin & Review* 20, 2 (01 Apr 2013), 228–242. https://doi.org/10.3758/s13423-012-0359-y

[14] Christian Kothe, David Medine, Chadwick Boulay, Matthew Grivich, and Tristan Stenner. 2021. Lab Streaming Layer. https://github.com/sccn/labstreaminglayer. Zugegriffen am: 14.11.2021.

[15] Akshay Kumar, Lin Gao, Elena Pirogova, and Qiang Fang. 2019. A Review of Error-Related Potential-Based Brain–Computer Interfaces for Motor Impaired People. *IEEE Access* 7 (2019), 142451–142466. https://doi.org/10.1109/ACCESS.2019.2944067

[16] Tian-jian Luo, Ya-chao Fan, Ji-tu Lv, and Chang-le Zhou. 2018. Deep reinforcement learning from error-related potentials via an EEG-based brain-computer interface. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 697–701. https://doi.org/10.1109/BIBM.2018.8621183

[17] Mostafa Mohammadpour and Saeed Mozaffari. 2017. Classification of EEG-based attention for brain computer interface. In *2017 3rd Iranian Conference on Intelligent Systems and Signal Processing (ICSPIS)*. 34–37. https://doi.org/10.1109/ICSPIS.2017.8311585

[18] Romain Mormont, Pierre Geurts, and Raphael Maree. 2018. Comparison of Deep Transfer Learning Strategies for Digital Pathology. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.

[19] Michael D. Mrazek, Dawa T. Phillips, Michael S. Franklin, James M. Broadway, and Jonathan W. Schooler. 2013. Young and restless: validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Frontiers in Psychology* 4 (2013).

[20] Henrique Oliveira-Junior, Wagner Casagrande, Fabiana Machado, Denis Delisle Rodriguez, Mariane Souza, Teodiano Bastos-Filho, and Anselmo Frizera. 2020. Towards an EEG-Based BCI System for Neurofeedback Assisted Rehabilitation of Attention Deficit Hyperactivity Disorder. In *Conference: X Congreso Iberoamericano de Tecnologías de Apoyo a la Discapacidad*.

[21] German I. Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. 2019. Continual lifelong learning with neural networks: A review. *Neural Networks* 113 (2019), 54–71. https://doi.org/10.1016/j.neunet.2019.01.012

[22] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural*

Information Processing Systems 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Eds.). Curran Associates, Inc., 8024–8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[23] Jonathan Peirce, Jeremy Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51 (02 2019). https://doi.org/10.3758/s13428-018-01193-y

[24] Felix Putze, Maximilian Scherer, and Tanja Schultz. 2016. Starring into the void? Classifying Internal vs. External Attention from EEG. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. 1–4.

[25] William J Ray and Harry W Cole. 1985. EEG alpha activity reflects attentional demands, and beta activity reflects emotional and cognitive processes. *Science* 228, 4700 (1985), 750–752.

[26] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping* (aug 2017). https://doi.org/10.1002/hbm.23730

[27] H. Si-Mohammed, C. Lopes-Dias, M. Duarte, F. Argelaguet, C. Jeunet, G. Casiez, G. R. Müller-Putz, A. Lécuyer, and R. Scherer. 2020. Detecting System Errors in Virtual Reality Using EEG Through Error-Related Potentials. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 653–661. https://doi.org/10.1109/VR46266.2020.00088

[28] Adilet Tuleuov and Berdakh Abibullaev. 2019. Deep Learning Models for Subject-Independent ERP-based Brain-Computer Interfaces. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*. 945–948. https://doi.org/10.1109/NER.2019.8717088

[29] Nayab Usama, Kasper Leerskov, Imran Niazi, Kim Dremstrup, and Mads Jochumsen. 2020. Classification of error-related potentials from single-trial EEG in association with executed and imagined movements: a feature and classifier investigation. *Medical & Biological Engineering & Computing* (08 2020). https://doi.org/10.1007/s11517-020-02253-2

[30] Lisa-Marie Vortmann, Felix Kroll, and Felix Putze. 2019. EEG-based classification of internally-and externally-directed attention in an augmented reality paradigm. *Frontiers in human neuroscience* (2019), 348.

[31] Lisa-Marie Vortmann and Felix Putze. 2020. Attention-aware brain computer interface to avoid distractions in augmented reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[32] Lisa-Marie Vortmann and Felix Putze. 2021. Exploration of Person-Independent BCIs for Internal and External Attention-Detection in Augmented Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 2, Article 80 (June 2021), 27 pages. https://doi.org/10.1145/3463507

[33] Lisa-Marie Vortmann, Moritz Schult, Mathias Benedek, Sonja Walcher, and Felix Putze. 2019. Real-Time Multimodal Classification of Internal and External Attention. In *Adjunct of the 2019 International Conference on Multimodal Interaction* (Suzhou, China) *(ICMI '19)*. Association for Computing Machinery, New York, NY, USA, Article 14, 7 pages. https://doi.org/10.1145/3351529.3360658

[34] Lisa-Marie Vortmann, Leonid Schwenke, and Felix Putze. 2021. Using Brain Activity Patterns to Differentiate Real and Virtual Attended Targets during Augmented Reality Scenarios. *Information* 12, 6 (2021), 226.

[35] Yu-Kai Wang, Tzyy-Ping Jung, and Chin-Teng Lin. 2015. EEG-Based Attention Tracking During Distracted Driving. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 23, 6 (2015), 1085–1094. https://doi.org/10.1109/TNSRE.2015.2415520

[36] Rozhin Yousefi, Alborz Rezazadeh Sereshkeh, and Tom Chau. 2019. Online detection of error-related potentials in multi-class cognitive task-based BCIs. *Brain-Computer Interfaces* 6, 1-2 (2019), 1–12. https://doi.org/10.1080/2326263X.2019.1614770 arXiv:https://doi.org/10.1080/2326263X.2019.1614770

Imaging Time Series of Eye Tracking Data to classify Attentional States

by

Lisa-Marie Vortmann, Jannes Knychalla, Sonja Annerer-Walcher, Mathias Benedek, and Felix Putze

published in: Frontiers in Neuroscience
2021, Volume 15, Article 664490
DOI: 10.3389/fnins.2021.664490

Abstract

It has been shown that conclusions about the human mental state can be drawn from eye gaze behavior by several previous studies. For this reason, eye tracking recordings are suitable as input data for attentional state classifiers. In current state-of-the-art studies, the extracted eye tracking feature set usually consists of descriptive statistics about specific eye movement characteristics (i.e. fixations, saccades, blinks, vergence, and pupil dilation). We suggest an Imaging Time Series approach for eye tracking data followed by classification using a convolutional neural net to improve the classification accuracy. We compared multiple algorithms that used the one-dimensional statistical summary feature set as input with two different implementations of the newly suggested method for three different data sets that target different aspects of attention. The results show that our two-dimensional image features with the convolutional neural net outperform the classical classifiers for most analyses, especially regarding generalization over participants and tasks. We conclude that current attentional state classifiers that are based on eye tracking can be optimized by adjusting the feature set while requiring less feature engineering and our future work will focus on a more detailed and suited investigation of this approach for other scenarios and data sets.

Keywords: Convolutional Neural Network, eye tracking, classification, Imaging Time Series, Augmented Reality, Gramian Angular Fields, Markov Transition Fields, attention

Contribution Statement: Conceptualization of the project including the formulation of research aims and goals; Development and design of methodology; Supervision of the implementation; Participated in the analysis; Discussion of the results; Writing and reviewing of the manuscript.

Combining Implicit and Explicit Feature Extraction for Eye Tracking: Attention Classification Using a Heterogeneous Input

by

Lisa-Marie Vortmann and Felix Putze

published in: MDPI Sensors
2021, Volume 21, Article 8205
DOI: 10.3390/s21248205

Abstract

Statistical measurements of eye movement-specific properties, such as fixations, saccades, blinks, or pupil dilation, are frequently utilized as input features for machine learning algorithms applied to eye tracking recordings. These characteristics are intended to be interpretable aspects of eye gazing behavior. However, prior research has demonstrated that when trained on implicit representations of raw eye tracking data, neural networks outperform these traditional techniques. To leverage the strengths and information of both feature sets, we integrated implicit and explicit eye tracking features in one classification approach in this work. A neural network was adapted to process the heterogeneous input and predict the internally and externally directed attention of 154 participants. We compared the accuracies reached by the implicit and combined features for different window lengths and evaluated the approaches in terms of person- and task-independence. The results indicate that combining implicit and explicit feature extraction techniques for eye tracking data improves classification results for attentional state detection significantly. The attentional state was correctly classified during new tasks with an accuracy better than chance, and person-independent classification even outperformed person-dependently trained classifiers for some settings. For future experiments and applications that require eye tracking data classification, we suggest to consider implicit data representation in addition to interpretable explicit features.

Keywords: eye tracking; attention; convolutional neural network; feature extraction; Markov transition fields; Gramian angular fields; heterogeneous feature sets; implicit feature learning

Contribution Statement: Conceptualization of the project including the formulation of research aims and goals; Development and design of methodology; Implementation of the data analysis; Discussion of the results; Writing and reviewing of the manuscript.

Multimodal EEG and Eye Tracking Feature Fusion Approaches for Attention Classification

by

Lisa-Marie Vortmann, Simon Ceh, and Felix Putze

published in: Frontiers in Computer Science
2022, Volume 4, Article 780580
DOI: 10.3389/fcomp.2022.780580

Abstract

Often, various modalities capture distinct aspects of particular mental states or activities. While machine learning algorithms can reliably predict numerous aspects of human cognition and behavior using a single modality, they can benefit from the combination of multiple modalities. This is why hybrid BCIs are gaining popularity. However, it is not always straightforward to combine features from a multimodal dataset. Along with the method for generating the features, one must decide when the modalities should be combined during the classification process. We compare unimodal EEG and eye tracking classification of internally and externally directed attention to multimodal approaches for early, middle, and late fusion in this study. On a binary dataset with a chance level of 0.5, late fusion of the data achieves the highest classification accuracy of 0.609 to 0.675 (95%-confidence interval). In general, the results indicate that for these modalities, middle or late fusion approaches are better suited than early fusion approaches. Additional validation of the observed trend will require the use of additional datasets, alternative feature generation mechanisms, decision rules, and neural network designs. We conclude with a set of premises that need to be considered when deciding on a multimodal attentional state classification approach.

Keywords: Feature Fusion, Convolutional Neural Networks, Attention, Eye Tracking, EEG, Markov Transition Fields, Gramian Angular Fields

Contribution Statement: Conceptualization of the project including the formulation of research aims and goals; Development or design of methodology; Implementation of the analysis; Discussion of the results; Writing and reviewing of the manuscript.

Real-Time Multimodal Classification of Internal and External Attention

by

Lisa-Marie Vortmann, Moritz Schult, Mathias Benedek, Sonja Annerer-Walcher, and Felix Putze

published in: ICMI '19 Adjunct
Adjunct of the 2019 International Conference on Multimodal Interaction, October 14–18, 2019, Suzhou, China
DOI: 10.1145/3351529.3360658

Abstract

The current attentional state can be divided into several categories, for example, the direction of attention. Often, this state is subconscious or its constant report impossible. Thus, an automated surveillance of the attentional state could be beneficial. In this paper, we performed a classification of multimodal data (EEG and eye tracking) to model internally- and externally-directed attention. 10 participants performed 6 different tasks of which 3 were associated with internal and 3 with external attention. In the first step, we showed that a combination of the two modalities led to an improvement of classification accuracy (average 72.67%) compared to single modality classifications. In a second step, the analysis was performed in real-time. The system was tested on one participant with an average accuracy of 60.87%. These results allow for an optimistic outlook on a reliable real-time multimodal classification system of internal and external attention.

Keywords: Multimodal data, classification, internal attention, external attention, EEG, eye tracking

Contribution Statement: Supervision of the analysis; Discussion of the results; Writing and reviewing of the manuscript.

Augmented Reality Interface for Smart Home Control using SSVEP-BCI and Eye Gaze

by

Felix Putze, Dennis Weiss, Lisa-Marie Vortmann, and Tanja Schultz

published in: IEEE SMC '19
Proceedings of the 2021 IEEE International Conference on Systems, Man, and Cybernetics, October 6–9, 2019, Bari, Italy
DOI: 10.1109/SMC.2019.8914390

Abstract

In this paper, we investigate the integration of eye-tracking and a Brain-Computer Interface into an Augmented Reality system to control a smart home environment. Through a head-mounted display, we present context-dependent control elements which the user selects by directing attention towards them. We show that the combination of both modalities leads to the most robust detection of selections and an interface which is accepted by its users.

Keywords: Augmented Reality, Smart-home, SSVEP, Brain-Computer Interface, eye tracking

Contribution Statement: Involved in the discussion of the results and reviewing of the manuscript.

Attention-Aware Brain Computer Interface to avoid Distractions in Augmented Reality

by

Lisa-Marie Vortmann and Felix Putze

published in: CHI '20 Extended Abstracts,
Conference on Human Factors in Computing Systems
April 25–30, 2020, Honolulu, HI, USA
DOI: 10.1145/3334480.3382889

ad

Abstract

Recently, the idea of using BCIs in Augmented Reality settings to operate systems has emerged. One problem of such head-mounted displays is the distraction caused by an unavoidable display of control elements even when focused on internal thoughts. In this project, we reduced this distraction by including information about the current attentional state. A multimodal smart-home environment was altered to adapt to the user's state of attention. The system only responded if the attentional orientation was classified as "external". The classification was based on multimodal EEG and eye tracking data. Seven users tested the attention-aware system in comparison to the unaware system. We show that the adaptation of the interface improved the usability of the system. We conclude that more systems would benefit from awareness of the user's ongoing attentional state.

Keywords: Attention; BCI; EEG; eye tracking; Augmented Reality

Contribution Statement: Conceptualization of the project including the formulation of research aims and goals; Development or design of methodology; Adaptation of the previously implemented systems; Data recording; Implementation of data analysis; Discussion of the results; Writing and reviewing of the manuscript.

Attention-Aware Translation Application in Augmented Reality for Mobile Phones

by

Lisa-Marie Vortmann, Pascal Weidenbach, and Felix Putze

Under Review

Abstract

As light-weight, low-cost EEG headsets emerge, the feasibility of consumer-oriented brain-computer interfaces (BCI) increases. The combination of portable smartphones and easy-to-use EEG dry electrode headbands offers intriguing new applications and methods of human-computer interaction. In previous research, Augmented Reality (AR) scenarios have been identified to profit from additional user state information - such as provided by a BCI. In this study, we integrated user attentional state awareness into a smartphone application for an AR translator. The attentional state of the user was classified in terms of internally and externally directed attention using the Muse 2 EEG headband with 4 electrodes. The classification results were used to adapt the behavior of the translation app, which uses the smartphone's camera to display translated text as augmented reality elements. A user study with 12 participants revealed that the EEG integration and system setup are promising paths toward mobile consumer-oriented BCI usage. For future studies, other use cases, applications, and adaptations will be tested for this setup to explore the usability.

Keywords: Translation, Augmented Reality, Brain-computer Interface, EEG, Smartphone, Attention

Contribution Statement: Conceptualization of the project including the formulation of research aims and goals; Development or design of methodology; Supervision of the implementation, data recording and analysis; Discussion of the results; Writing and reviewing of the manuscript.

Attention-Aware Translation Application in Augmented Reality for Mobile Phones

LISA-MARIE VORTMANN, Cognitive Systems Lab, University of Bremen, Germany
PASCAL WEIDENBACH, Cognitive Systems Lab, University of Bremen, Germany
FELIX PUTZE, Cognitive Systems Lab, University of Bremen, Germany

As light-weight, low-cost EEG headsets emerge, the feasibility of consumer-oriented brain-computer interfaces (BCI) increases. The combination of portable smartphones and easy-to-use EEG dry electrode headbands offers intriguing new applications and methods of human-computer interaction. In previous research, Augmented Reality (AR) scenarios have been identified to profit from additional user state information - such as provided by a BCI. In this study, we integrated user attentional state awareness into a smartphone application for an AR translator. The attentional state of the user was classified in terms of internally and externally directed attention using the Muse 2 EEG headband with 4 electrodes. The classification results were used to adapt the behavior of the translation app, which uses the smartphone's camera to display translated text as augmented reality elements. A user study with 12 participants revealed that the EEG integration and system setup are promising paths toward mobile consumer-oriented BCI usage. For future studies, other use cases, applications, and adaptations will be tested for this setup to explore the usability.

CCS Concepts: • **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; **Smartphones**; *Ubiquitous and mobile computing design and evaluation methods*; Empirical studies in ubiquitous and mobile computing; **User centered design**; *Interaction techniques*; • **Hardware** → Emerging tools and methodologies; • **Computer systems organization** → *Real-time systems.*

Additional Key Words and Phrases: Translation, Augmented Reality, Brain-computer Interface, EEG, Smartphone, Attention

1 INTRODUCTION

In our globalized age, many aspects of our lives, such as travel, advertising, business collaborations, signage, and information texts, are becoming increasingly international. Languages, however, remain very national and, along with culture, are often one of the most obvious differences between countries. They pose a big communication issue.
Encountering material in a foreign language, for example when traveling, makes information intake difficult and requires the use of a translator. While dictionaries were commonly used for such purposes in the past, translators on cell phones are now far more widespread in the digital age. In some circumstances, translating single words or sentences is sufficient; in others, entire documents must be translated into the native tongue. Copying such texts into a cell phone is time-consuming, and it becomes especially tough when the foreign characters differ greatly from your own. One solution to this problem are camera-based translators that can recognize the foreign language text and display translations using Augmented Reality (AR) [15]. Such AR translators can be used to swiftly construct a translated version of what the user is seeing. In real-time, text in an image is identified, translated, and replaced with visually corresponding translations. This produces an almost instantaneous AR illusion. It excludes the need for text to be typed and highly decreases the time required to obtain a translation. On the downside, this continuous updating of potentially salient visual content increases the chance for distraction from thought processes.

In this study, we explore how mobile AR translators for cell phones can be improved by adding attention-awareness. The attentional state of a user is estimated using an electroencephalography-based (EEG) Brain-Computer Interface

Authors' addresses: Lisa-Marie Vortmann, vortmann@uni-bremen.de, Cognitive Systems Lab, University of Bremen, Germany; Pascal Weidenbach, Cognitive Systems Lab, University of Bremen, Germany; Felix Putze, Cognitive Systems Lab, University of Bremen, Germany.

(BCI) the behavior of the application is adapted to the current state. It has been shown in several studies that an efficient use of AR can have positive effects on mental workload and task performance unless the distraction level of the virtual content is too high [21]. The virtual text overlay on the screen is usually updated regularly when using AR translators. This is required to compensate for minor unintentional camera movements and to update the translation in real time if new words are revealed. These updates can be a source of distraction, especially during times of internally directed attention (i.e. thought, mental task solving, memory) [8]. When users have their attention internally, the suggested system reacts by halting the translations, which would typically update continually.

Recent machine learning research showed that it is possible to separate attention into internal and external attention with a decent accuracy [45]. However, the applicability of this technology has mainly been demonstrated with BCIs that rely on a more or less stationary EEG setup, severely limiting the mobility of its users. Several companies have recently produced consumer-oriented EEG headsets that emphasize comfort and mobility. This technology represents a significant step toward the widespread application of BCIs. We considered an AR translator an appropriate use case scenario to demonstrate the enhancement of AR applications by adding attention-awareness with a mobile BCI because it is commonly used "on the road". Previous research has shown that the adaptation of an AR Smart-Home System based on internally or externally directed attention decreases the distraction and increases the usability compared to an attention-unaware system [46]. However, the study was performed inside a lab using a head-mounted AR device and an EEG cap with 16 electrodes that is meant for research purposes only. We will focus on a consumer-oriented BCI that is cheaper, has an easier setup and calibration, and can easily be used in the wild to improve the smartphone application.

The two main goals of this work were the implementation of the cell phone application in combination with the mobile BCI and a user study to evaluate our hypotheses about the application's improvement when attention-sensitivity is incorporated. The main contribution of this work is the purely smartphone-based setup using a light-weight consumer-grade EEG headset as the BCI data source. To test the system, we built the AR translator following the current-state-of-the-art apps. The adaptability to the user's attentional state was included as an optional parameter to compare two versions of the system - the attention-aware one and the standard one - in a user study. A travel-themed scenario was used for the study, in which participants had to read and understand foreign posters before answering questions on the material. The evaluations of the system will be based on achieved classification accuracies and user questionnaires rating the system's usability and level of distraction. The results will also provide insights regarding a design that accounts for false classifications. Assessing the potential benefits with the potential drawbacks and finding the right balance on when to adapt is key to maximizing the improvement.

2 RELATED WORK

Studies related to this work cover other AR translation applications, internal and external attention classification from EEG data, and mobile BCI setups. We are not aware of previous works that implemented and tested a mobile, light-weight BCI setup for cell phone applications that differentiates attentional states of the user.

2.1 AR Translation

Text can be translated from and to a language of choice using a translator, typically by inputting text manually. An AR translator directly projects translations onto existing text in the actual environment with the help of a display. The most well-known example is the Google Translate App [15]. There is no need for human text input because the translator detects text automatically. An AR translator must detect and recognize text, which is accomplished by optical

Table 1. All apps were found in the Google Play Store (Android) with the query "AR translate" and "Camera translate" on 26.02.2021. Many more translation apps are available that do not feature AR, so only the top ones were analyzed for this comparison. Langs. = Languages; RT = Real-Time

					Translation Mode		Text Replace	
App name	Author	Release	Downloads	Langs.	RT	Photo	AR	Text display
Google Translate	Google	2015 (AR)	500M+	109	yes	yes	yes	Replacement
Microsoft Translator	Microsoft	2015	50M+	22		yes	(yes)	Overlaid
Camera Translator	EVOLLY.APP	2017	10M+	56		yes	(yes)	Overlaid
Camera Translator	Fox Solution	2018	1M+	163		yes		Seperate
Translator for Texts, Websites & Photos	Octaviassil	2018	1M+	108		yes		Seperate
Cam Translate	Xiaoling App	2019	100 000+	28		yes		Seperate
Language Translator	Touchpedia	2021	50 000*	105		yes		Seperate

character recognition (OCR). Following that, the text is translated and projected back into position onto the original text, preferably with a text style that resembles the original for authenticity. To produce an authentic illusion of existence, the translated text must not only be applied in the precise location, but must also move with the camera and underlying text. This is accomplished by object tracking. The exact implementation of the application will be described in a later section.

The amount of research publications on AR Translators is limited. One of the latest papers concerning AR translators is from 2017 in which Tatwany and Ouertani [40] review AR use for text translation. They identified 12 papers relevant for their research. Most papers present translators that work by taking a picture. The picture is then analyzed with OCR and text is overlaid onto it, usually in a different location in a different style. Because translations are displayed on a fixed image, these translators cannot be considered true AR translators because they do not work in real-time.

Among all papers, only the TranslatAR (2011) [13] used real-time translations. Users translate text by tapping onto text, and translated versions with matching styles emerge at the same location after roughly one second. The translated text then moves in sync with the camera or the text, as if it were glued to the original text (textstickers). Using TranslatAR, textstickers move with the actual text, based on tracking. While HMD-based translators exist [41], smartphones as handheld AR devices dominate the research [40].

The latest publications on AR translators were from 2015 [37, 40], thus, it is interesting to see how consumer products evolved since. There are numerous consumer translation apps available now. Many of these allow you to translate text using your camera. Table 1 compares the top results from the Google Play Store. Some apps allow you to translate using your camera, however the translation is returned as plain text, which is shown independently from the image. As a result, these translators do not use AR to display translations. The majority of the apps evaluated provided this functionality, but only the top four of these apps were included in the Table. Only three apps were discovered to have AR capabilities, in which text is displayed within an image. Two of them, in particular, overlay the outcome on the text. The original text can still be seen behind the translations. Furthermore, because translations are displayed as images, these two translators are not real-time and hence do not fully meet the AR standards. Solely Google Translate [15] features real-time translations and is therefor the only true AR translation app. It is also the only consumer app with genuine textstickers. While TranslatAR [13] has this capability, it only translates and tracks a single word at a time, whereas

Google Translate replaces all words. Google's AR translator also allows you to manually pause translations, effectively freezing the screen. Google Translate launched in 2010, but it was not until 2015 that it featured AR translation [40]. Word Lens was the most popular AR translation app at the time and was a primary reference for most AR translation-related research [40]. Google Translate integrated Word Lens into their application in 2015, becoming the market's leading AR translation software with over 500 million downloads to date. According to download statistics, Google Translate has a significant market dominance, with ten times more downloads than its closest competitor, Microsoft Translate. The Google app is also the first translation app to appear in the Apple App Store when searching for "translator."

Google's AR translator is certainly the most advanced and cutting-edge, and it will be used as a model for this work.

2.2 Attention Classification

Attentional mechanisms are applied to filter task-relevant from task-irrelevant sensory input and mental processes at all times to deal with the constant information overload. Thus, Attention can be described as the mental process of concentrating on certain perceivable information. The processes, layers and dimensions of attention are numerous. One distinguishable aspect of attentional mechanisms are internally and externally directed attention [45]. In times of internal attention, sensory input is rated as task-irrelevant and the focus is on internally produced information (i.e. thoughts, memories, mental problem solving). External attention, on the other hand, is a selection and emphasis on information offered by our environment [8]. For the suggested use case, reading the translations required external attention and thinking about the question/task requires internal attention. In this study, we will interpret neurophysiological activity recorded by an EEG to produce a quantitative assessment of the given attentional state. Other studies instead use eye tracking data as classification input [1, 44, 44, 48]

Cooper et al. [10] found that alpha band amplitudes are higher during periods of internal attention, which they attribute to active blocking of external input. According to Benedek et al. [4], right-parietal alpha power increases with internal focus. Putze et al. [35] were the first to distinguish between internal and external attention using an EEG and Linear Discriminant Analysis (LDA) on a single trial, achieving an accuracy of up to 81.2%. Vortmann et al. [48] employed a multimodal setup to categorize internal and external attention in a real-time evaluation, employing both EEG and eye tracking, with an accuracy of 60.9%. In another study, Vortmann and Putze [46] improved the usability of a Smart-Home System with the multi-modal setup of the previous study, reaching a real-time accuracy of 65.7%.

As previously stated, we will use a light-weight, low-cost EEG headband for the BCI in this study in order to improve the usability and application possibilities of the proposed system. We will work with the Muse 2 Headband by InteraXon Inc. [29]. The number of electrodes in the Muse Headset is less than that in the aforementioned research. However, the right parietal area is partially covered by the TP10 electrode, which appears to be a significant area for separating attention into internal and external, as previously indicated. This gives reason to believe that attentional classification into internal and external attention is achievable, despite the fact that electrode coverage is restricted and data quality is worse. Previously, consumer-grade EEG was used to determine attention in passive BCI scenarios; however, classifications referred to attention levels in terms of focus and involvement [7, 22, 26, 42]. To the best of our knowledge, using a consumer-EEG, such as the Muse Headset, to classify internal and external attention was not done before. We will use the predicted attentional state of the EEG data classification to adapt the behavior of the AR translation application.

2.3 Mobile and passive BCIs

Brain-computer interfaces allow for direct communication between the brain and an external device. Mobile BCIs eliminate the need to stay stationary that traditional BCIs have due to their wiring. As BCI technology becomes more popular, several businesses are working to develop mobile, pleasant, and non-invasive EEG solutions. Because of its setup and method of data collection, only EEG meets the requirements for locomotion among non-invasive brain activity recording modalities [25]. Because it is a reasonably cheap and effective recording technology [31], EEG is one of the most popular types of BCI, having been employed in 60% of BCI research from 2007 to 2011 [18]. However, laboratory EEG necessitates a time-consuming setup and cleaning procedure [14]. A trained specialist is required to place the electrodes correctly [38]. A conductive gel or saline solution is routinely used to promote scalp connection, which improves data quality [14]. An mobile BCI, on the other hand, increases usability and aesthetics by eliminating the need for lengthy electrode placement procedures. It can also be used dry, avoiding the need for the time-consuming application and subsequent removal of conductive gel. This, however, comes at the expense of data quality. As previously stated, research EEG needs the placement of a qualified professional, whereas mobile BCI EEG can be placed by a novice. Consumer-oriented BCI solutions, such as headsets or headbands, are offered by companies such as "Emotiv (EPOC), Neurosky, Advanced Brain Monitoring (B-Alert X10), InteraXon (Muse), and Melon" [14]. The Muse headset was validated for EEG research by Krigolson et al. [24].

For a passive BCI, users do not need and should not alter their way of thinking actively [50]. It is meant as an implicit interaction in which the passive BCI picks up automatic, spontaneous brain activity for a background monitoring of cognitive and affective states. The implicity is defining for passive BCIs; the user should utilize the system as if there was no passive BCI [31].

passive BCIs have been utilized frequently in research over the last few decades, although largely in laboratories or tightly controlled situations [2]. Recently, research on applications in real-world settings is emerging. In Roy and Frey [36], passive BCIs help users under substantial stress and cognitive load, such as air traffic control or drone management. They accomplish this by modifying the information presented on an interface in order to reduce task complexity and hence workload. A passive BCI can also be used to update the user interface depending on user-experienced problems [50]. Zander et al. [49] assessed the use of a passive and mobile EEG (actiCAP Xpress dry-electrode) for autonomous driving in terms of signal quality and usability and discovered that the prerequisites for the development of actual systems were met. An autonomous automobile might utilize the data to determine if the driver is ready to take over control, or to assess the user's mood or attentional state.

In this work, we will make use of a passive mobile BCI together with a smartphone application aiming at an easy fast setup and effortless usability to decrease the distraction caused by AR applications.

3 THE MOBILE BCI-SMARTPHONE SYSTEM AND APPLICATION

The AR translator implemented for this work is a replication of the Google Translate app, because it proved to be the most advanced and state-of-the-art. We implement our own version of the app to make be able to make the necessary changes for the BCI integration and because the source code is not publicly available. Translated text is presented dynamically in real-time using textstickers that create a believable AR experience. However, the possibly distracting and unnecessary updating of the textstickers while users are not reading is an excellent use-case to test a passive BCI

for attention-sensitivity. When users read and process a translated text, they alternate between internal and external attention. We will determine the attentional state and modify the interface's behavior accordingly. The update frequency, appearance of textstickers, location of textstickers, and changes in appearance and location between updates determine the level of distraction. Our application can suspend the process of updating textstickers when the attention is internally focused and as soon as the user turns their attention externally, the updating of textstickers can resume. We will use a Muse 2 Headset to avoid limiting the mobility of users and for a fast and easy setup.

3.1 Translation App Implementation

We had to prioritize rapid computation and execution times when developing the real-time AR translator. Balancing performance and efficiency was key when selecting and developing algorithms. The main app was written using Java and C++. The app operates using an image resolution of 540x960 pixels, which was found to be the optimal compromise between performance and quality.[1] The essential features of the AR translator are provided in the following sections to provide a brief structural overview.

The activity diagram in Figure 1 visualizes the structure of the AR translator. After the user launches the app, the main processes are started.

3.1.1 Textsticker creation. First, the textsticker production process begins with a timer, which serves two purposes: as a delay for the camera to establish focus and as a wait between updates for the duration of the app's existence. The time it takes to create new textstickers ranges between 300 and 3000ms. It is expected that enough time has passed for the textstickers to be changed after this amount of time. The open source PaddleOCR, which is neural network based and under active development with frequent releases, was used in this work for optical character recognition (OCR) [33]. The ultra lightweight OCR is fast and has a high accuracy [12], even on scene text, making it ideal for this application. To generate the texts, an image is sent into PaddleOCR's model, which does text detection, recognition, and classification.

Following that, the discovered labels are forwarded to Google's translation service through REST API. The translated labels and the original image are combined to create textstickers with the same text and background color, font size and thickness as the translated text. For each label detected by OCR, a region containing the label's location in the image is returned. This is accurate, however because OCR and textsticker production take time, the text may have changed in the meantime, invalidating the previously accurate detected areas. As a result, the identified text regions are updated after the generation of the textstickers to reflect reality. After textsticker creation is complete, the image utilized for the OCR model is compared to the initial picture. Using feature matching, the label regions from the OCR image are redetermined in the latter picture [23]. This method significantly reduces the latency to reality, allowing optical flow to be exploited for real-time tracking. Optical flow attempts, at best, to determine where points from one image went to in another similar image [23]. Following that, any existing textstickers are replaced with the textstickers generated by this update.

3.1.2 Textsticker Update. The updating of textstickers is accomplished by the use of optical flow. It begins by locating four tracking points within a label region for each textsticker. These tracking points can be identified again in a new image, and the placements and shapes of the textstickers are modified based on the difference. To help overcome the

[1] The images of the app presented in the following were taken using a higher resolution (720x1280).

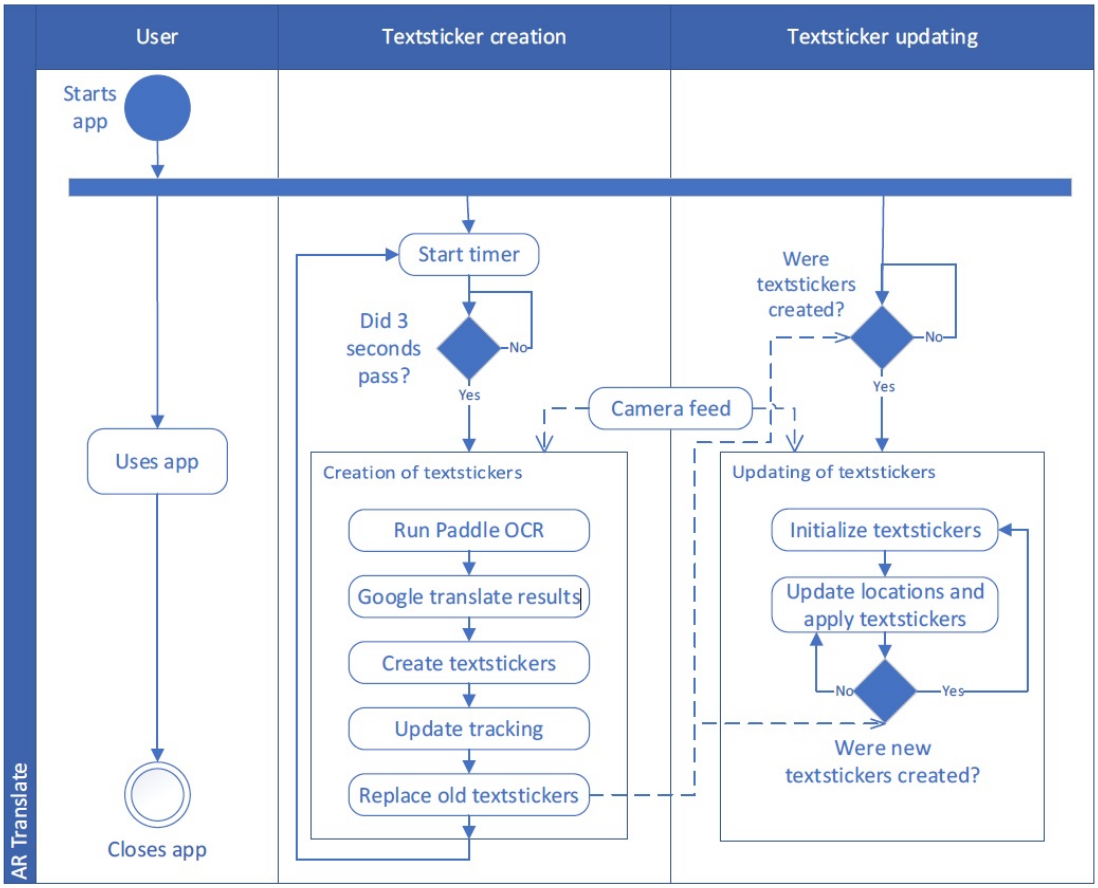


Fig. 1. Activity diagram for the AR translator

latency caused by feature matching, a wider search window size is used for the initial application of optical flow after feature matching. Following that, a reduced window size is used to boost speed, allowing for a higher frame rate. This is repeated frame after frame until the textstickers are changed with updated ones.

Figure 2 shows examples of original text, our AR translation application and the Google translate results.

3.2 Brain-Computer Interface

In the following, the addition of the BCI based adaptive pausing is described. This includes how the data from the Muse Headset is transferred to the app, how that data is processed and classified into either external or internal attention.

Figure 3 gives an overview of the data flow structure from the Muse Headset to the app, which will be described in the following. The data is first sent from the Muse Headset to an app called Mind Monitor [27]. Once linked to the Muse Headset, the app displays several graphical representations of the EEG frequency spectra as well as an indicator of the headset's fit, which is useful for determining whether the data from Muse is being captured adequately

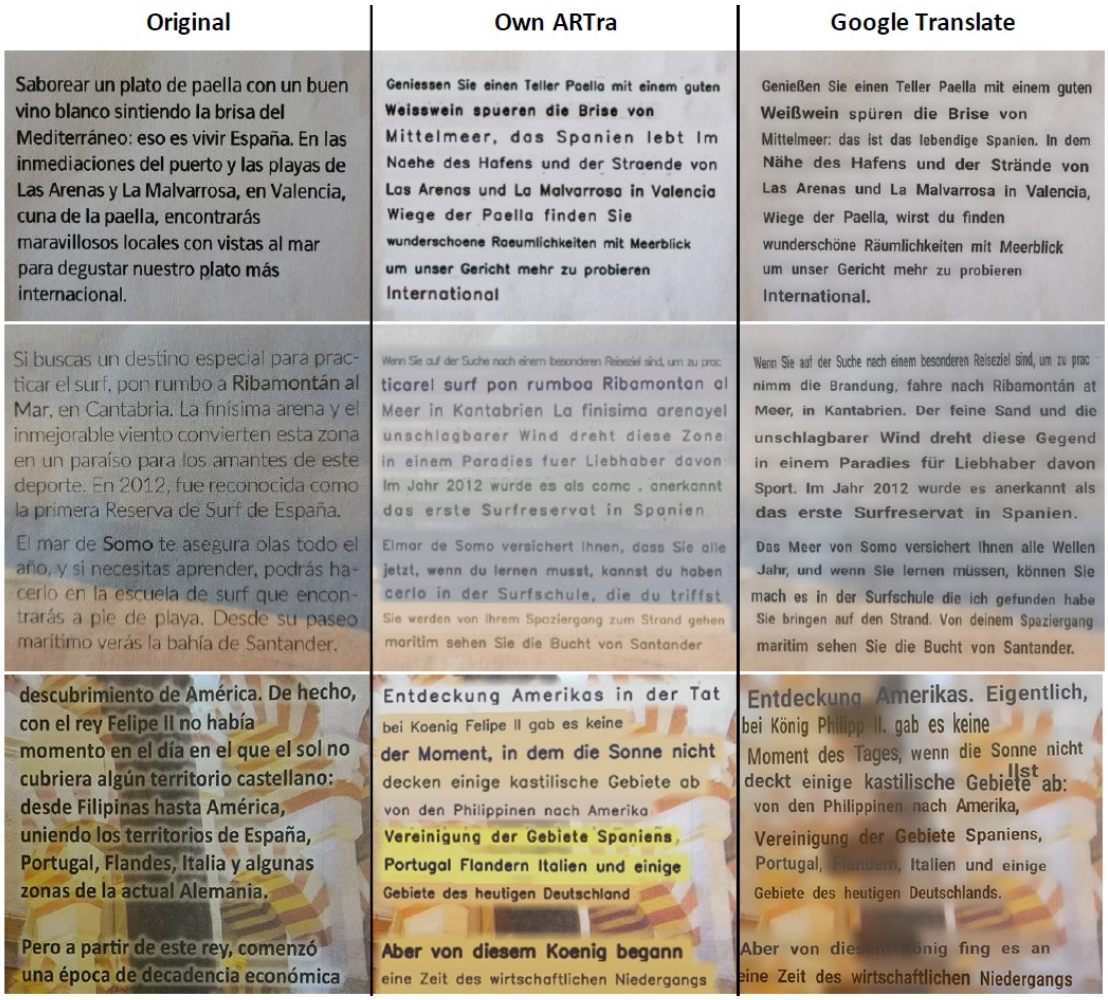


Fig. 2. Comparison of the implemented AR translator (ARTra), Google Translate and the original text.

and the device is being worn correctly. Furthermore, Mind Monitor applies a 50Hz notch filter to remove power line noise.

Mind Monitor streams the data via Open Sound Control (OSC) using User Datagram Protocol (UDP), acting as an interface to the Muse Headset's data. The OSC data is streamed to a device via WiFi by specifying the receiver's IP address. Mind Monitor is running from a separate device

- to reduce interference from the OS Android (the Mind Monitor should run in foreground)
- because the AR translate app is already computationally demanding, it makes sense to not put any extra strain on the device.
- because it enables continuous monitoring during the execution of the study, to ensure that there are no issues with the connection between Muse and Mind Monitor.

Fig. 3. Muse headset data flow

- the creator of Mind Monitor mentioned that the Muse Headset has connection issues with the Bluetooth module of Huawei phones, which was used for the translation app devices.

To receive the OSC data from Mind Monitor, the library JavaOSC [20] is used. When OSC streaming is enabled in the app, the Mind Monitor app provides multiple data streams over OSC pathways. The subscribed items are

- **EEG** absolute values of delta, theta, alpha, beta and gamma frequency bands as four float values[2]
- **Horseshoe** indicating fit of the electrodes
- **Battery info**
- **Gyroscope** measuring or maintaining orientation and angular velocity
- **Accelerometer** measuring acceleration
- **Blink**
- **Jaw Clench**

The absolute power band frequencies will be used for the attention classification process.

The Muse Headset can be seen in Figure 3. It is a four-channel EEG with two silver sensors on the forehead and two conductive silicone-rubber sensors on the ears. It can communicate data wirelessly via Bluetooth 4.0 and samples data at 256Hz. The electrodes cover areas of the brain in the anterior frontal (AF7, AF8) and temporoparietal (TP9, TP10)

[2]Frequency ranges: δ(1-4Hz), θ (4-8Hz), α (7.5-13Hz), β (13-30Hz), γ (30-44Hz)

lobes. It also has a 3-axis accelerometer and gyrometer for tracking head motions. [19].

The Muse Headset is adjustable in length and so suits a wide range of head sizes. It is a low-cost EEG designed to be used as a personal meditation helper. With a price of 250$, it is far less expensive than research EEG devices (e.g., ActiChamp 75,000$). In terms of data validity, Krigolson et al. [24] assessed the Muse data of 1000 participants in varied contexts and demonstrated its robustness and accuracy in a visual oddball task. The Muse Headset's detection patterns were found to be identical to those of the research-grade wet EEG systems actiCHamp and g.Tec [9]. In noisy environments, Przegalinska et al. [34] critique low data quality. Additional Muse Headset problems were discussed in those studies: Bluetooth-related delay and jitter (20-40ms delay, jitter 5ms), a temporal unstable beginning of consecutive samples (time difference of -10ms to 150ms), and missing samples (0.01-0.05 percent missing samples across all participants).

3.2.1 Classification. The adaptation of the AR translation application behavior was based on the classified attentional state. We differentiated internally and externally directed attention based on the recorded EEG data. Following the results of Putze et al. [35] and Vortmann and Putze [47], we chose 4-second data windows for the feature extraction of single trials. The classification was performed person-dependently and 400 seconds of training data were collected per person to train a Linear Discriminant Analysis (LDA). The training phase of the classifier will be explained in more detail in the user study. The extracted features were provided by the MuseIO [30]. To calculate the absolute band powers of the aforementioned frequency bands, a Fast Fourier Transform is used on a 256 sample Hamming window, sliding at 10 Hz. For each electrode and frequency band, one sample is extracted every 100 ms for the previous second. Thus, a total of 20 features is contained in the feature set and the windows overlap with 90%. As 400 seconds of training data were recorded per participant, a total of 4000 feature vectors were available per participant. The LDA was implemented using the standard scikit-learn parameters and settings. To calculate the training accuracy for each participant that was used to evaluate the quality of the calibration and training phase, a 5-fold cross-validation was performed on the data and the average was calculated over all folds. The results will be reported later. For the final training of the LDA, all feature vectors were used.

Apart from the notch filter, no additional preprocessing of the data was performed to keep the computation times at a minimum.

3.2.2 Prediction Integration. In the following, the strategy to adapt the BCI behavior based on the classification results of the BCI data will be explained.
In times of internally directed attention, salient changes of external stimuli, such as the smartphone display, can be distracting. Thus, whenever the attentional state of the user is presumably "internal", the updating of textstickers is paused. However, the pausing of translation updates should not occur in times of externally directed attention. The classifier predicts the attentional state every 100 ms based on the last second. We compile the predicted labels of 4 seconds to make a decision whether to pause the translation updates or not. Within the 4 seconds, at least 60% of the predictions (n=40) have to be labeled "internal" for the translations to pause. This threshold was chosen because wrongful pausing during external attention must be avoided, whereas missing pausing during internal attention can be tolerated.

4 USER STUDY

In a user study, the AR translator was tested with and without the addition of attention-based pauses. This gave information about the performance and usefulness of the AR translator, as well as the modality preferences of the participants. Furthermore, the modality with BCI was examined in terms of attentional classification accuracy and distribution, as well as the length and correctness of pauses.

The study's hypotheses are stated first, followed by the methodology, which explains how the study was carried out. Following that, the results are displayed and discussed in the last section, highlighting potential causes of faults and recommendations to enhance the system.

4.1 Hypotheses

One of the goals of this research was to evaluate the usability of a mobile BCI for attention-awareness. Several hypotheses concerning this purpose were considered.

4.1.1 Main hypotheses. The main hypotheses relate to the positive (*H1*) and negative (*H2*) effects that the pausing of translation updates may cause. It is believed that while users have internal focused attention, the pausing of translations is perceived as positive (*H1*). Contrary to that, pauses during phases of external attention, which occur during reading, are believed to have a negative effect (*H2*).

H1. The larger the percentage of pauses during thinking, the higher the experienced usability and the lower the task load.

H2. The larger the percentage of pauses during reading, the lower the experienced usability and the higher the task load.

As the app makes decisions for the user that are not always correct, attempting to help the user by pausing when they are thinking inevitably comes with negative effects while reading. *H1* and *H2* with regard to the attentional prediction accuracy gives insights to how high the accuracy needs to be to improve the app, as the accuracy is a deciding factor for how the usability regarding the pausing is perceived by participants.

4.1.2 Other hypotheses. The other hypotheses relate to whether text appearance or content may influence the experienced usefulness of the pausing.

H3. The more distracting the displayed translation results are, the more helpful pauses are during thinking.

H4. The more demanding the combination of text and question is, the more helpful pauses are during thinking.

Answering these hypotheses shows for which types of text the pausing is most useful. Another key goal of the study is to test the AR translator, for which it makes sense to test various texts that differ in visual and contentual complexity, ergo *H3* and *H4* can be tested with little extra effort. Another purpose of the study, as previously stated, is to evaluate the AR translator. This comprises not only the operating principle of the AR translator, but also the two modalities of pausing and not pausing. This assists in determining where the AR translator can be enhanced the most and where it already performs adequately. Evaluating the two modalities in relation reveals whether pausing is advantageous or not, and why.

4.2 Methods

The research was conducted following a within-subject design. Each participant tested the app with and without the Muse Headset. The evaluation of the hypotheses were mainly based on questionnaire results. To test the translation app, eight posters with Spanish texts were created. For each participant four posters were used for the version with Muse Headset and four for the version without. The four permutations of these combinations were labeled modalities A, B, C, and D, which were rotated throughout the study's runs. The participant count was multiple of four. Thus, the poster distribution across both versions was balanced.

4.2.1 Participants. Fourteen healthy participants between 16 and 59 (mean age 27.1, SD 9.8) with normal or corrected eyesight who were all german native speakers participated in the study. Two participants were excluded because the attention prediction was excessively unbalanced, making the BCI version unsuitable (nearly always paused) or too similar to the BCI version (almost no pauses). This reflects the expected rate of people with BCI illiteracy [39]. To keep the equal distribution of poster combinations, two new volunteers were selected for the outliers' experiment versions. Eight of the remaining participants were men, while four were women. The majority of participants did not speak Spanish at all, and those who did had just a rudimentary understanding. Two of the participants had previously used an AR translation software (both Google Translate). Each participant was informed about the collection of EEG data and written participation consent was collected. The data was anonymized by assigning six-digit participant IDs at random and the study was approved by the local ethics committee.

4.2.2 Procedure. Figure 4 shows the general structure of the study, which will be explained in the following.

Participants were introduced to the study verbally and through an introductory document, and then provided written consent after being informed about the goal of the study and the data that would be collected. Following that, a demographic questionnaire and the Mind Wandering Questionnaire (MWQ) [28] were completed. The MWQ was collected to find suspiciously high mind wandering scores because these participants would possibly influence the study results. Afterwards, the app was tested in the participant specific order. Half of the participants began with BCI, while the other half did not. The poster sets were the same; half began with poster set A and half with poster set B, each of which featured four posters. The production and selection of posters will be discussed more below.

For the run with BCI, a calibration of the EEG headset and the training of the classifier were required. The initial calibration was performed using the built-in impedance measurement of the Muse headset. The training data collection included four German texts that were read via AR (externally directed attention) and four text specific questions after each text (internally directed attention). Each of the eight parts lasted roughly 50 seconds (400 seconds of training data in total per participant), after which the app informed the participant that enough data had been collected for that part. Participants were instructed not to talk during the training data collection parts and to refrain from touching the Muse Headset during the remainder of the study, as this could impair the prediction accuracy.
After the calibration, for the training data collection, participants were asked to imagine that they are tourists in a foreign country, as that is a fitting use-case scenario for an AR translation app. Virtual posters were displayed on the smartphone screen. To match the overall study narrative during this phase, the content of the texts was related to traveling or about facts of countries. To ensure that participants had something to ponder about within the 50 seconds, the questions were relevant to the texts and were either open-ended or asked for a lot of information about the text. For example, after reading a paragraph describing Chinese customs, participants were asked to think about both Chinese

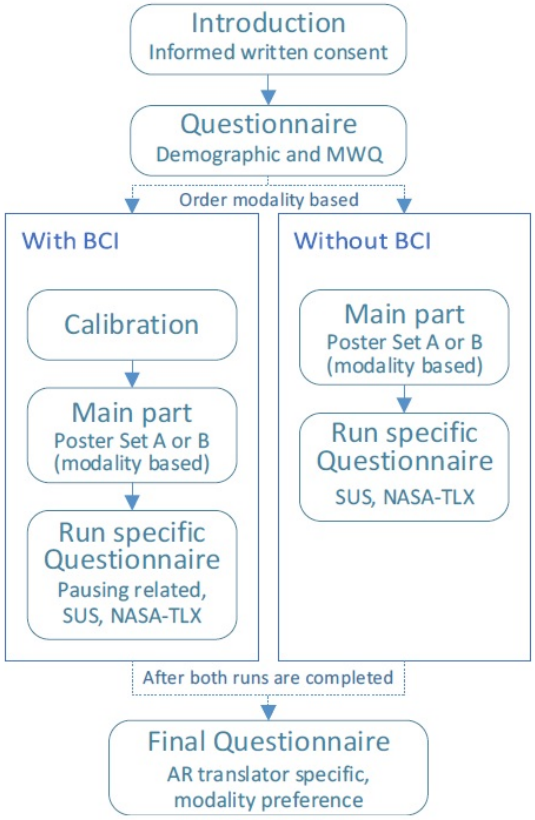


Fig. 4. User study overview

and other foreign customs.

During the study, the participants' main task was to read the translations of posters using the AR translation app and then answer questions about the content. The arrangement was identical to the training phase, except that actual posters were pinned to the pinboard and had to be scanned with the camera.

Following the reading (externally directed attention), the task phase began, in which participants were given a question to think about (internally directed attention). For the task, participants were instructed to consider the question and mentally respond it to themselves. Participants were also given the option of rereading the text during the task, however they were instructed to first think about the question. Participants were also instructed to rethink their answers completely if they reread the material. Participants were prompted verbally for their replies after each question to reward them and to subjectively judge the difficulty of posters and the quality of answers. The participants were not timed and could spend as much time as they needed to read the poster and complete the exercise.

The main part structure was communicated verbally and visually to the participants using a guidance document that included screenshots of each step of the main part. This added an explanation of when and why the application

paused in the version with BCI. Participants were informed that pausing may occur during reading and that waiting or attempting to read may cause the pause to be unpaused. Participants completed a run-specific questionnaire after each of the two main component runs (with and without BCI). Following the completion of both runs, a final questionnaire was completed.
Participants spend an average of 98 minutes on the study. To maintain excellent signal quality, the Muse Headset was routinely cleaned before each research.

4.2.3 Questionnaires. Before testing the app, participants filled a demographic questionnaire and the Mind Wandering Questionnaire[28] translated to German. The latter is based on a Likert scale and includes questions to generate a score that shows how likely the person is to start mind wandering. Participants completed a run-specific questionnaire after each run in the main section. The well-known system usability scale (SUS) [6] and the short form of the NASA Task Load Index [17], known as Raw TLX [16], were filled out in this questionnaire (German translations for SUS [5] and Raw TLX [32]). In addition, for the version with BCI, participants answered Likert-scale based questions on translation pauses. These questions were designed to provide answers to the four study hypotheses. Finally, following both runs of the main part, participants completed a questionnaire to rate their overall experience of the AR translator. These questions used a Likert scale and dealt with display smoothness, tracking, visual authenticity, accurate positioning, and contentual correctness. In the last section of the study, participants choose which version, with or without BCI, they preferred and may provide an open-ended response as to why.

4.2.4 Stimuli. An example poster is shown in Figure 5. The choice of posters for the study plays an important role in testing the mobile BCI for the AR translator, as well as being able to answer *H3* and *H4*. The behavior of the AR translator greatly depends on the appearance and structure of texts. Thus, to sufficiently test the BCI version, a variety of different texts needed to be used to obtain a more holistic review. The creation of the eight posters for the study was done by defining criteria based on what was needed to answer *H3* and *H4*. As *H3* relates to the appearance of texts, answering it also requires a variety of different text appearances similar to what is needed to test the AR translator more holistically.

Criteria for the Distractiveness of Translation Updates of Posters. H3 pertains to the distractiveness of textsticker updates, which depends on the appearance of the text and background on posters. To test if the BCI version of the AR translator may be considered more helpful for text/background combinations that elicit a lot of translation updates, several posters differing in design aspects have to be used. von Mühlenen et al. [43] show that a change in text color strongly captures attention. This effect is likely to be amplified if the background color changes as well. A smaller disparity in textsticker changes is less obvious and consequently less distracting. The visual complexity of a text and its background has a direct relationship to the consistency of textsticker generation. The AR translator replicates a text with an equally colored background more reliably than a text with alternating background colors, because noise and more color shades lead the determined background colors to stray a lot more. This results in a more prominent and attention-grabbing appearance over time, as the switching of textsticker colors between updates is unavoidable.

Another issue is the consistency with which the textstickers are correctly positioned. If the location of textstickers is not consistent between updates, they will shift a lot. Because movement attracts attention [11], consistency in location lowers the distractiveness of updates. A background that is equally colored has fewer potential features for the feature matching procedure. Even if the background lacks good features, the text itself can be used as a reference point. Furthermore, text with poor contrast to the background is unreadable by OCR and even by humans. Hence, it is not

Translation

RIDE THE WAVES IN
SOMO-RIBAMONTÁN

If you're looking for a special surfing destination, head for **Ribamontán al Mar** in Cantabria. The fine sand and superb wind make this area a paradise for lovers of this sport. In 2012, it was named Spain's top surfing resort.

The sea at **Somo** guarantees year-round waves, and if you're a beginner you can take lessons at the surf school you'll find on the seafront. You can see the Bay of Santander from the seafront promenade.

Whatever your level, you'll find waves to suit everyone in other villages and towns in the area like **Loredo**, **Längre** and **Galizano**.

Fig. 5. Example poster that was used in the study and the translation (in a separate box for showcase purposes)

possible to create posters with deviating consistency of textsticker placements without significantly compromising legibility and poster authenticity.

The frequency with which textstickers are updated has a direct effect on distractiveness. Changes to the text become more obvious as they occur more frequently, increasing the distractiveness of the previously described distractors. When the text is spread out, less text is in the camera view at the same time, lowering the complexity of the calculations necessary and hence the computation time. As a result, for less dense texts, the update frequency may be higher, increasing distractiveness. Larger typefaces, as well as text layout that spreads out text, are two ways to create lower text density.

Other elements influence distractiveness (change in font size between updates, word ambiguity causing various words to appear between updates), but the ones presented were thought to be the most effective and sufficient to design posters with enough variation.

The criteria are not binary in nature. As a result, in order to make four posters that rise in distractiveness, they must be created in relation to one another. The first poster should be the least distracting, while the last should be the most distracting.

Based on the criteria, four pairs with similar characteristics were created. It was attempted to similarly increase the level of the distraction per difficulty (e.g., gradually increasing font size and spacing between lines to reduce text density from poster 1 to 4).

Criteria for the Difficulty of Text and Question Combinations. To evaluate *H4*, the combination of text and question needs to be of varying mental demand. However, this did not need to be as nuanced as the appearance of the posters, as it is only relevant for *H4*. The length of the text is one consideration. The mental demand is often increased when more text provides more information, especially when open-ended questions are asked, where large sections of the text contain part of the answer. Those that require lengthier answers are likely to be more difficult than questions that can be answered in a few words since the participant must recall bigger portions of the text. Two difficulties were created based on the criteria above:
Simple text and question combinations had a word count around 100 - 150 words. The questions had simple, clearly defined answers (information that is contained in just one or two sentences of the text). The difficult text and question combinations had a word count of around 200 - 250 words. The questions were open-ended and ask for information contained in large parts of the text. Potentially, the participant needs to reread parts of the text to solve the question, as it is difficult to remember everything needed to solve the question. It should be noted that when using an AR translator, the process of reading is not as straight forward as compared to regular reading. Thus, the general demand is higher than it may seem.

In total, eight posters were used for this study. The posters were split into two groups of four of which one was used for the configuration with BCI and one for the configuration without BCI (balanced across participants). Within each group the posters had 4 different levels of difficulty regarding the distractiveness and two levels of difficulty regarding the text and question combinations.

4.3 Results

To evaluate both the application and the user study, and to test our hypotheses, we analyzed the classification accuracies and the questionnaire results. For all significance tests, an alpha level of 0.05 was assumed. All correlations were tested using the Pearson correlation and differences were assessed using dependent t-tests. To increase the readability, all six-digit participant IDs were exchanged by numbers 1-12.

4.3.1 AR Translator Application Rating. One post-session questionnaire was specifically designed to rate different aspects of the AR Translation application, independent of the BCI aspect. The questions and answers can be seen in Figure 6. For questions 1, 2, 3, and 5, the participants had opposing opinions, however for questions 4 and 6, they agree more. The latter concern the correct arrangement of translations (4) and the matching colors of translations (6), which are likewise the highest rated aspects (mean 5.6 and 5.4 respectively). The smoothness of translations (1) and the correctness of translations (5) are the lowest rated features (mean 4.0 and 3.9, respectively). Surprisingly, participants rated "translations matching colors" (6) higher than "translations replaced text visually authentically" (3). (means: 5.4 versus 4.8). When asked directly about the app, many participants stated that the translations were not perfect, making the texts difficult to read in some sections. Nonetheless, the participants were able to understand the core of the texts

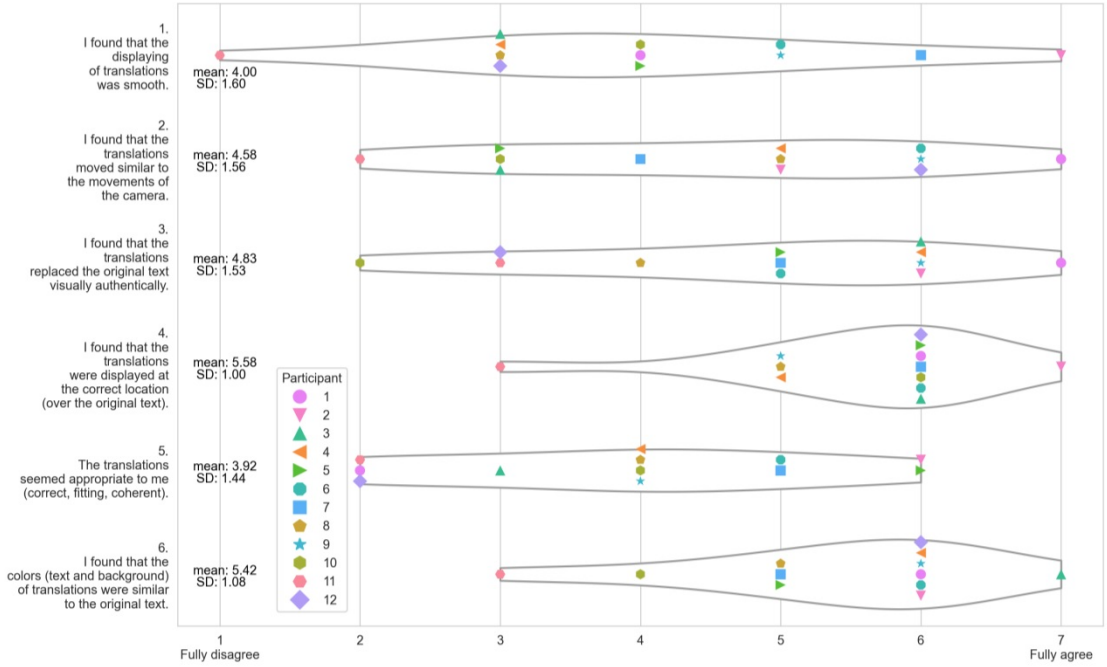


Fig. 6. Questionnaire results per participant for the AR translator application rating.

and, as a result, could respond to questions that frequently covered the most relevant components. Following the experimenter's subjective judgment, 72% of the answers were rated good (the majority of the questions were answered), 24% okay (about half of the questions were answered), and just 4% poor. In the NASA-TLX, participants responded to the question "How successful were you in performing the task?" "How satisfied were you with your performance?" The average result for both modalities in this performance category was 5.5 (SD 2.34), placing them directly in the middle of the scale. Although the majority of the answers were regarded good, the participants did not appear to agree as much on average. Some participants expressed a desire to be able to interrupt the translations at any time, a function available in Google Translate. The preference of modality had no significant nor interesting correlations with the MWQ.

In Conclusion. The most criticized aspects of the AR translator were the smoothness and the correctness of translations. The AR translator seems to take too long to display translations. Also, translations were confusing at times, as the they are created line by line, lacking coherence between lines. The placement and colors of textstickers were rated the best among the categories, even though some posters were designed to create difficulties regarding color determination for the AR translator. Even though some participants had problems, they were generally able to complete the tasks. Some participants stated that they would not be able to understand anything without the app, but with the AR translator, the Spanish sentences became understandable. We conclude that the quality of the implemented AR translator was good enough to test our hypotheses regarding the mobile BCI integration for attention-awareness.

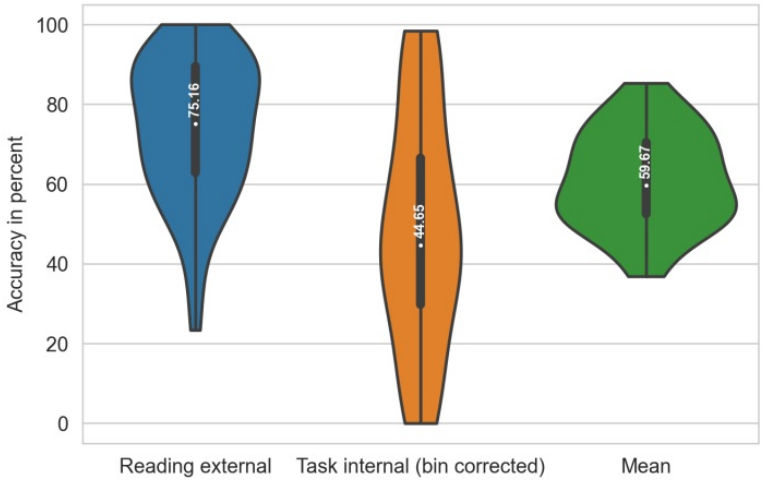


Fig. 7. Prediction accuracies for reading, bin corrected task and mean of both.

4.3.2 Classification Accuracies and Pauses. For the BCI version of the AR translator, we evaluated the classification accuracies and analyzed the pauses. These results shine a light on how well the implemented system worked in terms of "attention-awareness" and will help explain and discuss the results of the questionnaires.

The accuracy of the attention classification is largely influenced by the calibration and training phase accuracy. On average, the accuracy during the classifier setup was 82.5% ± 8.2% (calculated using a 5-fold cross-validation as described).

As a ground truth, it was assumed that participants' attention was directed externally in times of reading and internally during question answering. However, because the participants had the option to review the text while answering the question (which would result in externally directed attention within a phase of internally directed attention), we analyzed the reading phases in more detail and found that thinking and rereading resulted in longer task solving times for some trials. Additionally, throughout the course of the question answering phase, the classification accuracy dropped significantly in the middle for some trials. This indicates that the assumed ground truth (internally directed attention) is wrong and need to be corrected for our offline analysis of the classification accuracy. For trials identified as "thinking and rereading", phases (each trial was separated into 5 bins) with a high probability for external attention were excluded for this analysis (bin corrected). Figure 7 shows the distribution and mean of classification accuracies per trial during the reading and task separately, as well as the mean accuracy per viewed poster (n=48 for BCI condition).

The mean classification accuracy in the main part of the study was 59.67% which is comparable to the real-time attention classifier in Vortmann et al. [48]. The classification accuracy in time of external attention (75.16%) was higher compared to the internally directed attention (44.65%). Label noise for this analysis has to be assumed for both trial parts.

Figure 8 shows the correlation between the classifier calibration accuracy after the training phase and the classification calculated in the offline analysis on the assumed ground truth labels. As $p = 0.16$, the weak positive correlation of

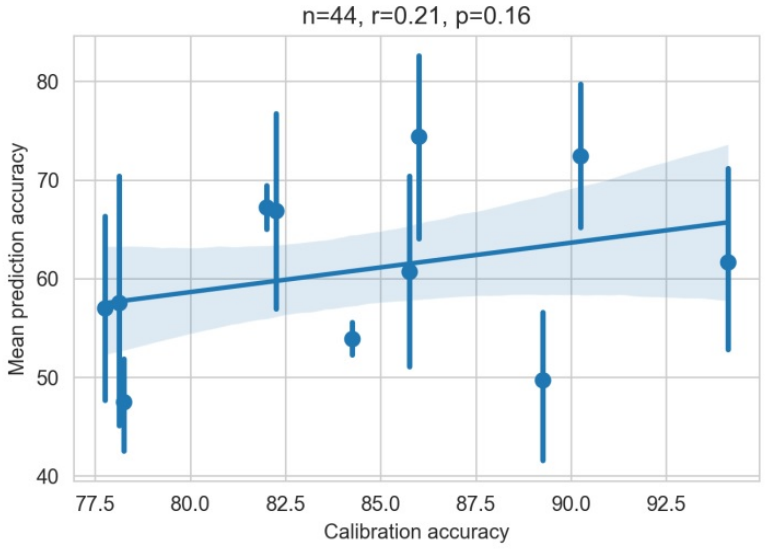


Fig. 8. Pearson correlation of calibration and mean prediction accuracy. Each dot with line represents the four posters for one participant. The dot is the mean and the lines are the min and max values of the four posters.

$r = 0.21$ is not significant. For most participants, there are also large fluctuations in mean prediction accuracy between the posters.

With BCI, participants needed 29 seconds longer (+23.7%) to read the posters, a statistically significant difference ($p < 0.01$). Surprisingly, this is also true when comparing the modalities of people who favored the BCI (M=36 seconds, $p < 0.01$). Participants needed slightly less time with BCI for the task, although this is not statistically significant ($p = 0.32$). In a next step, we analyzed the distribution and length of paused and unpaused translations. Each continuous time interval of either paused or unpaused translations were rated as one block and categorized depending on their length. Figure 9 shows the share of each block length depending on the current trial part. The translations were paused for 17% of the total time participants spend reading the posters. Almost 80% of these pauses were shorter blocks of below 3 seconds or between 3 and 10 seconds. The largest part of reading time the translations were not paused for intervals longer than 30 seconds. In times of internal attention (only thinking), the translations were paused for 35% of the total time. This was reduced to 33% in the trials there were later identified as thinking and rereading. It can also be noted that the length of the continuous pauses is longer for task parts than for reading parts.

The effect of the translation pauses during the reading part on the total task solving time was analyzed using a correlation analysis (see Figure 10). The moderate positive correlation of $r = 0.51$ is highly significant with $p < 0.001$. Thus, longer translation update pauses during reading significantly increased the total task solving times.

In Conclusion. For the modality with BCI, the attentional classifier reached a median accuracy of around 60%. However, this value is likely affected by label noise, as the study design made a clear separation into thinking and reading difficult. Regarding pauses, the percentage of time paused during reading was around half of that of the task (reading pauses

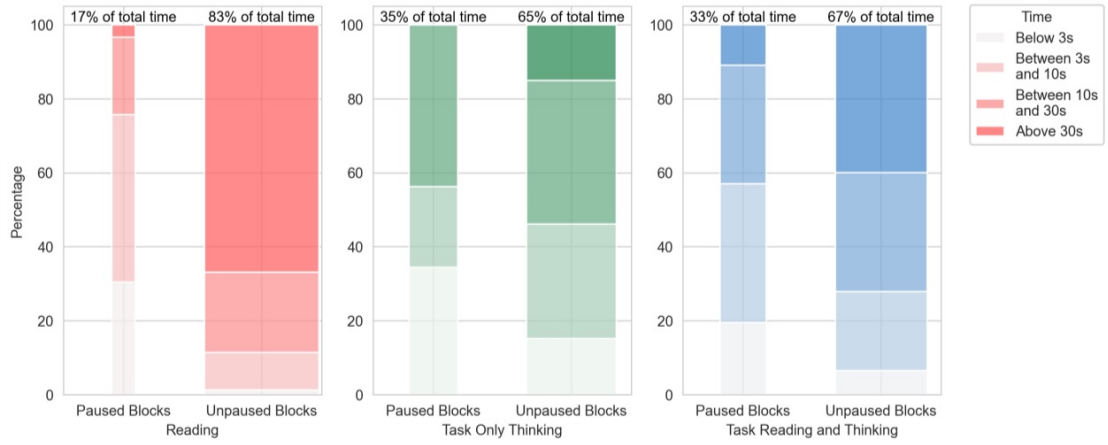


Fig. 9. Block during which translations were paused or unpaused separated by reading and task. The task is split into "Only thinking" and "Thinking and reading". The data uses the mean of the distributions of participants, so that the length of a run does not increase its weight; each participant had the same weight.

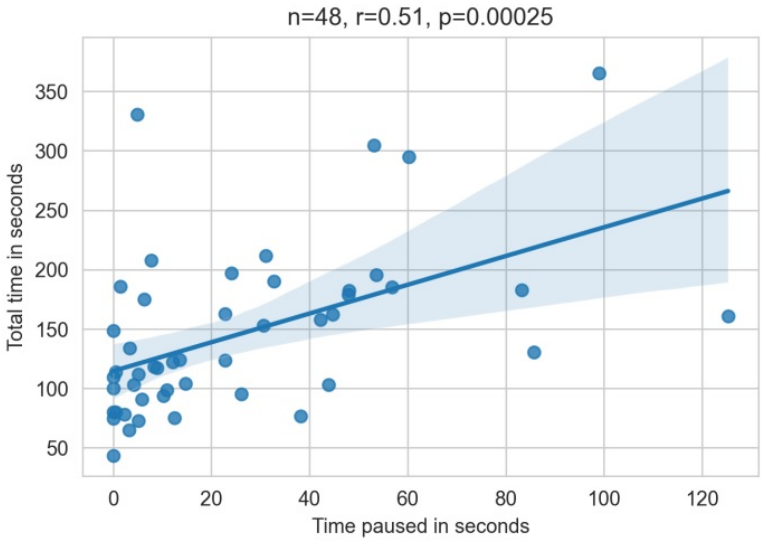


Fig. 10. Pearson correlation of the pause time during reading and the total time.

17%, task only thinking 35% and task thinking and reading 33%). Although a clear difference in pause percentages exists between reading and task, the paused percentage during the task is generally rather low as not even half of the task time was paused. Participants reported to be more affected by the reading pauses.

4.3.3 SUS and NASA-TLX. The SUS and NASA-TLX questionnaires were provided after each version of the AR translators separately and will be compared to rate the usability of the attention-aware translator compared to a

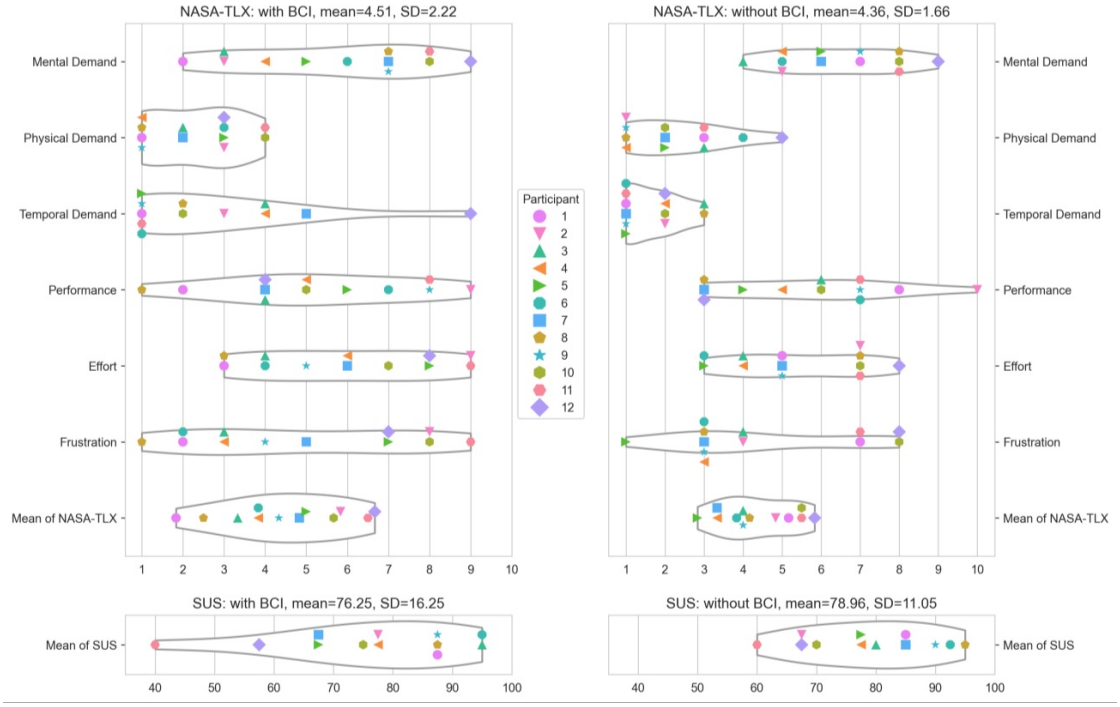


Fig. 11. NASA-TLX and SUS ratings per participant for BCI and BCI-less versions of the AR translator app.

regular AR translator. The version without BCI achieved higher results on the NASA-TLX and the SUS, with a reduced standard deviation, for both questionnaires. The higher SD for the version with BCI could be attributable to the fact that some people found the pause useful, while others found it annoying, resulting in polarization. The BCI version of the translator achieved an average SUS score of 76 , while the BCI-less version achieved a score of 79. According to Bangor et al. [3], these values fall between between "good" (71.4) and "great" (85.5). The version without BCI is slightly closer to "excellent," while the version with BCI is slightly closer to "good." The answers per person for both questionnaires can be seen in Figure 11.

To evaluate the differences between the NASA-TLX categories for both versions of the system, paired t-tests were performed but no significant differences were found. The biggest rated differences were present for the mental and temporal demand (see Table 2. The overall NASA-TLX and SUS score ratings were also not significantly different.

In Conclusion. The SUS and NASA-TLX scores were better for the version with BCI. The biggest difference for the NASA-TLX was the temporal demand, which might be due to the fact that participants required on average 29 seconds longer to read posters with the BCI ($p < 0.01$). The SUS score with BCI is 79 and without 76, both can be considered “good” according to Bangor et al. [3].

4.3.4 Hypotheses. The main aspect of this study was to find out whether an attention-aware AR translator would be superior to an attention-unaware system. When asked which modality they preferred, 58.3 percent chose the version without BCI and 41.7 percent liked the version with BCI. The major reasons for preferring an unaware version were

Table 2. Mean and Standard Deviation of the NASA-TLX categories

NASA-TLX category	With BCI		Without BCI		Difference
	M	SD	M	SD	
Mental Demand	5.75	2.3	6.5	1.57	-0.75
Physical Demand	2.33	1.15	2.33	1.3	0
Temporal Demand	2.83	2.4	1.67	0.78	1.17
Performance	5.25	2.45	5.75	2.22	-0.5
Effort	6	2.22	5.41	1.72	0.58
Frustration	4.92	2.78	4.5	2.35	0.41

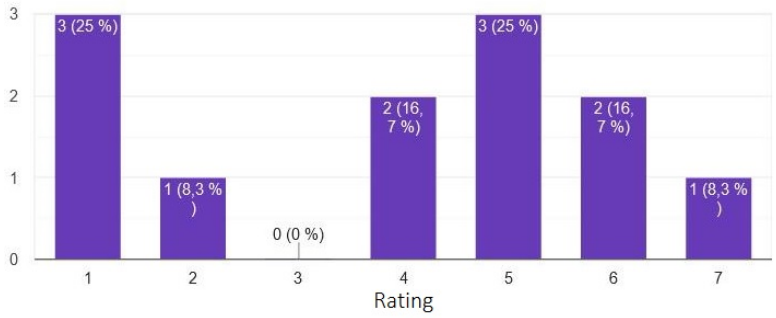


Fig. 12. Answers to the statement “I found the automatic pausing to be more distracting than helpful.” 1 = fully disagree, 7 = fully agree; n=12

that pausing interfered with the reading and the Muse headset was distracting. Those who favored the attention-aware version said the text was more legible since it did not update as frequently, and it seemed less stressful because the app adapted to their cognitive state. Figure 12 shows the participants agreement with the statement "I found the automatic pausing to be more distracting than helpful.” Participants who liked the BCI version rated the question with an average of 2.8, whereas the other group gave it a rating of 4.7. The latter is closer to a neutral value of 4. The participants who liked BCI leaned more towards 1 than the other group did towards 7, indicating a clearer preference.

H1 suggested that "The larger the percentage of pauses during thinking, the higher the experienced usability and the lower the task load". To evaluate the hypothesis, we calculated the correlation of the SUS and NASA-TLX scores with the task pauses percentage (see Figure 13). Regarding the task load (represented by the NASA-TLX), there was no correlation between the scores and the percentage of pauses during the task. For the usability assessment (represented by the SUS), we found a weak positive correlation of $r = 0.29$ which is however not significant ($p = 0.36$). We would have expected a positive correlation between pauses and the SUS score and a negative correlation between the pauses and the NASA-TLX. With the current results, *H1* can not be supported but the SUS results are promising for a larger test set.

H2 stated "The larger the percentage of pauses during reading, the lower the experienced usability and the higher the task load". To evaluate this hypothesis, we calculated the correlation of the SUS and the NASA-TLX scores with the translation pasues during reading (see Figure 14). While both correlations are not statistically significant with p-values of 0.2 and 0.12, they do demonstrate a tendency that supports *H2* with absolute r-values greater than 0.4. The bigger the

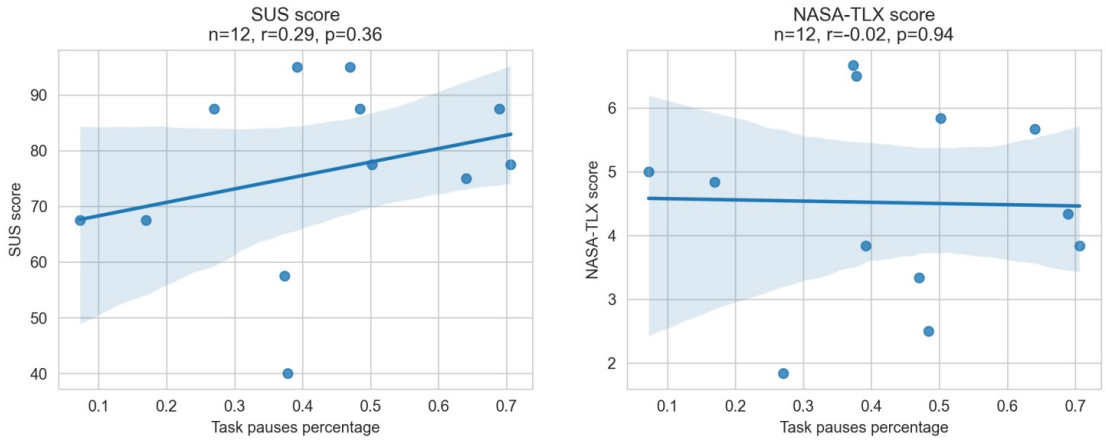


Fig. 13. Questionnaire scores of SUS and NASA-TLX correlated (Pearson) with the percentage of pauses during task. For the task, the data of OT and the first bin of TR was used, as for these it is more likely that participants were thinking.

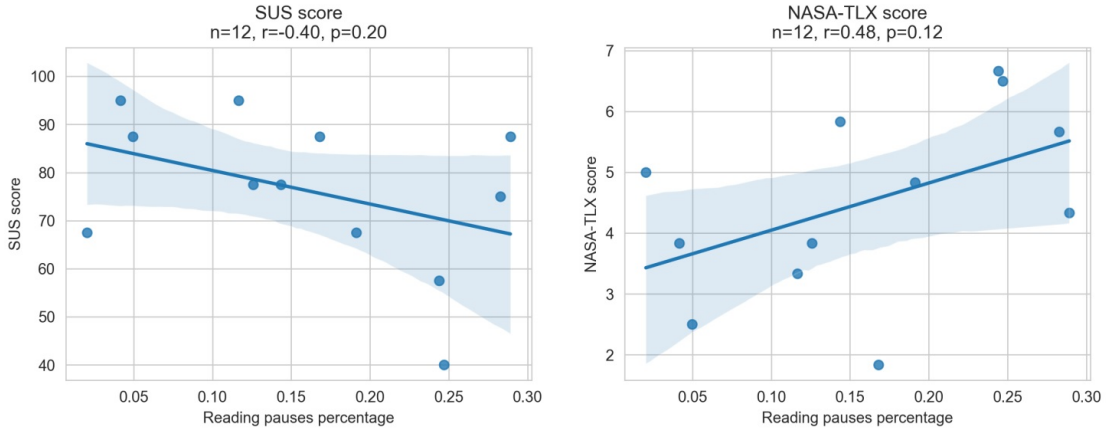


Fig. 14. SUS and NASA TLX correlated (Pearson) with the percentage of time that translations were paused during reading. The pauses percentage is the mean of the four posters.

percentage of pauses, the worse the score for both questions (higher score for NASA-TLX is worse). As a result, it is likely that *H2* is proven to be true with more participants. At this point, the results do not support *H2*.

For the evaluation of *H3*, participants were asked how much they agreed with the statement "The more distracting translations were during thinking, the more helpful the pausing was for me," which they would agree with if H3 were accurate. The results are shown in Figure 15. The average score was 4.3 (SD 1.7), indicating that participants tended to agree more than disagree. This supports *H3*. Participants were also asked if they thought the pausing was more distracting than useful (see Figure 12), and those who thought it was more distracting tended to disagree with the statement in Figure 15. The Spearman's correlation yielded an r-value of -0.78 with a p-value of 0.01 indicating that participants appeared to agree with this statement if they regarded the pausing to be useful.

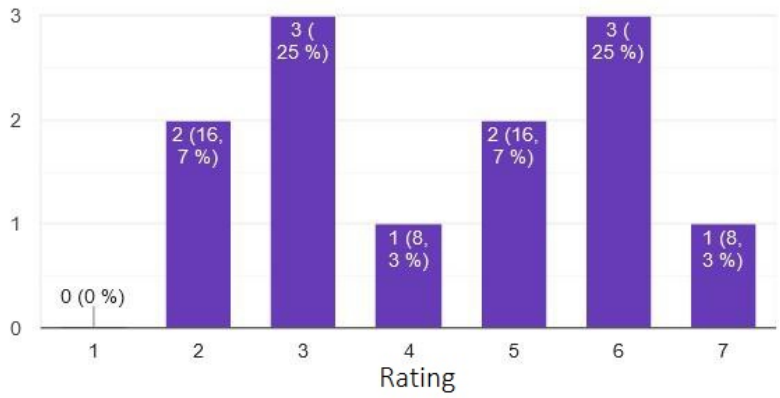


Fig. 15. Answers for agreement to the statement related to *H3* "The more distracting translations were during thinking, the more helpful the pausing was for me". 1 = fully disagree, 7 = fully agree; n=12

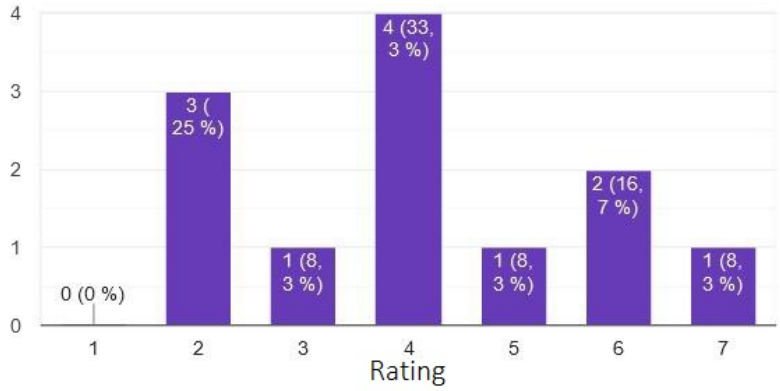


Fig. 16. Answers for agreement to the statement related to *H4* "The more I had to think during a question, the more helpful the pausing was for me". 1 = fully disagree, 7 = fully agree; n=12

Finally, *H4* was evaluated using another questionnaire question. *H4* stated that "The more demanding the combination of text and question is, the more helpful pauses are during thinking". The agreement to the statement "The more I had to think during a question, the more helpful the pausing was for me" can be seen in Figure 16. The mean of 4.1 (SD 1.7) is lower than for the *H3* statement. The results for this statement were also correlated with the results for the statement of whether participants felt the pausing to be more distracting than useful, and a negative correlation ($r = -0.63, p < 0.05$) was found.

In Conclusion. There is evidence for both *H1* and *H2*, since task pauses correlate with high SUS and NASA-TLX scores and reading pauses with low SUS and NASA-TLX scores, albeit this was not statistically significant. While *H3* and *H4* are likewise difficult to answer, participants who preferred the version with BCI appeared to agree with the hypotheses' statements.

5 DISCUSSION

We performed a user study to test the effects of the BCI integration into the AR translator application. The main novelty and contribution of this research was the mobile setup consisting of a smartphone and a consumer-grade EEG headset. We hypothesized that attentional state adaptation of text updates reduces distraction caused by the AR elements and increases the usability of the system. These hypotheses could not yet be proven with the current data. However, there are several other inferences and results, as well as indications that attention sensitivity can be helpful.

5.1 Use Case

As an authentic mobile BCI setting use case, we decided to use an AR translator because it is frequently used "in the wild". Using a smartphone to display the AR content is automatically less visually distracting than a head-mounted display that is typically used for Augmented Reality. The usability of such AR-glasses could probably be improved even more by adding attention awareness compared to the cell phone application.

The content for the translations was made up of Posters containing travel related information. As previously stated, only contiguous texts were employed in the study; nevertheless, discontiguous texts may also be translated by users. And these texts are likely to have a distinct distribution of reading and thinking time, because disjoint texts require less reading (e.g., book covers, shop window labeling, menus, consumer product labels, maps) and thinking is significantly dependent on the user. It is conceivable that a user studies a menu and spends considerable time inwardly discussing which food to order. In such instance, the AR translator would have more chances to halt when thinking than while reading. Such a use case might be even more appropriate to test the enhancement through attention-sensitive BCIs.

5.2 AR Translator Application

Participants were generally able to use the translator to understand texts of foreign language, judging by the subjective rating of their answers to the questions (72% of answers considered good). The majority of participants preferred the version without BCI (58.3%, n=7). The main reason for the preference of modality was the pausing during reading. For some participants, the reading pauses made the text more legible, as the text did not switch as much. Other participants disliked the pauses, as they interrupted the reading flow. The pausing during thinking did not play a deciding role, however it seems that it was generally regarded as positive. Because of the study design, the appearance of textstickers played only a minor role, which is presumably why the main complaints participants indicated in the AR translator questionnaire were the smoothness of translations and the quality of translations, both of which relate to understanding. Needless to say, the legibility of textstickers was critical for understanding in the study, although it did not really matter if a textsticker had red or black text to understand it. It is interesting that participants evaluated "matching colors of translations" higher than "translations replaced text visually authentically" (means: 5.4 versus 4.8). The purpose of an AR Translator is to create visually accurate textstickers, and matching textstickers merely in color does not appear to be sufficient. It should be emphasized that half of the posters were intended to cause problems for the Ar Translator. Posters 3 and 4 of both poster sets have letters with low contrast to the backdrop, while the background is an image with many varied colors that alternate often throughout the image. This is particularly difficult for OCR, and likely negatively affected the ratings of participants regarding the answers related to visual effects. Another issue raised by participants was the quality of translations. Because the OCR recognized text line by line, the translations were done line by line. It would make sense to utilize an OCR that returns text paragraph by paragraph, as the context between

lines is critical for translation tools to construct meaningful translations; however, PaddleOCR does not support this. Grouping lines after OCR is a non-optimal workaround, as paragraph detection within OCR is the faster method. This is because the image is already being processed by the OCR, making it more resource-efficient.
Overall, the visual authenticity of textstickers was rated well but there is still room for improvement of the AR translator application in general. This is most obvious if it is used to understand relatively large texts. However, the quality of the implemented application was good enough for our purposes and the mentioned shortcomings most likely did not severely interfere with the study outcome.

5.3 BCI Integration

The general setup using two smartphones and the Muse headset worked very well and calibration times were fast and easy. The system was rather comfortable to wear and the data quality seems sufficient for these purposes. The achieved classification accuracy of almost 60% was comparable to other real-time BCI applications in lab settings [46]. The main issue in this study was the label noise due to the rereading of text in time of internally directed attention. For future studies, it should be assured that externally and internally directed attention are clearly separable for the generation of ground truths to evaluate the classifier. Another study design related issue is the learning effect that may cause the second modality in execution order to be rated better. That is why looking at individual participants may not be expedient. Furthermore, judging upon the subjective answer quality, the difficulty of the two poster sets appears to differ slightly.

It is possible to improve the accuracy by adjusting the calibration configuration.Participants used augmented reality to read German literature. While it is encouraging that the text was read with AR in a manner similar to that of the AR translator, there is likely still a difference in elicited brain signals when comparing the calibration reading to the reading with the AR translator. Reading with an AR translator differs from conventional reading in that the flow of reading is frequently disrupted, either because translations swap while reading or because they do not always make sense. This is likely to result in confusion, possibly even frustration, or other reactions. A more specific calibration, in which participants read using the AR translator, would most certainly improve overall accuracy. Similarly, the thinking phase of the calibration could entail needing to glance at the text while the AR translator is paused. The calibration and main section of the study would then be more comparable.
In the long run, however, it would be desirable to exclude the need for collecting person-dependent training data to set up the classifier. This would increase the usability of a mobile BCI smartphone system by a lot. Vortmann and Putze [47] addressed the idea of person-independent EEG-based classification for internally and externally directed attention and found that Neural Networks outperform an LDA approach for 4 second data windows in such cases. The analysis in their paper were performed offline and a lot of training data was available for the classifier. No information about the computation time is given. For a training-free real-time application the choice of classification algorithm would have to be reevaluated. Again, the goal would be to find a compromise between computation time and classification accuracy.

5.4 Attentional State Application Adaptation

The comparison of both versions of the system showed that participants decided for or against a version based on the pausing during reading. While 4 participants expressed that the pausing interrupted their flow of reading, another 4 participants expressed that it helped them to focus on the reading. Only one participant commented on the pauses while thinking, stating that the app met their needs. Another participant claimed that pausing while thinking was

beneficial, but because they were focused on the work, it had no effect on them. This opinion was echoed by several subjects following the study when they were interviewed.

It was not anticipated while preparing the questionnaire or planning the study that participants would find the pause while reading useful. As a result, some questions did not directly inquire about pausing when thinking, but rather about halting in general. As a result, it is unclear whether participants felt the pausing to be beneficial because of the reading or because of the thinking, and by how much. According to the open-ended responses of those who preferred the version with BCI, pausing during reading was more useful than pausing during thinking.It is unknown how much of their choice is dependent on the pauses during thought. When asked about the pauses during thinking, the participants' responses were typically positive. However, because no participant explicitly said this in their open-ended response in the final questionnaire, those pauses do not appear to be a determining factor.

The way the classification result was integrated into the behavior of the application sometimes led to pauses of over 30 seconds before a new update would appear. This is far too long and should be restricted in later experiments.

5.5 Conclusions

This was the first attempt at developing a mobile BCI system using a smartphone and a low-cost EEG headset with few electrodes to add attention-awareness to an Augmented Reality application. Previously, the idea that the application's attention-awareness would improve an AR application was only tested for head-mounted displays and in a laboratory setting. We chose an AR translator inspired by Google Translate as a use case because it is a popular AR smartphone app that is frequently used on the road and provides several switches between internally and externally directed attention. A travel scenario was used for the user study, and foreign tests had to be read and understood.

The findings of our user study did not entirely support the claim that the BCI improves usability. However, the classification accuracy and ease of setup demonstrated that the system design is promising. Other use cases, usage scenarios, display devices, or applications should be investigated for future studies aimed at mobile attention-aware AR systems. For example, the precise nature of the interface's attention-adaptation or system behavior must be thought through and efficient in order to reduce distractiveness.

REFERENCES

[1] Sonja Annerer-Walcher, Simon M Ceh, Felix Putze, Marvin Kampen, Christof Körner, and Mathias Benedek. 2021. How Reliably Do Eye Parameters Indicate Internal Versus External Attentional Focus? *Cognitive Science* 45, 4 (2021), e12977.

[2] P Aricò, G Borghini, G Di Flumeri, N Sciaraffa, and F Babiloni. 2018. Passive BCI beyond the lab: current trends and future directions. *Physiological measurement* 39, 8 (2018), 08TR02.

[3] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[4] Mathias Benedek, Rainer J Schickel, Emanuel Jauk, Andreas Fink, and Aljoscha C Neubauer. 2014. Alpha power increases in right parietal cortex reflects focused internal attention. *Neuropsychologia* 56 (2014), 393–400.

[5] Bernard Rummel 2021. *System Usability Scale – jetzt auch auf Deutsch.* Retrieved November 12th, 2021 from "https://blogs.sap.com/2016/02/01/system-usability-scale-jetzt-auch-auf-deutsch/"

[6] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[7] Ok-Hue Cho and Won-Hyung Lee. 2014. BCI sensor based environment changing system for immersion of 3D game. *International Journal of Distributed Sensor Networks* 10, 5 (2014), 620391.

[8] Marvin M Chun, Julie D Golomb, and Nicholas B Turk-Browne. 2011. A taxonomy of external and internal attention. *Annual review of psychology* 62 (2011), 73–101.

[9] Colin David Conrad and Michael Bliemel. 2016. Psychophysiological measures of cognitive absorption and cognitive load in e-learning applications. (2016).

[10] Nicholas R Cooper, Rodney J Croft, Samuel JJ Dominey, Adrian P Burgess, and John H Gruzelier. 2003. Paradox lost? Exploring the role of alpha oscillations during externally vs. internally directed attention and the implications for idling and inhibition hypotheses. *International journal of psychophysiology* 47, 1 (2003), 65–74.

[11] Jody Culham. 2003. Attention-grabbing motion in the human brain. *Neuron* 40, 3 (2003), 451–452.

[12] Yuning Du, Chenxia Li, Ruoyu Guo, Xiaoting Yin, Weiwei Liu, Jun Zhou, Yifan Bai, Zilin Yu, Yehua Yang, Qingqing Dang, et al. 2020. PP-OCR: A practical ultra lightweight OCR system. *arXiv preprint arXiv:2009.09941* (2020).

[13] Victor Fragoso, Steffen Gauglitz, Shane Zamora, Jim Kleban, and Matthew Turk. 2011. TranslatAR: A mobile augmented reality translator. In *2011 IEEE workshop on applications of computer vision (WACV)*. IEEE, 497–502.

[14] Leo Galway, Paul McCullagh, Gaye Lightbody, Chris Brennan, and David Trainor. 2015. The potential of the brain-computer interface for learning: a technology review. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*. IEEE, 1554–1559.

[15] Google Translate Application 2021. *Camera-based AR translation for mobile phones.* Retrieved November 9, 2021 from https://translate.google.com/intl/en/about/

[16] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[17] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[18] Han-Jeong Hwang, Soyoun Kim, Soobeom Choi, and Chang-Hwan Im. 2013. EEG-based brain-computer interfaces: a thorough literature survey. *International Journal of Human-Computer Interaction* 29, 12 (2013), 814–826.

[19] InteraXon Muse 2 2021. *TECHNICAL SPECIFICATIONS, VALIDATION, AND RESEARCH USE.* Retrieved November 12th, 2021 from "https://images-na.ssl-images-amazon.com/images/I/D1RREdoENNS.pdf"

[20] JavaOSC 2021. *library that gives JVM language programs the ability of working with OSC content format.* Retrieved November 12th, 2021 from "https://github.com/hoijui/JavaOSC"

[21] Nor Farzana Syaza Jeffri and Dayang Rohaya Awang Rambli. 2021. A review of augmented reality systems and their effects on mental workload and task performance. *Heliyon* 7, 3 (2021), e06277.

[22] Mark Joselli, Fabio Binder, Esteban Clua, and Eduardo Soluri. 2017. Concept, development and evaluation of a mind action game with the electro encephalograms as an auxiliary input. *SBC Journal on Interactive Systems* 8, 1 (2017), 60–73.

[23] Adrian Kaehler and Gary Bradski. 2016. *Learning OpenCV 3: computer vision in C++ with the OpenCV library.* " O'Reilly Media, Inc.".

[24] Olave E Krigolson, Chad C Williams, Angela Norton, Cameron D Hassall, and Francisco L Colino. 2017. Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. *Frontiers in neuroscience* 11 (2017), 109.

[25] Vojkan Mihajlović, Bernard Grundlehner, Ruud Vullers, and Julien Penders. 2014. Wearable, wireless EEG solutions in daily life applications: what are we missing? *IEEE journal of biomedical and health informatics* 19, 1 (2014), 6–21.

[26] Koji Mikami, Kunio Kondo, et al. 2017. Adaptable Game Experience Based on Player's Performance and EEG. In *2017 Nicograph International (NicoInt)*. IEEE, 1–8.

[27] mindmonitor 2021. *Real time EEG graphs from your Interaxon Muse headband.* Retrieved November 12th, 2021 from "https://mind-monitor.com/"

[28] Michael D Mrazek, Dawa T Phillips, Michael S Franklin, James M Broadway, and Jonathan W Schooler. 2013. Young and restless: validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Frontiers in psychology* 4 (2013), 560.

[29] Muse 2 Headset 2021. *Consumer-grade EEG Headband.* Retrieved November 12th, 2021 from "https://choosemuse.com/muse-2/"

[30] MuseIO 2021. *Muse for developers.* Retrieved November 12th, 2021 from https://web.archive.org/web/20181105231756/http://developer.choosemuse.com/tools/available-data#Absolute_Band_Powers

[31] Chang S Nam, Anton Nijholt, and Fabien Lotte. 2018. *Brain–computer interfaces handbook: technological and theoretical advances.* CRC Press.

[32] NASA-TLX 2021. *(Kurzfassung deutsch).* Retrieved November 12th, 2021 from "interaction-design-group.de/toolbox/wp-content/uploads/2016/05/NASA-TLX.pdf"

[33] PaddleOCR 2021. *Light-weight optical character recognition using neural networks.* Retrieved November 12th, 2021 from "https://github.com/PaddlePaddle/PaddleOCR"

[34] Aleksandra Przegalinska, Leon Ciechanowski, Mikolaj Magnuski, and Peter Gloor. 2018. Muse headband: Measuring tool or a collaborative gadget? In *Collaborative Innovation Networks*. Springer, 93–101.

[35] Felix Putze, Maximilian Scherer, and Tanja Schultz. 2016. Starring into the void? Classifying Internal vs. External Attention from EEG. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. 1–4.

[36] Raphaëlle N Roy and Jérémy Frey. 2016. Neurophysiological markers for passive brain–computer interfaces. *Brain–Computer Interfaces 1: Foundations and Methods* (2016), 85–100.

[37] Ana Rita de Tróia Salvado. 2015. *Augmented reality applied to language translation.* Ph. D. Dissertation.

[38] Sumit Soman, Siddharth Srivastava, Saurabh Srivastava, and Nitendra Rajput. 2015. Brain computer interfaces for mobile apps: State-of-the-art and future directions. *arXiv preprint arXiv:1509.01338* (2015).

[39] Desney Tan and Anton Nijholt. 2010. Brain-computer interfaces and human-computer interaction. In *Brain-Computer Interfaces*. Springer, 3–19.

[40] Lamma Tatwany and Henda Chorfi Ouertani. 2017. A review on using augmented reality in text translation. In *2017 6th International Conference on Information and Communication Technology and Accessibility (ICTA)*. IEEE, 1–6.

[41] Takumi Toyama, Daniel Sonntag, Andreas Dengel, Takahiro Matsuda, Masakazu Iwamura, and Koichi Kise. 2014. A mixed reality head-mounted text translation system using eye gaze input. In *Proceedings of the 19th international conference on Intelligent User Interfaces*. 329–334.

[42] Gabriel Alves Mendes Vasiljevic and Leonardo Cunha de Miranda. 2020. Brain–computer interface games based on consumer-grade EEG Devices: A systematic literature review. *International Journal of Human–Computer Interaction* 36, 2 (2020), 105–142.

[43] Adrian von Mühlenen, Mark I Rempel, and James T Enns. 2005. Unique temporal change is the key to attentional capture. *Psychological Science* 16, 12 (2005), 979–986.

[44] Lisa-Marie Vortmann, Jannes Knychalla, Sonja Annerer-Walcher, Mathias Benedek, and Felix Putze. 2021. Imaging Time Series of Eye Tracking Data to Classify Attentional States. *Frontiers in Neuroscience* 15 (2021), 625.

[45] Lisa-Marie Vortmann, Felix Kroll, and Felix Putze. 2019. EEG-based classification of internally-and externally-directed attention in an augmented reality paradigm. *Frontiers in human neuroscience* 13 (2019), 348.

[46] Lisa-Marie Vortmann and Felix Putze. 2020. Attention-aware brain computer interface to avoid distractions in augmented reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[47] Lisa-Marie Vortmann and Felix Putze. 2021. Exploration of Person-Independent BCIs for Internal and External Attention-Detection in Augmented Reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 2 (2021), 1–27.

[48] Lisa-Marie Vortmann, Moritz Schult, Mathias Benedek, Sonja Walcher, and Felix Putze. 2019. Real-time multimodal classification of internal and external attention. In *Adjunct of the 2019 International Conference on Multimodal Interaction*. 1–7.

[49] Thorsten O Zander, Lena M Andreessen, Angela Berg, Maurice Bleuel, Juliane Pawlitzki, Lars Zawallich, Laurens R Krol, and Klaus Gramann. 2017. Evaluation of a dry EEG system for application of passive brain-computer interfaces in autonomous driving. *Frontiers in human neuroscience* 11 (2017), 78.

[50] Thorsten O Zander, Christian Kothe, Sabine Jatzev, and Matti Gaertner. 2010. Enhancing human-computer interaction with input from active and passive brain-computer interfaces. In *Brain-computer interfaces*. Springer, 181–199.

B

Addendum

List of Publications

The following bibliography lists all publications of which the author is author or co-author.

Conference Proceedings

2022 **Lisa-Marie Vortmann**, Moritz Schult, and Felix Putze (2022). *Differentiating Endogenous and Exogenous Attention Shifts Based on Fixation-Related Potentials.* 27th International Conference on Intelligent User Interfaces. New York, NY, USA: Association for Computing Machinery.

2021 **Lisa-Marie Vortmann**, Jonas Klaff, Timo Urban, and Felix Putze (2021). *SSVEP-Aided Recognition of Internally and Externally Directed Attention from Brain Activity.* Proceedings of the 2021 IEEE International Conference on Systems, Man, and Cybernetics. IEEE.

2020 Felix Putze, Merlin Burri, **Lisa-Marie Vortmann**, and Tanja Schultz (2020). *Model-Based Prediction of Exogeneous and Endogenous Attention Shifts During an Everyday Activity.* Companion Publication of the 2020 International Conference on Multimodal Interaction. New York, NY, USA: Association for Computing Machinery.

Lisa-Marie Vortmann and Felix Putze (2020). *Attention-Aware Brain Computer Interface to Avoid Distractions in Augmented Reality.* Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems.

2019 **Lisa-Marie Vortmann** (2019). *Attention-driven Interaction Systems for Augmented Reality.* 2019 International Conference on Multimodal Interaction. New York, NY, USA: Association for Computing Machinery.

Felix Putze, Dennis Weiß, **Lisa-Marie Vortmann**, and Tanja Schultz (2019). *Augmented Reality Interface for Smart Home Control using SSVEP-BCI and Eye Gaze.* Proceedings of the 2019 IEEE International Conference on Systems, Man, and Cybernetics. IEEE.

Lisa-Marie Vortmann, Moritz Schult, and Felix Putze (2019). *Real-Time Multimodal Classification of Internal and External Attention.* Adjunct of the 2019 International Conference on Multimodal Interaction. New York, NY, USA: Association for Computing Machinery.

Journal Publications

2022 **Lisa-Marie Vortmann** and Felix Putze (2022). *Multi-Modal EEG and Eye Tracking Feature Fusion Approaches for Attention Classification.* Frontiers in Computer Science, 4(780580).

2021 **Lisa-Marie Vortmann**, Leonid Schwenke, and Felix Putze (2021). *Using Brain Activity Patterns to Differentiate Real and Virtual Attended Targets during Augmented Reality Scenarios.* Information, 12(6).

Lisa-Marie Vortmann and Felix Putze (2021). *Exploration of Person-In-dependent BCIs for Internal and External Attention-Detection in Augmented Reality.* Proceedings of the ACM Interaction of Mobile Wearable Ubiquitous Technology, 5(2).

Lisa-Marie Vortmann, Jannes Knychalla, Sonja Annerer-Walcher, Mathias Benedek, and Felix Putze (2021). *Imaging Time Series of Eye Tracking Data to Classify Attentional States.* Frontiers in Neuroscience, 15.

Lisa-Marie Vortmann and Felix Putze (2021). *Combining Implicit and Explicit Feature Extraction for Eye Tracking: Attention Classification Using a Heterogeneous Input.* Sensors, 21(24).

2019 **Lisa-Marie Vortmann**, Felix Kroll, and Felix Putze (2019). *EEG-Based Classification of Internally-and Externally-Directed Attention in an Augmented Reality Paradigm.* Frontiers in Human Neuroscience, 13.

Under Review

2022 **Lisa-Marie Vortmann**, Pascal Weidenbach, and Felix Putze (2022). *Attention-Aware Translation Application in Augmented Reality for Mobile Phones*.

Lisa-Marie Vortmann, Timo Urban, and Felix Putze. (2022). *Machine Learning from Mistakes: Self-correcting Attention Classifier using Error-Potentials.*