Workload-Aware Systems and Interfaces for Cognitive Augmentation
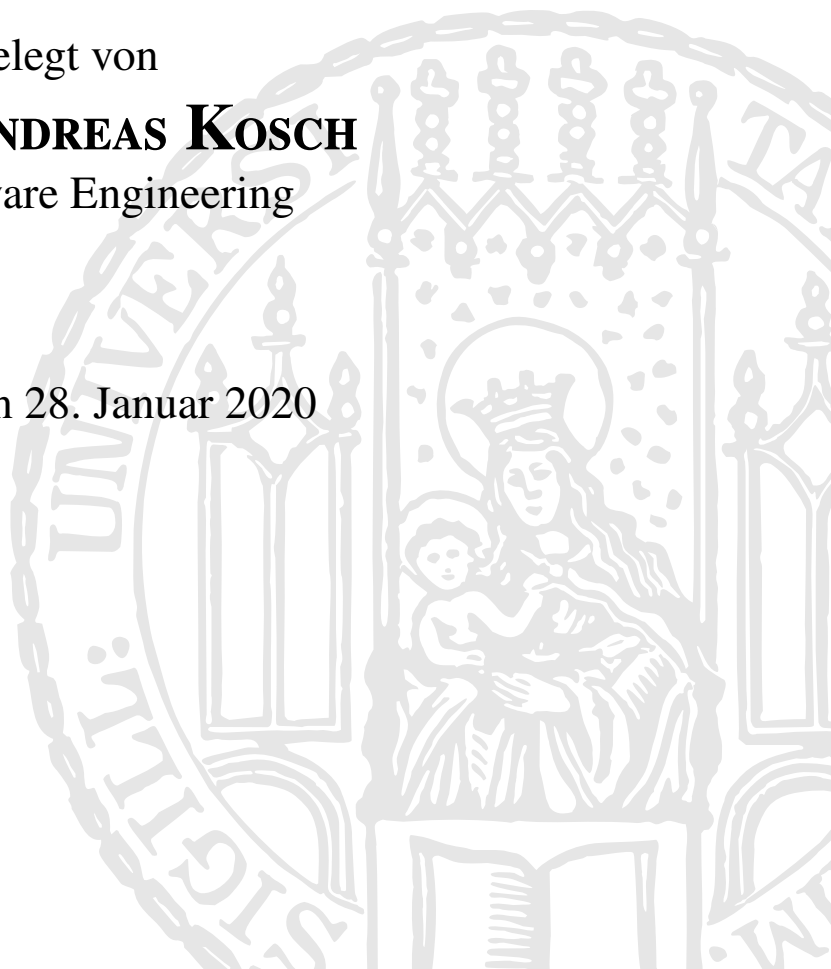
Dissertation

an der Fakultät für Mathematik, Informatik und Statistik
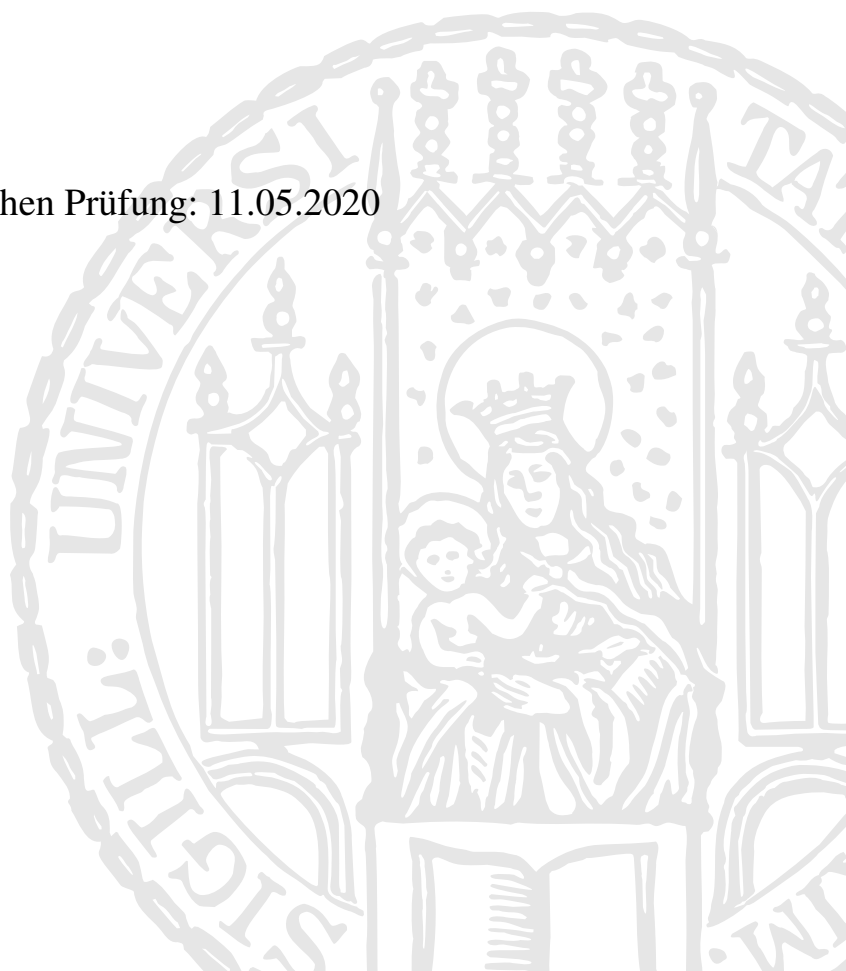der Ludwig-Maximilians-Universität München

vorgelegt von

Thomas-Andreas Kosch
M.Sc. Software Engineering

München, den 28. Januar 2020

Erstgutachter: Prof. Dr. Albrecht Schmidt
Zweitgutachter: Prof. Dr. Fabio Paternò
Drittgutachter: Prof. Dr. Jonathan Lazar

Tag der mündlichen Prüfung: 11.05.2020

ABSTRACT

In today's society, our cognition is constantly influenced by information intake, attention switching, and task interruptions. This increases the difficulty of a given task, adding to the existing workload and leading to compromised cognitive performances. The human body expresses the use of cognitive resources through physiological responses when confronted with a plethora of cognitive workload. This temporarily mobilizes additional resources to deal with the workload at the cost of accelerated mental exhaustion.

We predict that recent developments in physiological sensing will increasingly create user interfaces that are aware of the user's cognitive capacities, hence able to intervene when high or low states of cognitive workload are detected. In this thesis, we initially focus on determining opportune moments for cognitive assistance. Subsequently, we investigate suitable feedback modalities in a user-centric design process which are desirable for cognitive assistance. We present design requirements for *how* cognitive augmentation can be achieved using interfaces that sense cognitive workload.

We then investigate different physiological sensing modalities to enable suitable real-time assessments of cognitive workload. We provide empirical evidence that the human brain is sensitive to fluctuations in cognitive resting states, hence making cognitive effort measurable. Firstly, we show that electroencephalography is a reliable modality to assess the mental workload generated during the user interface operation. Secondly, we use eye tracking to evaluate changes in eye movements and pupil dilation to quantify different workload states. The combination of machine learning and physiological sensing resulted in suitable real-time assessments of cognitive workload. The use of physiological sensing enables us to derive *when* cognitive augmentation is suitable.

Based on our inquiries, we present applications that regulate cognitive workload in home and work settings. We deployed an assistive system in a field study to investigate the validity of our derived design requirements. Finding that workload is mitigated, we investigated how cognitive workload can be visualized to the user. We present an implementation of a biofeedback visualization that helps to improve the understanding of brain activity. A final study shows how cognitive workload measurements can be used to predict the efficiency of information intake through reading interfaces. Here, we conclude with use cases and applications *which* benefit from cognitive augmentation.

This thesis investigates how assistive systems can be designed to implicitly sense and utilize cognitive workload for input and output. To do so, we measure cognitive workload in real-time by collecting behavioral and physiological data from users and analyze this data to support users through assistive systems that adapt their interface according to the currently measured workload. Our overall goal is to extend new and existing context-aware applications by the factor *cognitive workload*. We envision *Workload-Aware Systems* and *Workload-Aware Interfaces* as an extension in the context-aware paradigm. To this end, we conducted eight research inquiries during this thesis to investigate how to design and create workload-aware systems.

Finally, we present our vision of future workload-aware systems and workload-aware interfaces. Due to the scarce availability of open physiological data sets, reference implementations, and methods, previous context-aware systems were limited in their ability to utilize cognitive workload for user interaction. Together with the collected data sets, we expect this thesis to pave the way for methodical and technical tools that integrate workload-awareness as a factor for context-aware systems.

ZUSAMMENFASSUNG

Tagtäglich werden unsere kognitiven Fähigkeiten durch die Verarbeitung von unzähligen Informationen in Anspruch genommen. Dies kann die Schwierigkeit einer Aufgabe durch mehr oder weniger Arbeitslast beeinflussen. Der menschliche Körper drückt die Nutzung kognitiver Ressourcen durch physiologische Reaktionen aus, wenn dieser mit kognitiver Arbeitsbelastung konfrontiert oder überfordert wird. Dadurch werden weitere Ressourcen mobilisiert, um die Arbeitsbelastung vorübergehend zu bewältigen.

Wir prognostizieren, dass die derzeitige Entwicklung physiologischer Messverfahren kognitive Leistungsmessungen stets möglich machen wird, um die kognitive Arbeitslast des Nutzers jederzeit zu messen. Diese sind in der Lage, einzugreifen wenn eine zu hohe oder zu niedrige kognitive Belastung erkannt wird. Wir konzentrieren uns zunächst auf die Erkennung passender Momente für kognitive Unterstützung welche sich der gegenwärtigen kognitiven Arbeitslast bewusst sind. Anschließend untersuchen wir in einem nutzerzentrierten Designprozess geeignete Feedbackmechanismen, die zur kognitiven Assistenz beitragen. Wir präsentieren Designanforderungen, welche zeigen *wie* Schnittstellen eine kognitive Augmentierung durch die Messung kognitiver Arbeitslast erreichen können.

Anschließend untersuchen wir verschiedene physiologische Messmodalitäten, welche Bewertungen der kognitiven Arbeitsbelastung in Realzeit ermöglichen. Zunächst validieren wir empirisch, dass das menschliche Gehirn auf kognitive Arbeitslast reagiert. Es zeigt sich, dass die Ableitung der kognitiven Arbeitsbelastung über Elektroenzephalographie eine geeignete Methode ist, um den kognitiven Anspruch neuartiger Assistenzsysteme zu evaluieren. Anschließend verwenden wir Eye-Tracking, um Veränderungen in den Augenbewegungen und dem Durchmesser der Pupille unter verschiedenen Intensitäten kognitiver Arbeitslast zu bewerten. Das Anwenden von maschinellem Lernen führt zu zuverlässigen Echtzeit-Bewertungen kognitiver Arbeitsbelastung. Auf der Grundlage der bisherigen Forschungsarbeiten stellen wir Anwendungen vor, welche die Kognition im häuslichen und beruflichen Umfeld unterstützen. Die physiologischen Messungen stellen fest, *wann* eine kognitive Augmentierung sich als günstig erweist.

In einer Feldstudie setzen wir ein Assistenzsystem ein, um die erhobenen Designanforderungen zur Reduktion kognitiver Arbeitslast zu validieren. Unsere Ergebnisse zeigen, dass die Arbeitsbelastung durch den Einsatz von Assistenzsystemen reduziert wird. Im Anschluss untersuchen wir, wie kognitive Arbeitsbelastung visualisiert werden kann. Wir stellen eine Implementierung einer Biofeedback-Visualisierung vor, die das Nutzerverständnis zum Verlauf und zur Entstehung von kognitiver Arbeitslast unterstützt. Eine abschließende Studie zeigt, wie Messungen kognitiver Arbeitslast zur Vorhersage der aktuellen Leseeffizienz benutzt werden können. Wir schließen hierbei mit einer Reihe von Applikationen ab, *welche* sich kognitive Arbeitslast als Eingabe zunutze machen.

Die vorliegende wissenschaftliche Arbeit befasst sich mit dem Design von Assistenzsystemen, welche die kognitive Arbeitslast der Nutzer implizit erfasst und diese bei der Durchführung alltäglicher Aufgaben unterstützt. Dabei werden physiologische Daten erfasst, um

Rückschlüsse in Realzeit auf die derzeitige kognitive Arbeitsbelastung zu erlauben. Anschließend werden diese Daten analysiert, um dem Nutzer strategisch zu assistieren. Das Ziel dieser Arbeit ist die Erweiterung neuartiger und bestehender kontextbewusster Benutzerschnittstellen um den Faktor *kognitive Arbeitslast*. Daher werden in dieser Arbeit *arbeitslastbewusste Systeme* und *arbeitslastbewusste Benutzerschnittstellen* als eine zusätzliche Dimension innerhalb des Paradigmas kontextbewusster Systeme präsentiert. Wir stellen acht Forschungsstudien vor, um die Designanforderungen und die Implementierung von kognitiv arbeitslastbewussten Systemen zu untersuchen.

Schließlich stellen wir unsere Vision von zukünftigen kognitiven arbeitslastbewussten Systemen und Benutzerschnittstellen vor. Durch die knappe Verfügbarkeit öffentlich zugänglicher Datensätze, Referenzimplementierungen, und Methoden, waren Kontextbewusste Systeme in der Auswertung kognitiver Arbeitslast bezüglich der Nutzerinteraktion limitiert. Ergänzt durch die in dieser Arbeit gesammelten Datensätze erwarten wir, dass diese Arbeit den Weg für methodische und technische Werkzeuge ebnet, welche kognitive Arbeitslast als Faktor in das Kontextbewusstsein von Computersystemen integriert.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

List of Acronyms

AI	Artificial Intelligence
ANOVA	Analysis of Variance
AR	Augmented Reality
BCI	Brain-Computer Interface
BMBF	German Federal Ministry of Education and Research
CNN	Convolutional Neural Network
DAEM	Designing Assistive Environments for Manufacturing
DALI	Driver Activity Load Index
DL	Deep Learning
ECoG	Electrocorticography
EDA	Electrodermal Activity
EEG	Electroencephalography
ELSI	Ethical, Legal and Social Implications
EMG	Electromyography
EOG	Electrooculography
ERP	Event-Related Potential
FBN	Functional Brain Network
FFT	Fast Fourier Transform
fMRI	Functional Magnetic Resonance Imaging
fNIRS	Functional Near-Infrared Spectroscopy
GDPR	General Data Protection Regulation
GWW	Gemeinnützige Werkstätten und Wohnheime GmbH
HbO	Oxygenated Hemoglobin
HbR	Deoxygenated Hemoglobin
HCI	Human-Computer Interaction
HMD	Head-Mounted Display
HR	Heart Rate
HRV	Heart Rate Variability
Hz	Hertz
KoBeLU	Context-Aware Learning Environment
ML	Machine Learning
NASA-TLX	NASA Task Load Index
PI	Performance Index
RQ	Research Question
RSVP	Rapid Serial Visual Representation
RSME	Rating Scale of Mental Effort
SVM	Support Vector Machine

TCT	Task Completion Time
UI	User Interface
VR	Virtual Reality
WoZ	Wizard of Oz

I

INTRODUCTION AND FOUNDATIONS

Chapter 1

Introduction

The world has undergone a substantial change in using computing systems for information intake. Since the beginning of mankind, the speed at which information is delivered and consumed has increased dramatically and keeps accelerating with the increasing availability of past information platforms and the proliferation of modern media. Local transfer systems such as speaking and writing have evolved into a global information network with the advent of the World Wide Web. Thus, extensive knowledge has become available for the public through different modalities. Buckminster coined this as the *knowledge doubling curve* [126].

The distribution of information has started to form a society that obtains their knowledge from other sharing peers. The amount of available information is expected to rise further with ubiquitous tools that proliferate into our daily routine. The consistent availability of information requires constant attention and strains the cognitive demand of users. Various stimuli compete for the attention of people and it is up to the user which ones are eligible to be processed, requiring strategies to decide on how much of the available cognitive resources should be spent on specific information channels. Miksch and Schulz elaborate this aspect in their work *Disconnect to Reconnect*, where they found that the reduced use of digital technologies is associated with improvements in *performance*, *self-control*, *well-being*, and *maintenance of real-life relationships* [268].

Additionally, information has become more accessible with the proliferation of mobile ubiquitous computing devices. This includes smartphones, smartwatches, notebooks, and smart eye-wear technologies. These devices contain rich capabilities to provide output in mobile scenarios. This enforced a shift on how humans process information, since content can be consumed on-the-go or in any other stationary setting. For example, popular applications

include reading (*e.g.*, Kindle), auditory skimming (*e.g.*, Blinkist), language learning (*e.g.*, Duolingo), or activity on social media platforms (*e.g.*, Instagram).

However, with the establishment and an increasing number of apps that compete for cognitive resources, full attention cannot be paid to multiple stimuli that demand the same perception channels [327]. This results in mental fatigue, difficulties with concentration, or decline in attention. This ongoing trend is expected to continue with an increasing number of devices, which negatively impacts the operation of user interfaces in short and long-term usage. Cognitive overload due to complicated user interface operations or extended multitasking as well as cognitive underload because of boredom is the result. While user interfaces in personal home environments can be disposed of, safety and time-critical environments depend on reliable operators that require few cognitive resources for operation. The *Hawaii False Missile Incident* in 2018 [17], in which a false missile alert was issued due to a series of operation errors, shows the need for interfaces that are aware of the user's cognitive states to avoid far-reaching interaction mistakes.

Since cognitive resources change throughout the day, users find themselves at times where they can accomplish challenging tasks as well as times where they are unable to concentrate. Systems that implicitly assess cognitive states already exist in driving contexts where they evaluate eye movements to infer the driver's fatigue level. However, these are restricted to specific application scenarios. Mobile and stationary applications do not take the current task difficulty and the user's cognitive demand into account, and therefore, rely on explicit user actions (*e.g.*, user intervention) or trigger interruptions (*e.g.*, notifications) at times where users are focused, thus interrupting the workflow.

This thesis investigates how ubiquitous technologies can be used to sense the current level of cognitive demand of users while they interact with user interfaces. Users seldom express their mental demands explicitly, hence implicit detection methods for cognitive workload provide the potential for continuous monitoring instead of selective sampling. Ubiquitous technologies become equipped with rich sensing techniques that can determine cognitive and physical states. Human bodies emit physiological signals that mediate mental states, such as stress level or current level of physical activity. Advances in physiological sensing enable the quantification of the concurrent cognitive workload levels. Therefore, this thesis focuses on how systems achieve awareness about the user's perceived workload and explore applications that provide cognitive augmentation. This includes the design of assistive technologies that mitigate or foster cognitive workload (*i.e.*, investigating *how* to assist), evaluation of physiological sensing modalities that provide real-time metrics for cognitive workload (*i.e.*, creating methods *when* to assist), and the presentation of applications that benefit from these insights (*i.e.*, in *which* application scenarios to assist). This **workload-awareness** is envisioned as an extension of the context-aware computing paradigm and enables the development of novel application scenarios.

Workload-awareness allows applications to match the user's current available cognitive resources with the complexity of the presentation of a user interface. By estimating their available cognitive resources, opportune moments for cognitive augmentation can be selected

and matched with the preferred exhibited mental demand. By matching task requirements and cognition, peak performance levels are achieved while avoiding frustration and boredom. The scheduling of tasks towards a currently perceived or expected cognitive workload level can be arranged by modifying the schedule, adapting the presentation of the current task to the cognitive context, provide visualizations that inform the course of cognitive demand throughout the day, and enable designers to evaluate their user interfaces regarding the placed cognitive demand.

To build workload-aware systems that use physiological sensing as a marker for cognitive effort, we initially focus on the phenomena of cognitive capabilities and effective perception of information. At its core lies the theory of cognition, a crucial factor for perceiving, processing, memorizing, and responding to information. This thesis presents a series of studies that investigate the design requirements of systems that utilize cognitive workload and leverage physiological sensing to detect cognitive workload as a strategic metric for user interface adaptations and evaluation. We explore physiological characteristics in combination with machine learning techniques to reliably detect changes in cognitive demand across individuals. To complement, we provide use cases and implementations that quantify and use workload in real-time.

We develop tools that can be applied by researchers and engineers to create workload-aware interfaces that take the cognitive context into account. Therefore, we present algorithms and methods to detect the current level of mental demand, which includes workload-awareness as an additional variable within context-aware computing. We create and evaluate concrete workload-aware systems: the concepts developed in this thesis include the use of an in-situ Augmented Reality (AR) interface that supports persons with cognitive impairments during daily tasks, visualizations that include direct representation of how the brain is affected by cognitive processing, and the assessment of text reading efficiency. Taking cognitive capacities into account, interface designers receive novel tools to engage with users, implicitly evaluate interfaces, and control the information flow.

1.1   Vision: Workload-Aware Interfaces

Sensing cognitive workload has become relevant in different scenarios, such as workplaces or in educational settings, to provide self-reflection or content-based adaptation. This requires ubiquitous methods to record and analyze user actions. Since we are integrating technologies in our daily lives, clothes, environments, or even our bodies, seamless physiological data collection and analysis can be carried out. We believe that such technologies can support people relative to workload measurements. In the following, we outline our vision of workload-aware systems that facilitate user interaction throughout the day.

The goal of technology is to integrate seamlessly into the environment and support users [381]. Thus, the more devices know about physical and physiological constraints or states, the better they can provide contextual assistance. A fully integrated piece of tech-

nology knows *when* to approach users and *how* to provide adaptive support with a *suitable* adaptivity level to lower or increase cognitive workload. This process includes multiple steps. First, user states have to be detected. Physical and physiological data is collected and analyzed in real-time to estimate the current level of cognitive workload. Second, the user needs to be approached at the right time to avoid a mismatch between user expectancy and system. The final step copes with the content modification that assesses environmental factors suitable for displaying the newly adapted content. Reinforcement-based assessment algorithms determine if an improvement in physical and physiological measures has taken place. If not, the agent can modify spatial and temporal factors to optimize the user experience. Repeating this process enables the agent to optimize workload-related parameters for the individual user. We define workload-aware systems in the following:

> **Definition** *Workload-Aware System*
>
> *Workload-Aware systems provide interfaces that enable implicit and explicit sensing of cognitive workload. Workload-aware interfaces integrate cognitive workload as a dimension into the paradigm of context-aware computing which can be taken into account for user interface evaluation, workload reflection, or the creation of adaptive user interfaces.*

A workload-aware system needs information about the users' current actions as well as tasks that have to be accomplished to fulfill the definition. Cognitive default states (*i.e.*, resting or engaged states) need to be monitored across time to detect deviations. Furthermore, the system must be aware of its environment to provide support at the right place and at the right time. The idea of context-aware computing is to use several extrinsic and intrinsic factors to provide support for users just-in-time. Workload-aware systems focus on cognitive workload to augment the user's mental processing. Those systems can sense highs and lows in cognitive workload throughout the day and can enhance cognition. Systems that are aware of the users' current workload throughout the day can support and enhance the cognitive capabilities in various ways:

User Interface Adaptation

Being aware of the current cognitive capacities of the user can be used as a parameter to influence the representation of a task. An adaptive user interface can provide more complex task representations to engage the user and increase the overall task success rate if sufficient cognitive capacities are left or untouched. Providing adaptive in-situ assistance during daily tasks such as cooking and at manual assembly workplaces can provide cognitive alleviation. Switching assistance modalities in real-time according to the current cognitive workload enables users to effectively engage and solve tasks in a shorter amount of time. At the same time, a system can detect boredom, reduce the provided assistance, and increase the commitment of users' to the current task.

Evaluation Tool

Subjective measures, such as questionnaires, interviews, or think-aloud protocols are heavily employed to gain insights into the usability of user interfaces. For example, questionnaires have established as a measure for workload where cognitive demand represents a subset for quantification [161, 297]. However, these methods are prone to subjective perception, require participants to memorize their experience regarding workload, and do not allow real-time insights into cognitive processing during an experiment. For example, knowledge about the current cognitive state allows experimenters and user interface designers to find distracting or unclear elements. A workload-aware system that points this out can be used by the experimenter afterward to explore difficulties in the User Interface (UI) design in post hoc interviews.

Reflection

Reflecting states of cognitive workload is the initial step to self-improvement. By making cognitive states spatially and temporally accessible, users can reflect on their performance throughout the day to schedule demanding tasks at suitable time slots. Hence, productive phases can be identified and exploited. On the other hand, tasks that do not require cognitive attention can be scheduled to time slots where the user is aware of having low cognitive capacities. For example, a user or a system can decide when it is the best time to whether to read a complex scientific article or when to wash the dishes. Synchronizing task difficulty and cognitive workload have the potential to reduce frustration as well as boredom. Instead of communicating cognitive workload only with the user, it can be shared with other persons and agents. This enables collaborative scenarios, in which multi-agents schedule tasks, assign tasks automatically to other persons, match collaboration partners, or communicate the perceived cognitive workload with the environment. This provides other persons and the environment with the ability to reflect on the cognitive workload of other individuals.

Individual Assistance

Workload-aware systems may become individual assistants who use recent advances in Artificial Intelligence (AI) to create models from cognitive and behavioral patterns. These models are beneficial when it comes to schedule tasks and appointments according to the required cognitive performance. Such devices weave altogether with the required performance and upcoming tasks. By sharing these models, others can synchronize their capacities in workload to optimize the outcomes. The fluctuation of workload depends on multiple parameters, such as nutrition, quality of sleep, workout routine, personal wellbeing, upcoming daily chores, or incoming information consumption. For example, a meeting can be scheduled when a set of parameters that influence workload are optimized among all participants. Teaching technology how cognitive workload affects the priority of tasks throughout

the day goes beyond simple context-aware applications. Workload-aware interfaces ubiquitously provide cognitive alleviation by matching chores to cognitive workload and structure tasks in a way to optimize the overall performance. This has the potential to help users to be efficient at their tasks, reduce boredom as well as frustration, and increase well-being. Going one step further, workload-aware interfaces can autonomously assess the difficulty of a task while being aware of the user's available cognitive resources. These tasks can then be matched to the current cognitive fitness to ease its processing, hence providing a workload-aware task schedule.

Interruption Management

Interruptions in physical as well as virtual environments drain attention and reduce the capability to remain focus on a task. Workload-aware systems become an interface between a user and other individuals to communicate the current state of workload. Users may not be interrupted by notifications or inconvenient office visits. For example, a workload-aware system can mediate the current cognitive state using lights or heat on the door handle. Furthermore, senders of messages may be notified that the current message is delayed and should be forwarded only under important circumstances. Flow states are fostered due to times in which high focus is required, but also for situations in which interruptions are inappropriate. For example, disruptions during highly focused conversations between individuals can be proactively prevented and delayed to a more suitable time slot. Workload-aware systems can, therefore, serve as a potential protection mechanism for periods that require high focus.

1.2   Research Questions

Three pivotal aspects have to be considered before integrating workload-awareness into ubiquitous computing technologies: the users and their context, available workload sensing modalities, and the application. Table 1.1 provides an overview of the Research Question (RQ) that are addressed within the scope of this thesis.

Overflowing users with information and support-related stimuli may reduce the capability to process further information. The result can be boredom, frustration, or excessive workload. Therefore, it is important to understand **human cognition** and the design requirements of workload-aware interfaces before augmenting workload. To identify opportune moments regarding workload assistance, we employ qualitative inquiries to derive design implications for workload-aware systems (**RQ1**). Here, we observe and conduct semi-structured interviews with tenants as well as the labor of sheltered living facilities. This specific user group is used to report the need for cognitive assistance during all facets of their daily life. Hence, it allows us to sketch and weigh a sequence of daily life tasks according to their cognitive demand, leading to design guidelines for future workload-aware systems. Upon deriving

RQ	Research Questions	Chapter
	Human Cognition	
RQ1	What are the design requirements for systems that provide opportune cognitive augmentation?	3
RQ2	What are suitable in-situ feedback modalities for cognitive support?	4
	Sensing Cognitive Workload	
RQ3	Does electroencephalography provide measures of cognitive workload for user interface evaluation?	5
RQ4	Do eye gaze metrics enable the classification of cognitive workload states?	6 & 7
	Applications	
RQ5	Does in-situ feedback provide cognitive support during a cooking task?	8
RQ6	How can complex physiological signals be visualized to be utilized by non-expert users?	9
RQ7	How can RSVP reading parameters be selected based on cognitive workload?	10

Table 1.1: Overview of the investigated research questions. The first part investigated opportune moments and suitable feedback modalities for workload-aware environments. The second part looked at physiological modalities for workload quantification. Finally, the third part combines both concepts and presents workload-aware applications.

these design guidelines, we investigate which output modalities are suitable to communicate assistance using tactile, auditory, and visual feedback (**RQ2**). This part describes *how* cognitive assistance can be provided.

Ubiquitous awareness of cognitive states requires applications that are **sensing cognitive workload** to provide relevant insights in real-time. First, we leverage an established method to estimate the level of cognitive processing using EEG measures (**RQ3**). Second, we investigate how eye activity reflects cognitive workload as a contactless alternative to EEG. We exploit smooth pursuit as specific eye movement which requires participants to lock on a target and closely follow moving objects. Furthermore, we investigate how daily math tasks influence the fluctuation of the pupil size. Both approaches provide a robust real-time classification (**RQ4**). This part investigates sensing modalities that reveal *when* cognitive assistance should be delivered.

The **application** section represents the final part of this thesis. Addressing workload-awareness as a promising dimension for future context-aware applications, we focus on applications that aim to amplify or improve the overall user experience. In a user study, we show how cognitive assistance is provided using in-situ assistance that comprises auditory and visual feedback (**RQ5**) and investigates how cognitive processes can be visualized to the individual user. Individuals gain the possibility to review the level of workload during specific tasks or certain points of time. Furthermore, current brain areas that are responsible for cognitive processing can be reflected in real-time to raise self-awareness (**RQ6**). Finally, we assess EEG as a tool for evaluating information processing parameters that influence the space and time of Rapid Serial Visual Representation (RSVP) presentations. We show how EEG can be used as an implicit workload measure from which adaptive systems and user interface designers benefit when designing interfaces that represent textual information (**RQ7**). Finally, this part is concerned with use cases *which* benefit from cognitive augmentation.

1.3 Challenges and Contribution

The extensive consumption of information is increasingly proliferating throughout society through several modalities, such as mobile or wearable devices. At the same time, these interfaces require our full attention while our time and resources are limited. With workload-aware systems that are conscious regarding the user's cognitive capabilities, content can be mapped suitable to the mental state. Thus, we investigate how technologies can help us to unload cognitive workload and enable more efficient interaction with the user. Thereby, we put the focus within this thesis on the following three challenges:

1. Users, although aware of their mental capacities, do not expect the user interface to react to the current level of workload. With individual cognitive capabilities changing throughout the day, such systems have to employ cognitive assistance *just-in-time* when needed. When succeeding, such assistance can result in higher efficiency and focus due to mental offloading. However, this process has to be seamless. The cognitive augmentation and "disaugmentation" through computing systems without making the user aware of it is a pressing challenge. Furthermore, new upcoming tasks can be scheduled depending on the currently available cognitive capacities. Two decisive parameters are commonly manipulated within the domain of adaptive user interfaces: the presentation modality, such as auditory, visual, or tactile feedback, and the respective representation of the modality such as a tactile pattern or specific auditory alert. However, how the prior mentioned parameters are adjusted depends (a) on the personal preferences of the user and (b) on the task the user is currently working on. A system that chooses sub-optimal values for these parameters may frustrate the user due to placing additional cognitive workload on it or provoke boredom through excessive assistance. Finding the sweet spot that provides an engaging user experience is a key challenge for workload-aware systems.

2. Ubiquitous technologies – which end users can acquire nowadays – become more and more context-aware. Today computing devices and their accompanying sensors are gaining computation power while being reduced in their size. As an example, devices such as fitness trackers support health goals, yet operate without regarding the cognitive state of the user. Technologies that quantify cognitive workload as an additional dimension provide the possibility to adapt the current user interface, give user interface designer insights into their design, and enable task scheduling that is mapped between task difficulty and cognitive capacity. The challenge is to exploit technologies that extract diurnal measures in real-time that predict the currently perceived workload so that user interfaces can operate accordingly.

3. Finally, little research has been conducted on manipulating the self-awareness and reflection of cognitive demand throughout the day. Stationary, mobile, and wearable technologies can sense user activities and behavior constantly and may reflect cognitive states to the user. This, combined with a recording of conducted activities

10

Human Cognition	Physiological Sensing	Applications
Context Content-Awareness Feedback Opportunities **Feedback** Modalities Perception	**Measures** Cognitive Workload **Methods** Physiological Pipeline	**Support** Cognitive Augmentation **Evaluation** System Assessment **Presentation** Reading Optimization
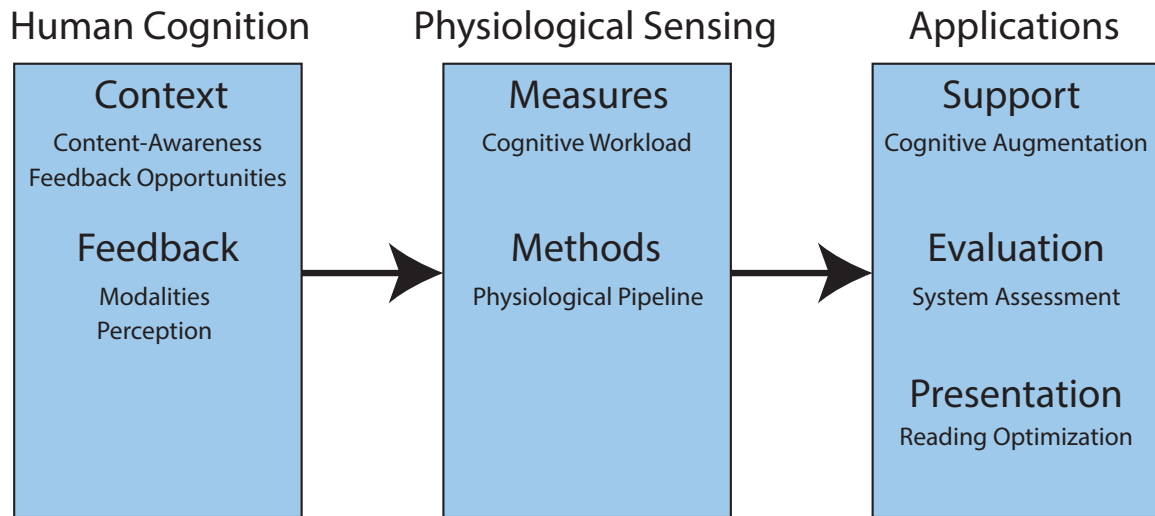
Figure 1.1: Overview of the contributions and challenges tackled during the course of this thesis. Design implications and the efficiency of feedback modalities are investigated in the *Human Cognition* section. We then apply AI-based *Physiological Sensing* to objectively measure cognitive workload. Finally, *Applications* are sketched which benefit from workload-awareness.

throughout the day, allows users to optimize their schedule to match their availability of cognitive resources to a suitable task. Moreover, recent advances in deep learning pave the way for automatic scheduling of tasks depending on the required cognitive demand. Smart assistants can assess that information to provide implicit assistance in task scheduling for cognitive assistance.

In this thesis, we tackle the aforementioned challenges by applying theories of cognitive psychology and physiological sensing on technology to provide cognitive assistance. This includes working memory theories, the utilization of ubiquitous sensing, constant availability of physiological data, and applying recent advances in AI. More specifically, we set the goal of

1. objective real-time quantification of mental workload to identify the need for cognitive assistance during the use of technology as an additional dimension within the paradigm of context-aware computing.

2. deciding and identifying opportune moments that require technologies to deliver content that is tailored to the users' remaining mental capacities.

3. visualizing cognitive demand with the intent to provide user reflection or deeper insights for user interface designers that cope with the creation of mentally gentle or demanding UIs.

This thesis focuses on end-user evaluations; in particular workload augmentation and the application layer within the domain of Human-Computer Interaction (HCI), cognitive psychology, and AI including Machine Learning (ML) as well as Deep Learning (DL). A human-centered approach is selected by conducting qualitative inquiries a well as lab and in-situ studies, and feedback for future workload-aware systems is collected and evaluated within technological and algorithmic probes. Table 1.2 provides a summary of the conducted research inquiries while Figure 1.1 illustrates the structure of the thesis. The following scientific contributions address the aforementioned goals:

Design Requirements for Workload-Aware Interfaces

The design requirements and modalities for technologies that adapt during opportune moments were evaluated in a field study. Through qualitative inquiries, design guidelines were derived to match the expectation of the users. Specifically, households that accommodate people with cognitive impairments that require cognitive assistance were targeted [233]. Supervised cooking emerged as an important communal activity where cognitive assistance is frequently required but not always available due to shortages of specially trained staff. If not mentally supported, tenants tend to make mistakes that may cause injuries or leave their current task "as it is" including the risk of unexpected consequences (*e.g.*, leaving the cooktop on). Our methodology employed observations to create an initial question catalog which was refined in an iterative process. Next, semi-structured interviews were conducted with caretakers of the sheltered living facility to establish design implications for workload-aware interfaces that support tenants during the cooking process. Regarding the preferred feedback modalities of computerized assistance, an additional study revealed significant preferences for auditory and visual feedback during opportune moments for assistance.

Using Physiological Sensing to Assess Cognitive Workload

We build upon previous theories of cognitive psychology and the field of HCI to quantify the demand of user interfaces on end-users. An objective real-time assessment of cognitive workload is important since questionnaires or think-aloud protocols, the traditional way of workload assessment, are prone to subjective perception, interfere with the experimental workflow, and require participants to remember their experience. We exploit physiological sensing as an alternative to traditional workload measures for two reasons: First, UI designer and experimenter receive in-situ feedback about their novel system while participants use it. Secondly, users benefit from self-reflection and UI adaptations that prevent boredom or frustration due to UI difficulty adjustments. In this thesis, we evaluate the concepts of eye tracking [223, 226] and EEG [221] as direct workload measurements that can be employed in real-time.

Identifying Opportune Moments for Cognitive Assistance

Based on in-situ observations, we identified how user interfaces can adapt to the current cognitive state by evaluating visual, auditory, and tactile feedback. First, we reviewed the
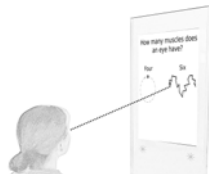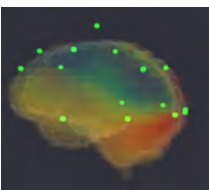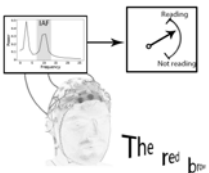
	Inquiry	Description	Chapter
Human Cognition		We investigated the design implications of cognitive assistance through qualitative inquiries in a sheltered living facility and derive five design implications for the design of workload-aware systems.	Chapter 3
		To evaluate the impact of different stimuli on cognitive workload, we performed a study with three different output modalities to communicate errors in an assembly scenario.	Chapter 4
Sensing Cognitive Workload		We present the use of EEG features to assess the workload induced by two different instruction modalities in a manual assembly scenario and present a EEG-based evaluation pipeline.	Chapter 5
		We exploit the performance of smooth pursuit eye movements as an indicator of cognitive workload. Furthermore, we explore the classification performance for potential workload-aware systems.	Chapter 6
		Changes in the pupil diameter during different cognitive workload states. We evaluate the performance of cognitive workload classification in an adaptive math task scenario.	Chapter 7
Applications		We evaluate an assistive system based on the previous Chaptersthrough a long-term study in a sheltered living facility.	Chapter 8
		We present the implementation of a brain visualization that displays dipoles from which electrical activity is generated. We show how this approach can be combined with a neurofeedback loop.	Chapter 9
		We present a study which investigates the influence of different RSVP reading parameters on information and knowledge acquisition. We derive a predictive model that utilizes EEG that forecasts the suitable reading parameters.	Chapter 10

RESEARCH INQUIRIES

Table 1.2: Research inquiries developed within the scope of this thesis. Each research implementation is described in detail in a dedicated Chapter or section within this thesis.

aforementioned modalities in a user study [228] and found a preference for visual and auditory feedback. We thus implemented a system using the preferred stimuli in an in-situ study to determine the feasibility of providing adaptive assistance suited to the current level of cognitive demand [231].

Tools and Frameworks for Building Workload-Aware Systems

Tools and frameworks are essential for the development of context-aware systems. Such tools can support the proliferation of a new dimension in the paradigm of context-aware computing. We present visualizations that support user interface designers to understand how their interfaces affected cognitive processes on a cortical level. Additionally, users can self reflect on their past performance [225]. In this context, we publish the tools and source code we have implemented. This is complemented by the collected data sets that can be used to innovate or extend existing research.

1.4 Ethics

All studies reported in this thesis were conducted in accordance with the declaration of Helsinki [12]. Furthermore, we followed the aspects of the Ethical, Legal and Social Implications (ELSI) [276]. This includes the submission of a detailed study plan to individual project partners that were involved in the related study. Within the context of studies that require proactive involvement and acquisition of participants, consent for data collection was given according to the General Data Protection Regulation (GDPR). Personal data, such as demographic data, photos, and videos, were stored securely to prevent access by third parties. No data was collected by any probes before the agreement of the involved users. The studies conducted in Chapter 3, 4, and 8 included tenants with cognitive impairments from a sheltered living facility. Thereby, the study procedure followed an additional examination process of the German Federal Ministry of Education and Research (BMBF)[1] and the Gemeinnützige Werkstätten und Wohnheime GmbH (GWW).

1.4.1 Consent

Written and informed consent was received, when possible, by the participants themselves. Participants were informed about their rights and compensation within the study. To ensure proper consent from participants with cognitive impairments, we submitted our study plan to the BMBF and the corresponding living facility GWW. Upon gaining approval from both institutions, the tenants of the sheltered living facility were informed about the course of study and were asked for their potential participation. This was achieved by asking participants directly for consent. If this was not possible due to the level of cognitive impairment, the

---

[1] `www.bmbf.de/en/index.html` - last access 2020-01-28

associated legal guardians were asked to provide written consent for the participant. Regardless of the aforementioned procedure, participants were always informed about the course of study, collected data, and their rights to abort the study or request a deletion of their data.

1.4.2 Physiological, Behavioral, and Task-Related Data

During our research, we collected a rich amount of individual physiological, behavioral, and task-related data from participants. This was required to feed the data into ML and DL algorithms that learn or derive patterns. However, participants might behave differently during an experiment to hide their regular behavior due to privacy concerns. We designed our experimental procedure transparently by showing what can be inferred from physiological, behavioral, and task-related data. Furthermore, participants were introduced into the anonymization process to provide evidence about the intractability of their identity. When employing systems in the wild, a "*Privacy by Design*" [236] paradigm was employed to ensure privacy in protected places.

1.4.3 User-Awareness

When technology tracks mental states, as well as behavioral data in private settings, the challenge of workload-aware interfaces, is to provide seamless interaction while communicating with the user how certain decisions were made. Users may be unaware of certain triggers that cause an unexpected system behavior based on their cognitive state or behavior.

1.5 Research Context

This thesis presents work that was carried out in the context of different scientific institutes and research partners for four years. The presented research was carried out over two years in the Human-Computer Interaction and Cognitive Systems group at the Institute for Visualization and Interactive Systems in Stuttgart as well as in the Human-Centered Ubiquitous Media group at the Ludwig Maximilian University of Munich. Both groups are supervised by Prof. Dr. Albrecht Schmidt. Several publications resulted in a variety of collaborations.

1.5.1 Project *KoBeLU*

The major part of this thesis was conducted in the context of the project Context-Aware Learning Environment (KoBeLU) (Grant No. 16SV7599K) funded by the BMBF within the project call "Erfahrbares Lernen". Including the Ludwig Maximilian University of Munich, eight project partners (AUDI AG, MAHLE GmbH, International Center for Ethics in Research (IZEW), GWW. Offenburg University of Applied Sciences (HSO), Stiefel Eurocart

GmbH, User Interface Design UID) took part in the project. KoBeLU aimed to develop and evaluate assistive systems within the learning contexts. Specifically, the influence of in-situ assistance was subject to studies for different stakeholders in various learning environments. By integrating assistive technologies, this 3-year project (Aug. 2016 - Nov. 2019) researched the degree of cognitive assistance that is provided in real-time. The collaboration between research and application resulted in conjoint publications [135, 218, 224, 228, 231, 233, 281].

1.5.2 Project *be-greifen*

Within the project call "Erfahrbares Lernen", collaborations were done with the project "be-greifen" (Grant No. 16SV7527). This is reflected in several publications with the research associate Pascal Knierim [204, 205, 206, 207, 208, 209, 222, 229]. The work there was related to cognitive assistance using Head-Mounted Displays (HMDs).

1.5.3 Project *AMPLIFY*

Several projects were supported by the European Union's Horizon 2020 Programme under the ERCEA AMPLIFY (Grant No. 683008). This is reflected in publications with the research associates Jakob Karolus [189, 190, 246] and Matthias Hoppe [172, 173, 174, 175, 392]

1.5.4 University Collaborations

Starting in 2017, the workshop series Designing Assistive Environments for Manufacturing (DAEM) was launched at the conference PETRA in cooperation with the Ostwestfalen-Lippe University of Applied Sciences. The workshop is centered around the development and evaluation of assistive technologies in manufacturing environments. This includes the evaluation of technologies that support and foster the cognitive abilities of workers, resulting in a design space for industrial assistive computing [59]. To date, the workshop was renewed for a fourth iteration in the year 2020.

Furthermore, this thesis presents work concerning people with cognitive impairments. Since this represents a marginalized group, traditional qualitative methods are difficult to apply. To overcome this, we consulted the expert Erin Brady for this research project as she provides the necessary research qualifications. The conducted project is concerned about the behavior of persons with cognitive impairments during communal activities, particularly cooking. This enabled us to derive design implications for assistive kitchen technologies that provide mental assistance for people with cognitive impairments [233] and point out the potential abuse of assistive technologies in working environments [232].

1.6 Thesis Outline

This thesis comprises **eleven chapters** and is divided into **six parts**. The structure of this thesis closely follows the depiction of Table 1.2. The first part denotes background knowledge and differentiates existing work from the novel contributions presented in this thesis. The second part provides a series of studies that investigate the design requirements and implications of workload-aware systems. The third part introduces studies about physiological sensing modalities that enable in-situ measures of workload states. The fourth part presents how the aforementioned studies provide cognitive assistance and enable users to reflect on workload states. Finally, the fifth part concludes with guidelines and provides an outlook for workload-aware systems in the domain of computational interaction. The sixth and last part includes the bibliography and listings.

1.6.1 Part I: Introduction and Foundations

Chapter 1 – Introduction

The first Chapter describes the motivation and vision for workload-aware interfaces, presents use cases, states the context in which this thesis was conducted, states collaborations, and summarizes challenges as well as opportunities which were faced throughout our research.

Chapter 2 – Background and Foundations

The second Chapter introduces key concepts of human cognition and the concept of cognitive workload. We present state-of-the-art measures that facilitate explicit and implicit measures for cognitive workload. This is followed by the presentation of recent developments of ubiquitous assistive technologies that support workers and persons with cognitive impairments. Ethical considerations are discussed since sensor-dependent systems need to be viewed critically from a data collection perspective.

1.6.2 Part II: Human Cognition

Chapter 3 – Opportune Cognitive Augmentation

Detecting suitable situations for cognitive augmentation is crucial to avoid context-aware mismatches. Therefore, we first focus on the disclosure of opportune moments for cognitive assistance. We assess the characteristics of diurnal chores that benefit from cognitive support in a case study with persons with cognitive impairments.

Chapter 4 – Feedback Modalities

In this chapter, we evaluate the efficiency of feedback modalities after detecting situations that require assistance. We conducted a lab study that compared different visual, auditory, and tactile feedback patterns.

1.6.3 Part III: Sensing Cognitive Workload

Chapter 5 – Mobile Electroencephalography

Cortical activity provides markers for cognitive demand and exhaustion in real-time. In a lab study, we evaluated the cognitive demand of two different user interfaces by analyzing EEG data. We found that EEG is a reliable complementary measure for cognitive workload and proposes an evaluation pipeline that supports researchers and designers in their assessment of user interfaces regarding the placed mental demand. Machine learning techniques are applied to show potential real-time adaptations of future workload-aware visualizations.

Chapter 6 – Smooth Pursuits

Human eye gaze behavior is affected under different levels of cognitive workload. We investigate changes in smooth pursuit eye movements under different levels of cognitive workload. Applying machine learning techniques reveals a reliable distinction for different cognitive workload levels for interpersonal and intrapersonal classification.

Chapter 7 – Pupil Dilation

The pupil is an eye component that is heavily affected by changes in cognitive demand. We show how differences in pupil dilation can be utilized to build reliable on-the-fly workload-aware systems, and how eye-tracking data can be used for the physiological prototyping of workload-aware systems.

1.6.4 Part IV: Applications

Chapter 8 – Design Pipeline Evaluation

We evaluate the results and the design pipeline from Chapter 5 in a case study at a sheltered living facility. Using an exemplary assistive system that incorporates the aforementioned results, tenants with cognitive impairments reacted favorably to the provided cognitive augmentation and were unusually independent throughout the task.

Chapter 9 – Reflection

Communicating changes in cognitive demand help to understand how diurnal tasks can be mapped to available resources that vary throughout the day. We propose the implementation of a real-time visualization which we evaluate in the context of a neurofeedback loop.

Chapter 10 – Workload-Aware Reading Interfaces

This chapter describes the concept of an application scenario for workload-aware adaptive reading scenarios. By evaluating EEG, we propose strategies to adjust RSVP reading parameters depending on the current level of perceived workload. We propose predictive models that adjust the reading speed regarding the cortical activity.

1.6.5   Part V: Conclusion

Chapter 11 – Conclusion and Future Work

In this chapter, we summarize and conclude the findings of this thesis and revisit the research questions stated in the beginning. We reflect on the presented research approach, discuss future work, and present implications for workload-aware systems.

Chapter 2

Background and Foundations

My problem is that I have been persecuted by an integer. For seven years this number has followed me around, has intruded in my most private data, and has assaulted me from the pages of our most public journals.

George A. Miller

The research presented in this work is based on the field of ubiquitous computing, assistive technologies, and physiological sensing. To create and design workload-aware interfaces, we apply theoretical concepts from psychology, cognitive load theory, and current practices of employing sensing modalities in assistive technologies. This includes ethical considerations that arise with interfaces that can sense physiological states ubiquitously. Here, we provide readers with the necessary knowledge that deepens the foundations and selected methods for the studies presented in the upcoming chapters.

2.1 Ubiquitous Computing

The term *Ubiquitous Computing* describes the era in which computers started to "weave themselves into the fabric of everyday life until they are indistinguishable from it" [381]. This included a deviation from the many-to-one relationship, where many persons relied on the use of one computer. With the introduction of the *MITS Altair 8800* personal computer in the year 1975, computers became available for broad masses of people. In the following years, computers became smaller and their capabilities in sensing capabilities and computational power increased. Nowadays, the distribution of computers has changed into a many-to-one relationship, where many computational units are owned or operated by a

single person. These can sometimes be not distinguishable as they are seamlessly integrated into the environment: adaptive light switches, slide doors, or voice assistants are just a few examples of technologies that facilitate our life [382].

These technologies accomplish tasks for us which would usually drain our cognitive attention. They provide mental alleviation and allow us to schedule cognitive resources to other tasks that require human action. However, in emergency cases, ubiquitous devices might compete for our attention to take action, so we should be made aware of our cognitive states to provide selective assistance. In the following, we summarize previous research on cognitive workload theory.

2.1.1 Perception of Workload

Workload, regardless of mental or physical nature, is an important component throughout the perception of diurnal activities. The body-mind interaction is believed to be affected when strained or relaxed over a certain time span [265]. Hence, the workload is ubiquitously present and often associated with the accomplishment of daily tasks. Reducing workload to support users in daily tasks is an established way to increase a person's overall life quality. For example, robots and machines reduce physical workload by overtaking tasks that are dangerous or exhaustive for humans.

Cognitive workload is one type of workload that exhausts mental resources and that is attributed to individual properties. Past research aimed to reduce cognitive workload that arises through everyday tasks, such as digital calendars, task scheduling apps, or notification management systems [344]. Providing too much cognitive support might lead to boredom or loss of important information that comes along with a context-aware mismatch. No presence of cognitive support might lead to frustration due to failing in accomplishing the present task. Both aspects are detrimental for the user in terms of performance, user experience, and task performance which might elicit dangerous situations [139]. Finding the "sweet spot" for mental demand is necessary to keep the user in flow states, where flow states are an essential part of maintaining the users' engagement and task efficiency [84, 85]. We elaborate on the term cognitive workload and the relevant semantics for this thesis in the following.

2.2 Cognitive Workload

We process information by perceiving events in our environments. In this case, cognitive workload describes any information processing load placed on cognitive mechanisms. It is often associated with the level of measurable effort that is mobilized to accomplish a task. Even more often, cognitive workload is falsely used interchangeably with the terms *stress*, *fatigue*, or *physical workload* [158]. In the following section, we provide an overview of the semantics of the term *cognitive workload* and its theories which are deemed relevant for the scope of this thesis.

2.2.1 Working Memory

As humans perceive information in their environment, they are obliged to store relevant "*chunks of information*" in their short-term memory for further processing. This temporary storage, known as working memory [14], utilizes the five senses sight, touch, taste, smell, and hearing, to encode information. It is assumed that cognitively demanding tasks can only be efficiently accomplished with a sufficient amount of working memory. Thereby, the number of objects that can be stored simultaneously is limited to *seven* plus or minus two chunks [83]. Without sufficient capabilities to store information temporarily, cognitively demanding tasks might be completed with less efficiency. How large the capacity of working memory is and how it can be quantified is highly disputed in the field of Psychology. In common parlance, it is said that "seven chunks plus/minus two" can be memorized by an individual. Miller argues how this number might have emerged as the status quo in research [270]. Further research has been conducted in this regard, indicating that the number seven is the limiting factor that should be considered as critical when relying on temporary storage [325]. Thereby, large inconsistencies in a set of observed events provide a better memory performance.

2.2.2 Cognitive Workload Theory

Previous work has inquired into different theories on how cognitive workload and mental models work. The term "cognitive workload theory" has been designed to optimize knowledge intake and boost learning performance during information processing. It is assumed that working memory is limited in dealing with the perception through our five senses. Thereby the processing of visual, auditory, or tactile information is limited by the number of "chunks" an individual can remember and mentally process at the same time. The overall aim of cognitive workload theories is to provide guidelines that minimize the demands upon working memory [355]. In the context of HCI, designers could reduce the number of visual chunks a user has to remember (*e.g.*, menu items), or utilize the Gestalt laws [162] to group similar chunks as means to reduce the effort for item searching. John Sweller [353] presented a cognitive workload theory which divides workload, learning, and problem-solving into three components: *Intrinsic Workload*, *Germane Workload*, and *Extraneous Workload*, that describe factors that contribute to the demand for working memory.

Intrinsic Workload describes the inherent complexity of a task. *Intrinsic Workload* is not trivial to manipulate since the intrinsic task complexity is associated with the necessary individual cognitive demand to process and act upon specific information. If possible, *Intrinsic Workload* should be kept at a minimum to avoid unnecessary internal processing of users. However, this is considered difficult because the task itself needs to be modified to manipulate the complexity [398].

Germane Workload represents the effort to process patterns within a task. Realizing schemes that help to solve a task may increase task engagement and foster learning. Hence, maxi-

mizing the share of *Germane Workload* is emphasized as a crucial factor when designing engaging user interfaces.

Extraneous Workload is manipulated by the task representation that can be perceived with human senses. For instance, a well-designed visualization enhances the interpretation of data far better than bad design. Relative to the other workload types, *Extraneous Workload* is trivial to manipulate since the task representation can be exchanged through the appearance of the problem (*e.g.*, visually or auditory). *Extraneous Workload* should be minimized to avoid unnecessary allocations of cognitive resources required to resolve the intrinsic task complexity. Figure 2.1 illustrates the components of Sweller's cognitive workload theory.
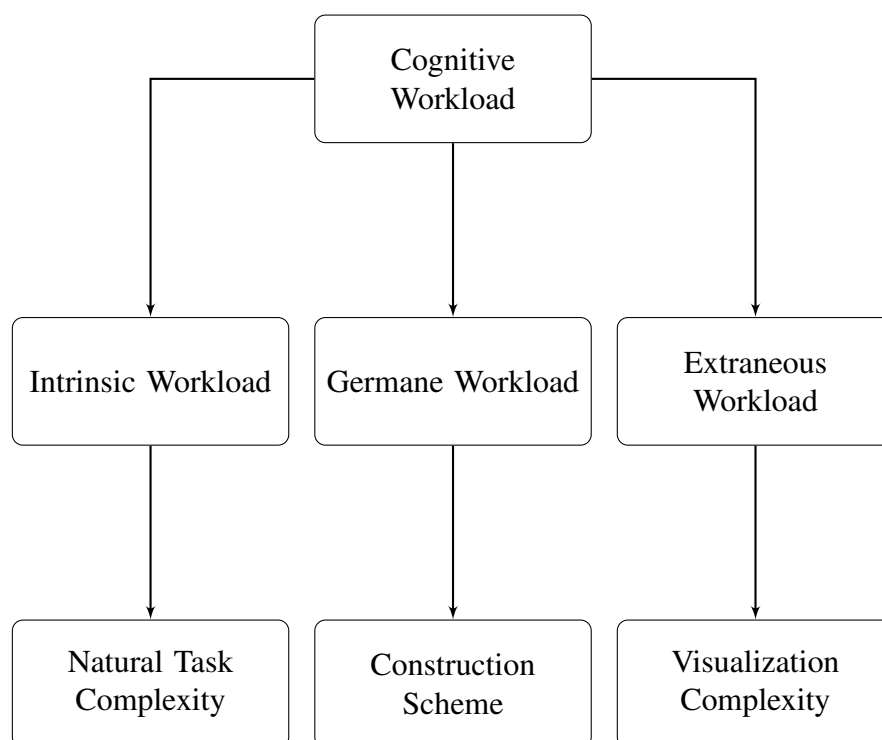
Figure 2.1: Graphical representation of the cognitive workload theory by Sweller [353]. *Intrinsic*, *germane*, and *extraneous* workload represent form the overall experienced mental demand. While *intrinsic* and *germane* workload cannot be easily changed due to the inherent task complexity and tools provided to the user, *extraneous* workload can be manipulated by the task representation.

2.2.3 Cognitive Workload Assessments

The assessment of the cognitive workload that is placed by user interfaces is a factor that has been viewed as crucial by designers. Different measures, ranging from direct subjective assessments to implicit physiological measures, have been employed to sense cognitive workload. Although being far from exhaustive, the following section informs about current
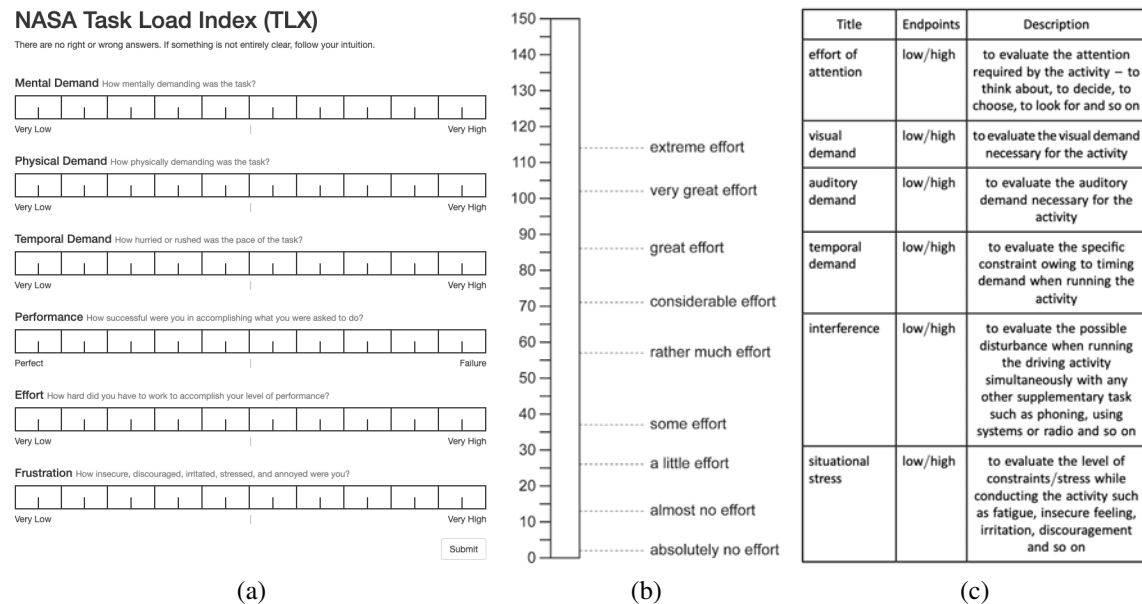
Figure 2.2: Example of questionnaires which measure cognitive workload. **(a):** NASA Task Load Index, **(b):** Rating Scale of Mental Effort, and **(c):** Driver Activity Load Index.

practices of workload measures that have gained popularity in HCI research. These focus primarily on self-rated measures, brain-sensing, eye tracking, and physiological measurement modalities.

Self-Rated Measures

Asking users in-situ or post hoc about their cognitive strains during interaction is a common way to evaluate user interfaces. Subjective inquiries benefit from being an uncomplicated measurement that can be used without the need for complicated hardware or foreknowledge. Popular questionnaires which are used among researchers are, for example, the NASA Task Load Index (NASA-TLX) [159, 161], Rating Scale of Mental Effort (RSME) [385, 401], or Driver Activity Load Index (DALI) [296] (see Figure 2.2). Questionnaires reign among the most used measures for cognitive workload since they are easy to employ and be applied during or after experiments. However, they face critical trade-offs: They are susceptible to the subjective perception of participants, they place additional cognitive workload on the participants by requiring them to remember their experience, and they lack concrete real-time assessments.

Electroencephalography

The human brain is the information and control unit of a human being. Thereby, approximately 100 billion neurons underlie human cognition [165]. Electrical activity in several brain regions can be measured by placing electrodes on the scalp. Neurons communicate by

| (a) | (b) |
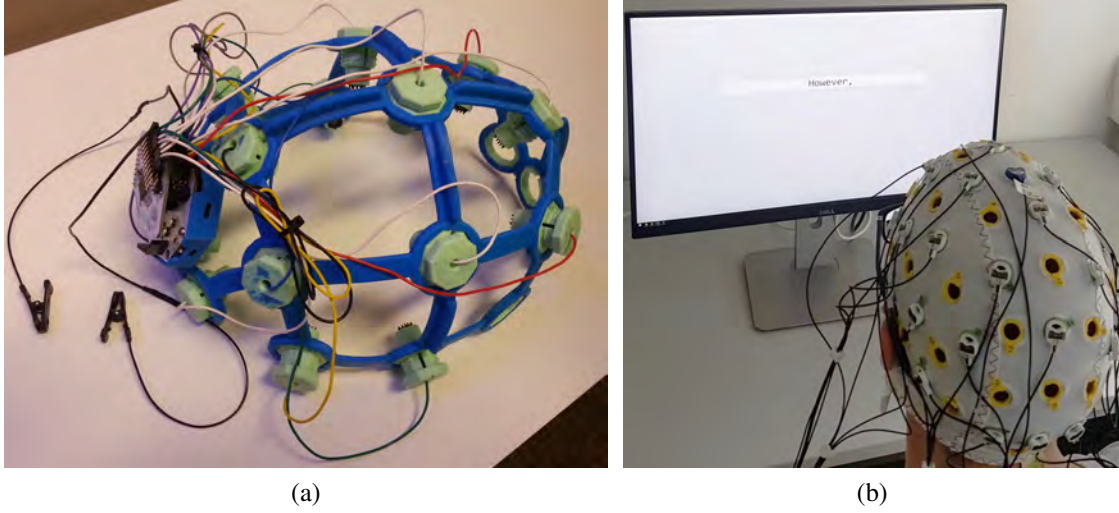
Figure 2.3: Different EEG setups. **(a):** Affordable off-the-shelf solutions can be achieved using the OpenBCI with a customized 3D printed headset. **(b):** More expensive devices enable integrated impedance checks and a dense electrode placement with up to 64 electrodes.

exchanging electrical activity using neurotransmitters, a chemical transferred between neurons. In this work, we focus on measuring this electrical activity via an EEG headset. EEG is commonly leveraged in clinical application and yields a non-invasive method to estimate cortical activity [261, 278, 390]. Electrical potentials between $1\,\mu v$ and $100\,\mu v$ (microvolts) are measured by placing conductive electrodes on a scalp. An additional electrode serves as a reference electrode, which can be placed on the earlobe or scalp [257]. The measured EEG potentials allow the processing and extraction of different features. Machine learning has been used on these features to retrieve insights into cognitive processes [251, 252]. This approach allows to discriminate between cognitive states [124, 152, 238] and enables EEG as a modality for user experience evaluation [123]. For example, changes in electrical potentials are observed by analyzing frequency bands. Previous work found a drop in frequencies of alpha (8 - 12 Hz) and an increase of theta (4 - 8 Hz) [199, 221] when subjects have to raise mental capacities. An alternative approach is the assessment of Event-Related Potentials (ERPs) to infer mental workload [46, 320]. To complement this, real-time brain visualizations enable deeper insights into sequences of neuronal activity [225].

However, EEG measurements are prone to noise. Head movements, muscle contractions, eye movements, and even eye blinks cause changes in the electrical field on the scalp. Researchers are concerned about this and invest great effort to reduce the number of measurement artifacts [86, 125, 216, 279]. Since artifacts cannot be completely avoided, the use of EEG often requires a controlled environment comprising minimal body movements of the user. Such conditions are often impractical for end-users due to their high experimental control. However, recent technical advances ameliorate these disadvantages [274] that allow artifact corrections for EEG during mobile settings [36]. The barrier EEG in real-world set-

tings has been lowered by making EEG headsets accessible to the consumer market. These are usually priced between \$249[1] and \$1600[2]. More affordable open source solutions can be acquired within the OpenEEG Project[3] (see Figure 2.3).

Frequencies

Different oscillations in EEG signals are attributed to several semantics of cognitive states. The six bandwidths delta (1 Hertz (Hz) - 3 Hz), theta (4 Hz - 7 Hz), alpha (8 Hz - 12 Hz), lower beta (13 Hz - 20 Hz), upper beta (21 Hz - 30 Hz), gamma (31 Hz - 100 Hz) power are usually investigated when analyzing frequencies. Changes in alpha and theta frequencies are correlated with the mental demand placed on working memory as well as increase or decrease in task engagement [146]. The alpha power is associated with changes in brain resting states. If the brain is resting, a modulation of alpha waves is achieved which decreases when users perform tasks that require their memory. Theta oscillations correlate with changes in task engagement at the frontal anterior cingulate cortex [55]. Thereby, an increase in theta power is associated with higher task engagement while lower theta power is an indicator of low task engagement. Since it is unknown if correlations between the demand of working memory (*i.e.*, alpha power) and task engagement (*i.e.*, theta power) might elicit flow states, researchers use the theta-alpha ratio as a coherent metric for cognitive demand.

Event-Related Potentials

ERPs are specific amplitudes that characteristically appear after users perceive a stimulus. Different types of amplitudes can be measured depending on the type of provided stimulus. For example, if a stimulus does not match the expectation of a user, a negative peak after approximately $400\,ms$ is measured in the EEG signal. Another example is the processing of stimuli that require the users' attention which typically resemble a positive peak after $300\,ms$ [47]. Smaller ERP amplitudes are expected when users pay less attention to a specific stimulus. Tasks that require large amounts of working memory are more difficult to process, thus resulting in smaller ERP amplitudes. Hence, ERPs can be measured to assess the mental vigilance towards auditory cues [147] or to detect vocabulary gaps [339, 340].

Eye Tracking

Eye tracking is a method to follow the gaze of a person. The gaze trail, a sequence of gazed objects, or the pupil diameter can be measured without great effort. Eye tracking has become an interesting topic for the HCI community since it provided the possibility to computationally link attention and objects of interest with visual perception. The *Mind-Eye Hypothesis* states that gazed objects are an indicator of visual attention [183]. This is not necessarily true: objects can be gazed without paying visual attention to them. However, eye

---

[1] `www.choosemuse.com` - last access 2020-01-28

[2] `www.futurehealth.org/bm_at1.htm` - last access 2020-01-28

[3] `www.openeeg.sourceforge.net` - last access 2020-01-28

(a)   (b)

Figure 2.4: (a): Stationary eye tracker which emits infrared light. This light reflects on the cornea which is captured by the device to estimate the gaze points. **(b):** Head-worn mobile eye tracker. An infrared camera captures the reflection of infrared light and maps the gaze to a world camera.

tracking represents an interesting tool to learn about human attention, focus, cognition, and potential modality for advanced user inputs [101].

Eye trackers are available in stationary and mobile settings. Stationary eye trackers are traditionally mounted below a display and emit infrared light which is reflected on the eye (see Figure 2.4a). This reflection, known as corneal reflection, is captured by the eye tracker to estimate the gaze point on the screen [308]. Additional variables that can be captured include the pupil diameter, gaze point differences between the left or the right eye, and environmental lighting conditions. In contrast, mobile eye trackers resemble goggle-like devices that are worn by users (see Figure 2.4b). Instead of putting the eye tracker below a monitor, the infrared emitter can be worn like glasses. Similar to stationary eye tracking, the corneal reflection is captured and the gaze is mapped to the currently perceived view. Mobile eye tracking is not limited to a physical screen and can be used to track the eye gaze throughout mobile scenarios, where the eye gaze is mapped on a recorded video feed [51]. Alternative approaches to mobile eye-tracking include the evaluation of Electrooculography (EOG). EOG describes electrical activity which is created by muscle contractions through eye movements. EOG goggles employ electrodes that have contact with the skin and serve as a measurement unit for eye-based activity [52]. The current eye gaze and movement directions can be evaluated by analyzing the emitted electrical activity. The incorporation of mobile eye-tracking has recently taken place in AR [369] or Virtual Reality (VR) [106] applications and is likely to become a popular usability tool.

We elaborate on well-known eye movements and metrics in the following. Although being far from exhaustive, we focus on metrics that are relevant for the understanding of this thesis, where a comprehensive review is provided by Duchowski [103]. **Fixations** are pauses over regions which attract the attention of users [326]. Thereby, the visual gaze remains over the

region for the duration of the fixation. **Saccades** describe rapid eye movements between fixations [326]. Saccades are very fast and are considered to be among the fastest movements a human body can produce. **Smooth Pursuits** are smooth eye movements that occur when closely following a moving target on a screen [374]. Calibrating an eye tracker, which is a challenging procedure that disrupts the user experience, is not necessary to detect smooth pursuits. On the contrary, smooth pursuits can be used to calibrate eye trackers [194]. Finally, the **Pupil Diameter** can be measured as an indicator for cognitive workload [304, 149]. However, pupil dilation is easily affected by external noise factors such as changing lighting conditions. Therefore, careful interpretations have to be derived when using pupil dilation as a measure for cognitive states.

Other Physiological Measures for Cognitive Workload

Alternative measures to sense cognitive workload have emerged besides the use of questionnaires EEG and eye tracking. **Functional Near-Infrared Spectroscopy (fNIRS)** is a non-invasive modality that uses near-infrared range light between $650\,nm$ and $1000\,nm$ to measure concentration changes in Oxygenated Hemoglobin (HbO) and Deoxygenated Hemoglobin (HbR) in the brain [375]. The concentration levels can be estimated by placing light emitters on the scalp that measure the returning light from an infrared round trip [170]. This method is popular for Brain-Computer Interface (BCI) applications since fNIRS is easy to set up and process [277]. However, fNIRS is known to have significant latencies since it takes between five to ten seconds until changes in oxygen are visible in the bloodstream [255]. Another non-invasive measure for brain activity is **Functional Magnetic Resonance Imaging (fMRI)**. fMRI uses magnetic resonance imaging to analyze blood flow levels [110]. In contrast to the presented non-invasive methods, invasive methods require electrodes to be placed below the scalp. Methods, such as **Electrocorticography (ECoG)**, require electrodes to be placed below the skull to capture electrical signals [243]. **Heart Rate (HR)** measures the heartbeat where changes in HR and Heart Rate Variability (HRV) are frequently used as an assessment for stress and cognitive workload [313]. **Electromyography (EMG)** measures electrical activity emitted from muscle contractions by body movements. Changes in muscular micro-movements are an indicator of cognitive workload [69]. **Electrodermal Activity (EDA)** describes the activity of sweat glands as an indicator of stress and workload [41, 227]. Cognitive workload can have an impact on **haptic interaction**, such as touch or pen input. Changes in force or interaction patterns can be measured when users experience cognitive workload [394]. Cognitive workload is also known to influence **Speech** regarding lexical density [195], tempo [71], pauses [196], or voice pitch [378]. **Body Temperature** changes have been correlated with cognitive load [331]. Finally, direct measures of cortisol levels have been used to measure cognitive workload [74]. However, cortisol levels require laboratory analysis and are not suitable for real-time analysis.

2.3 Assistive Technologies and Systems

Assistive technologies started as devices that aided persons with disabilities, such as voice recognition systems, braille, hearing aids, wheelchairs, or screen readers. This basic idea of using technologies to enhance human capabilities has attained popularity for the general population. Such assistive systems have become popular tools in home and work settings. Speakers, phones, and TVs have been enriched by technologies that enable them to communicate with users and provide information just-in-time. Assistive systems evolved from the need for explicit interaction (*e.g.*, mobile phones) to devices which require subtle interaction (*e.g.*, speech recognition). Cognitive assistance through in-situ feedback enables the ubiquitous availability of information that has to be memorized for a short amount of time. In particular, assistive systems have proliferated for two stakeholders: people working at production lines with frequently changing assembly batches, and therefore constant changes in assembly instructions, have cognitively benefited from visual in-situ assistance [129]. Furthermore, persons with cognitive impairments have shown positive effects by the use of assistive systems in home and work environments [136]. We summarize relevant research about the role of assistive systems for both user groups, focusing on lowering cognitive effort, fostering collaboration among peers, and providing assistance that is necessary to accomplish daily tasks. Thereby, we use the term *assistive technology* and *assistive system* synonymously to describe a technical artifact that supports individuals physically and mentally regardless of the presence of a cognitive impairment.

2.3.1 Assistive Systems at Workplaces

Small batch sizes are pivotal to the proliferation of assistive technologies at production lines due to the individual demand for goods. Instead of producing the same part over and over again, workers at production lines have to memorize variations of assembly instructions. The constant memorization of assembly instructions strains the working memory [127, 221] which eventually results in a larger number of errors and a decrease in manufacturing quality. Especially new workers that have to learn novel variations of assembly instructions are confronted with a steep learning curve. This has detrimental consequences for the assembly process: production times, error rates, and cognitive workload temporarily increase as new assembly instructions have to be memorized.

Several assistive systems exist to reduce assembly errors and task completion times by keeping the cognitive effort low. McCalla et al. [264] presented PHelpS, a system that can convey tasks to other colleagues who are willing to help. PHelpS consists of a repository that algorithmically matches colleagues that are suitable and eager to provide support. Klinker et al. [108] incorporated AR into a welding gun, using projection-based and HMD-based AR to support workers during their welding tasks. Another approach for assistive computing was presented by Gulevich et al. [155]. Their project *TeleAdvisor* uses a camera-projector-system to provide learners with real-time instructions in a TV setup scenario. Gauglitz et

al. [141] presented the use of a hand-held device for mobile AR assistance. AR instructions can be annotated to operate a Boeing 737. Funk et al. [133] presented motionEAP, a camera-projection system to assist with in-situ projections. Several studies found that in-situ projections provide a higher assembly efficiency and cognitive alleviation when projections are adapted to the current assembly performance [129, 134]. However, how to adapt to the individual perception and skill set remains an open question.

2.3.2 Assistance for Persons with Cognitive Impairments

Past work has recognized the potential of mental augmentation for persons with cognitive impairments. Today, approximately 20% of the world's population lives with some level of cognitive impairment [290]. This trend is expected to increase with an aging society. Sheltered work and living organizations, as well as researchers, have recognized this and begun to evaluate the efficiency of assistive systems.

Levine et al. [244] conducted an early review of how persons with cognitive impairments perceive several assistive systems. They present a model for the use of assistive technologies for people with different disabilities. Comprehensive literature reviews which present past work about the usability of assistive systems for persons with cognitive impairments have been presented in the past [338], showing that mental support through assistive technologies is considered as a cognitive prosthesis for persons with cognitive impairments. Funk and Schmidt [127] exploited the potential of several assistive systems for cognitive assistance at workplaces. A series of studies revealed that assistive systems contributed to lower task completion times, fewer errors, and enhanced engagement during an assembly at production lines [136, 137, 215].

Different types of cognitive impairments raise individual requirements for assistive technologies. For example, dementia-related impairments such as Alzheimer's or Parkinson's disease require prospective aids that amplify memory [31]. Using such memory aids helps affected persons to concentrate on the current task or remember important actions that lead to their accomplishment. This has to be implemented carefully: novel technologies are less frequently adopted by individuals with cognitive impairments and are more likely to be abandoned when direct advantages or rewards are not immediately visible [212, 337]. Assistive systems are often considered as complicated and likely to cause difficulties in the process of their adoption [90]. Hence, assistive systems should be designed in a user-centered design process that involves the relevant stakeholders accordingly. The type of cognitive impairments that were present in our studies were trisomy 21, autism, dementia, and general impairments that were legally defined as cognitive disabilities. Such impairments can, for example, influence how information is processed and retained, hence hindering independent living.

2.3.3 Ethical Considerations

Ethical issues mays arise with the collection and use of personal data through the mundane attunement of users, as the use of assistive systems is considered an authorization to implicitly use personal data. With this, ethics regarding participatory HCI research that includes persons with cognitive impairments has emerged as a relevant aspect [167]. Recent research in philosophy, as well as ethics, discussed the consequences of systems that can ubiquitously collect data and adjust their services [379, 380]. Research ethics follow the principles of "autonomy", "justice", "benevolence", and "non-maleficence", which are cornerstones of ethical theorizing [24]. We highlight the relevant ethical considerations *empowerment*, *autonomy*, *informed consent*, *privacy*, and *social interaction* in the following that have been considered in the previous research.

Empowerment is considered as an intuitive aspect for assistive computing: through assistive systems, people are physically and cognitively supported to live autonomously and accomplish tasks with less or without cognitive effort. This is achieved by technology compensating for certain forms of physical or cognitive demand. For instance, an assistive system could function as an externalized memory for a person with a memory weakness. The assistive system is assumed to work as cognitive support. But from a philosophical point of view, this raises interesting questions about the boundaries of the mind and also the ethical consideration of whether assistive systems could be paternalizing their users by nudging them into certain behavioral patterns [75].

There are further reasons to be hesitant about subscribing to *empowerment*. For instance, the underlying concept of cognitive disabilities presupposes a biomedical understanding of support, according to which a person has a situational or permanent impairment that is physically or cognitively compensated by assistive technologies [39, 157]. For persons with cognitive impairments, such a model tends to overlook the role of social environments in shaping or perpetuating disabilities, as well as the different forms of discrimination persons with cognitive impairments are confronted with. That notwithstanding, while there might be clear limits to how much empowerment might be achieved through assistive systems, its benefits for helping people with cognitive impairments in leading better lives cannot be denied.

Autonomy describes the capacity of a person to exhibit self-governance [57]. What that means is that if a person acts, the acting is motivated by personal reasons. An autonomous action might be contrasted with cases where a person is either coerced or manipulated to act in a certain way. The necessary conditions for being autonomous are much debated in the relevant philosophical literature [9, 56], nevertheless it is fair to assume that, to be autonomous, one must possess a certain set of psychological capacities such as the ability to self-reflect. This includes that a person's beliefs and intentions must be coherent and stable over time. In addition to that, an agent must possess higher-order evaluative judgments on what matters to them [9, 120]. For instance, if someone is persuaded by a fitness app to go for a run, the relevant action being autonomous hinges on whether running is an activity that the person value. In the context of assistive systems, *autonomy* as an ethical criterion is

supposed to ensure that the person and not the technology exhibits control [56]. There are some challenges for using *autonomy* as an ethical criterion within the context of people with cognitive impairments [89]. The conditions for *autonomy* are mentally and physically highly demanding, increasing the importance of *autonomy* as an ethical principle. For example, even if a person does not meet the requirements for full-fledged *autonomy*, they might still have certain beliefs, desires, and values that express what matters to them. Therefore, one ethical criterion for the development of assistive systems or technologies has been that these do not interfere with the basic values and desires of their users.

The problem of cognitive demand also applies to *informed consent* and *privacy* [115, 269]. Assistive systems can collect fine-grained data on a person and, therefore, might infringe upon their *privacy*. Whereas strict guidelines have been established, governing the collection and storage of personal data, it is still a problem to sensitize users to issues related to informational *privacy* [280]. One factor is the high degree of abstraction regarding *privacy* issues, where many users find it difficult to grasp the content of *informed consent* and terms of service documents or to appreciate their relevance [44, 368]. While many decisions concerning welfare are deferred to caretakers, we still consider it as crucial that a person with cognitive impairments has authority over their interests [319]. Thus, we are currently experimenting with developing easy to understand, *informed consent* documents in regards to *privacy* issues that make use of simple language and are accompanied by informational videos. Admittedly, this is still an area where a lot of empirical testing is required [19, 370].

We need to consider how assistive systems affect the *social aspects* of the lives of persons with cognitive impairments. While assistive systems have the potential to improve the life quality for persons with cognitive impairments, they should foster social and communal activities [37, 77, 93], which eventually results in better care work through freeing other resources of caretakers. Consequently, it is one of the critical ethical aims to ensure that not merely the technology, but also the social settings of the care- and tenant-relationship are designed in a way which promotes the welfare of users.

II

HUMAN COGNITION

Chapter 3

Opportune Cognitive Augmentation

> *Today we serve technology. We need to reverse the machine-centered point of view and turn it into a person-centered point of view: Technology should serve us.*
>
> Donald A. Norman

Cognitive workload, although different from the individual constitution and individual amount of required cognitive resources, is ubiquitously present when persons work on chores. The relationship between exerted cognitive resources and the detection of the need for cognitive assistance is an important factor that has to be considered when designing workload-aware interfaces. This poses the following question, assuming that effective workload measures are integrated into the environment: *when* is the right time to intervene with cognitive assistance?

The vision of workload-aware user interfaces requires careful investigations of design requirements and implications for the respective user group. Persons with cognitive impairments, who are living in sheltered living facilities, are a user group that requires cognitive assistance to achieve basic life skills. Such cognitive impairments include trisomy 21, autism, dementia, and general impairments that are legally defined as cognitive disabilities. As persons with cognitive impairments receive "*handcrafted*" ubiquitous assistance (*e.g.* through family members or trained labor) during daily chores, their barrier to express the need for cognitive assistance is lowered. This supports the social intention to support persons with cognitive impairments by researching workload-aware assistive technologies since the need for cognitive assistance is expressed early to family members or caretakers. This chapter focuses on persons with cognitive impairments to explicitly detect when manual mental augmentation through a supervising person is desired: these cues are potential entry points for workload-aware systems. Cognitive impairments can impact someone's ability to complete

traditional activities of daily living, such as cooking, bathing, or acquiring groceries [299]. Numerous government and volunteer-driven organizations exist to provide specialized training to people with cognitive impairments as they learn these independent living skills. As average lifespans increase, the number of people affected temporarily or permanently by cognitive impairment will continue to rise [94], and this training will be in higher demand. The world report on disability [290] approximately assesses 20% of the human population a cognitive impairment where the numbers are expected to rise.

In continental Europe, one venue for learning independent living skills are sheltered living facilities, which are housing communities where tenants with cognitive impairments live together with supervision from at least one caretaker. In these facilities, 20 to 30 people share living quarters and are coordinated in learning daily living skills by expert staff members and teams of volunteers. The ultimate goal of a sheltered living facility is to teach the inhabitants to live independently with other tenants who help and support each other. The organization supports tenants in moving to a shared apartment, where four to six people help each other without any caretaker supervision. Consequently, the facilities provide training in both the acquisition of life skills and the social organization of a community of tenants.

Communal cooking is one activity that integrates both of these aspects of living in sheltered housing facilities. Tenants learn to complete their kitchen duties to benefit other members of the community. In sheltered housing, instructors provide the group with specialized supervision on safe cooking methods and techniques to cook collaboratively. However, due to worker shortages in this field, individual instructions are rarely possible. Further, assistance in cooking may still be needed by tenants who have recently transitioned to an independent living facility.

Contextualized assistance (*e.g.*, delivered through displays [136], augmented reality [128], or context-aware support systems [358]) has shown to be effective in helping people with cognitive impairments perform chores independently. While the workload from caretakers may be alleviated when using digital assistants in communal cooking activities, support and learning can be provided independently from the sheltered living facility. In this chapter, we describe current training practices in sheltered living facilities. Furthermore, we explore potential challenges and benefits of providing *communal* contextual instruction and support for the development of independent living skills. Using contextual observation of communal cooking sessions in sheltered living facilities, we conduct supplementary interviews with staff members and volunteers. We chart the opportunities, trade-offs, and constraints involved in the design of assistive systems for communal kitchens for persons with cognitive impairments. In this chapter, we report a qualitative inquiry of current communal cooking practices in sheltered establishments. Based on the qualitative inquiries, we describe four themes regarding the design opportunities and constraints in a communal kitchen for persons with cognitive impairments. Finally, we conclude with five implications for assistive technologies supporting cooking in sheltered housing. The research question we answer is:

→ **RQ1**: What are the design requirements for systems that provide opportune cognitive augmentation?

This section is based on the following publication:

- Thomas Kosch, Paweł W. Woźniak, Erin Brady, and Albrecht Schmidt. Smart Kitchens for People with Cognitive Impairments: A Qualitative Study of Design Requirements. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI'18, pages 271:1–271:12, New York, NY, USA, 2018. ACM

3.1 Related Work

The work presented in this chapter bases on the use of assistive technologies by persons with cognitive impairments as outlined in Section 2.3.2. This section provides an overview of technologies supporting persons with cognitive impairments during everyday tasks and describes current research which aims to provide contextual assistance in kitchens.

3.1.1 Accessibility through Assistive Systems

Assistive technologies have proliferated into the home environments of people with mental and physical disabilities. Users of assistive technologies are ubiquitously surrounded by technologies that monitor them. This includes smartphones, tablets, ambient computers, or wearable devices, all of which have become increasingly popular. Assistive technologies intend to provide mental assistance for persons with cognitive impairments. Past research has investigated the design of assistive technologies for persons with cognitive impairments and elderly people that provide mental assistance, and to this end, reduce the complexity of the underlying system.

Bouchard et al. [40] developed a plan recognition framework in smart homes for people suffering from dementia. Using microsensors, the framework analyses previous actions to predict the original intention of the person. Pollack et al. [307] define three goals for assistive technologies when used by older people with cognitive impairments: providing assurance for elderly, use technologies to compensate for impairment, assess the users' status. Furthermore, three different types of assistive systems are defined: assurance, compensation, and assessment systems. Since smart environments need to be equipped with microsensors to measure contextual parameters of their surroundings [245], ethical implications for assistance systems in home environments were to be defined [349]. Assistive technologies provide several benefits, such as workload reduction, socialization, or care delivery. However, the use of such technology has ethical ramifications and considerations (see Section 2.3.3 for a comprehensive elaboration). To enable independent living for the people with cognitive impairments, a smart home system leveraging several sensors, such as infrared motion sensors, microphone arrays, and accelerometers, was used to sense contextual data [8]. The

collected data was used to train artificial intelligence, which provides adaptive assistance whenever required. However, challenges such as preserving privacy [81, 82] and acceptance by senior users [48, 92] have to be considered. Mihailidis et al. [267] present the concept of an assistive system that supports older adults with dementia during regular daily routine tasks. Specifically, this includes a study where washing hands were trained via an audio-based system.

Within workplaces in industrial areas, augmented reality was used to provide cognitive alleviation for workers with cognitive impairments in workplaces [67] and assembly lines [128]. They found that by projecting in-situ information about the current assembly step, increased efficiency concerning time and number of errors would be achieved, and concluded that assistive systems at workplaces foster learning by skill transfer while releasing cognitive resources at the same time [127, 136]. Gamification can be incorporated into assistive systems to maintain or increase motivation [212, 213, 214].

Research has also concerned notifying persons about events, such as warning of dangers or sending reminders. In previous research, we [217, 228] investigated how events of interest can be communicated efficiently with persons with cognitive impairments within workplaces. By comparing visual, auditory, and tactile feedback, their findings show that visual in-situ cues are perceived efficiently. A survey exploring how notifications in smart homes can be communicated was carried out by Voit et al. [377]. Leveraging an online survey, they investigated suitable devices and locations displaying notifications. Their results showed that smart home-related notifications should be received by mobile devices which would be easily perceivable when worn on the body. Wiehr et al. [387] define how the increasing number of notifications raises challenges regarding the design, implementation, and psychological factors of users.

3.1.2   Assistive Technologies in Kitchens

Cooking represents an important communal and social activity for people with cognitive impairments. Cooking entails a starting point (*e.g.*, starting to prepare the ingredients), a structured or unstructured processing pipeline (*e.g.*, cooking using a fixed or unstructured procedure), and an endpoint that facilitates the cooking result (*i.e.*, the cooked meal). Previous research has researched smart kitchens extensively as it revealed challenges and opportunities that are generalized to other assistive technologies.

Thereby, smart kitchens have been the focus of Blasco et al. [34]. They developed and assessed smart kitchens for older adults and evaluated the simulation of specific situations, such as making dinner or washing up. Besides assisting in kitchens, calorie and nutrition-aware contextual cooking plans can be provided [72, 73]. By displaying information about nutrition during cooking, healthier ingredients can be chosen by cooks. Hashimoto et al. [163] designed algorithms for smart kitchens, which recognize the users' cooking actions and food material. Since recipes play an important role in cooking, Schneider et al. [341]

developed a semantic cookbook. The semantic cookbook is a system enabling different parties to share their recipes among their smart kitchens to display it on an output device. As most handwritten recipes are passed down by previous generations, digital recordings can ensure the continued existence of recipes and can be easily shared. Cooking can also represent a method for social communication and interaction. Therefore, Terrenghi et al. [360] present the "Living Cookbook". In their work, cooking experiences are recorded to educate others, practice cooking techniques, or share cooking experiences. An evaluation of the "Living Cookbook" shows that its use increased motivation and improved social communication. Hooper et al. [171] investigates how the new material can be learned within instrumentalized environments to support task-based learning. During their research, the efficiency and design space of learning new languages within regular cooking tasks were evaluated. Bonanni et al. [38] evaluate several augmented kitchen interfaces regarding their usability. Their studies focus on usability as well as interfaces that are not demanding in terms of attention and cognitive workload. Miyawaki et al. [272] prototyped a kitchen for people with higher brain dysfunction. Olivier et al. [288] presents a lab-based replication of a smart kitchen, where designers can evaluate novel solutions. Also, Scheible et al. [334] show how social and emotional components can be incorporated into the cooking process.

3.1.3 Collaborative Accessibility

People with disabilities often *co-construct accessibility* in spaces alongside the other disabled or able-bodied peers who inhabit them [42]. This process of co-construction involves a large amount of collaboration among individuals as they stage environments to be more accessible; divide tasks and responsibilities based on ability, and intervene or request assistance for tasks that are not accessible.

Branham and Kane [42] described the collaborative accessibility practices of blind and sighted domestic partners in their home settings. Within these close partnerships, partners can set up plans for managing household tasks, object placement, and requests for assistance. However, this collaboration to make accessible environments can be less successful in work settings, where an intimate understanding of the disability is not possessed by a person with a disability or co-workers [43]; or in public spaces where accessibility cannot be co-constructed in advance [388].

The study of this collaboration work has been rare among people with cognitive impairments. In our observations in this study, we examined the collaboration between individual tenants with cognitive impairments and the neurotypical[1] instructors who supervised them; however, we also discuss opportunities for collaboration between multiple tenants with cognitive impairments without supervision.

---

[1] We use the term "neurotypical" here to describe individuals without any measurable level of cognitive impairment. This term is often used to differentiate people with a neurological disorder, like individuals on the Autism spectrum or with cognitive impairments, from individuals without a neurological disorder.

Overall, previous work has investigated effort into the development, design, and evaluation of assistive systems. However, the design space of smart kitchens for people with cognitive impairments has not been considered yet in related research. Through the execution of qualitative contextual inquiries with caretakers and observations of tenants in sheltered living organizations for people with cognitive disabilities, we close this gap by charting the design space for assistive technologies in smart kitchens for users with cognitive impairments.

3.2 Methods

Sheltered living facilities offer people with cognitive impairments assistance with learning everyday tasks, including cooking. The main goal is to teach elementary skills in a methodical way that can be reapplied in independent living environments. When tenants make progress in applying their skills without caretaker intervention in the facility, they can move to live more independently in houses shared with a small number of other people with cognitive impairments. The research for this project was conducted in collaboration with a sheltered living organization in Germany which operates both a sheltered living facility and independent houses for tenants who have completed their training.

3.2.1 Context: Communal Kitchens

The kitchen in the sheltered living facility and independent living homes use regular components necessary to cook a meal, such as an oven, stove, tabletop, and refrigerator. The kitchen from the sheltered living facility is shown in Figure 3.1.

Within the sheltered housing, tenants and caretakers cook dinner every weekday evening. If enough inhabitants stay at the sheltered housing during weekends, both lunch and dinner are cooked. The way supervision is provided depends on whether the inhabitants are living in sheltered facilities or residing in their own independent homes.

Major supervision during cooking is mostly provided by a caretaker within the sheltered housing. A sheltered living facility comprises 20 to 30 inhabitants depending on its size. Groups of four to six inhabitants cook together with at least two caretakers who explain the course of cooking in the beginning, distribute tasks, and make sure that the participants avoid serious injuries. Currently, no standardized assistive technology supporting caretakers is enrolled in such facilities. Dangerous tasks, such as operating the oven, are only performed by caretakers.

Minor supervision is necessary for persons with cognitive impairments living in independent housing. This independent place is managed by the sheltered living organization but spatially separated from the sheltered living facility. Such independent living places comprise between four and six inhabitants living together. This principle enables maintaining social settings where people can help each other. For cooking, the same rules apply for sheltered

Figure 3.1: A kitchen in a sheltered living facility, used by the tenants with cognitive impairments and the caretakers. Tenants overtake practiced tasks which include cutting ingredients or laying the table. Dangerous tasks, such as the operation of the oven or stove, are accomplished by the caretakers.

housing. However, only a single caretaker is present to avoid possible injuries and provide advice on demand. The caretaker also assesses the social setting, cognitive development, and skill progress of the inhabitants.

3.2.2 Data Collection

Data for this study was collected through two observation sessions of communal cooking in a sheltered living facility and an independent living facility in Germany and was supplemented by semi-structured interviews with staff members from those organizations. The tenants who participated in the cooking sessions ranged between 30 and 49 years old. Ethical approval for all components of the study was given by the sheltered living organization and the German Federal Ministry of Education and Research according to institutional guidelines.

We chose to conduct observations rather than interspersing interview questions throughout the task as in a contextual inquiry, due to the level of cognitive impairment among our participants and the time-based nature of the cooking tasks being taught. Cognitive impairment impacts executive functioning, like task planning and memory. Interruptions during a time-based task make it more difficult for a person with a cognitive impairment to hold onto their original intention, and less likely to return to the task after being distracted [362]. At the direction of the caretakers at the facility, we limited interaction with the tenants during the cooking tasks and used follow-up interviews with staff members to supplement our understanding of the observed data. The contextual inquiry was thus carried out with the caretakers, which have been observed and interviewed during this study [33]. Participants were monetary compensated through funding from a joint project.

3.2.3 Observations

We conducted two observations in communal kitchens as tenants prepared meals. We observed communal cooking processes among four to eight participants - the number of tenants in the kitchen at one time varied as they arrived and left the kitchen at their discretion, but a total of twelve participants were observed between the two sessions. Two rounds of observation were performed — one of a *supervised cooking session* in the sheltered living facility, and one of a *unsupervised cooking session* in an independent household managed by the same organization. Each observation session lasted at least one hour. A caretaker was present in each kitchen at all times during our observations but did not intervene in the cooking activities in the unsupervised session.

In the supervised cooking session, ten participants we observed were born with a cognitive impairment limiting their ability to understand and process information. In addition to this, two other participants were affected by light motoric impairments. All participants we observed in the unsupervised cooking session were affected by a cognitive impairment limiting their ability to understand and process information. None of the participants were affected by sensory impairments or dementia.

Our observations were documented through notes and photographs. We took pictures of the cooking environments before the tenants began cooking to limit task disruptions. During the observation, two researchers took detailed field notes on the session. Observations focused on how the cooking process was organized and the interactions between the tenants. In the observation of the structured cooking session, we also observed how the caretakers provided instruction to the tenants and how they directed their attention. We noted how physical artifacts in the kitchens were used, arranged and assigned by the tenants. In each of the two observations, we remained in the kitchen for the entire duration of the cooking process, from before the tenants started arriving in the kitchen to when they left the room after having eaten. We have not interacted with the tenants directly as we wanted to avert confusion and avoid disruptions during the regular cooking procedure.

3.2.4 Interviews

After the observations, we constructed specific semi-structured interview protocols for the caretakers to clarify and refine our understanding of the data collected during the cooking sessions. This enabled us to obtain information that could not be disclosed in the presence of the tenants and have the full attention of the caretaker. Table 3.1 summarizes the participants' role and working experience. The primary focus of these interviews was to better understand the work processes and roles assigned involved in the structured cooking sessions, such as the reasons for distributing particular tasks, rules for using appliances, and methods for distributing the workload among tenants. We were also able to inquire further about the role of cooking in the community of tenants in both structured and unstructured settings.

Participant	Role	Age	Experience
P1	Housing officer	33	15 years
P2	Sheltered caretaker	49	30 years
P3	Sheltered caretaker	20	0.5 years
P4	Independent caretaker	46	6 years

Table 3.1: Demographic data and working experience of the interviewed participants.

The interviews comprised one interview with the housing officer of a sheltered living institution (a person whose full-time job is to manage the facilities), two interviews with volunteer caretakers working at a sheltered housing facility, and one caretaker responsible for an independent group of persons living together. Overall, 5:45 hours of recordings were collected. The mean age ranged from 20 to 49 years (M = 37, SD = 13.29) and working experience of caretakers ranged from six months to 30 years (M = 12.88, SD = 12.87). P1, who is a housing officer and responsible for multiple living facilities, provided us with most data about the observed cooking processes as well as general information. The interviews from P2, P3, and P4 were largely used to gain new insights and to confirm the statements provided by the housing officer.

3.2.5 Data Analysis

We conducted a qualitative analysis of our observation notes and interview transcripts to understand the constraints and opportunities for communal cooking experiences for tenants. We used a team consisting of two coders working with the Atlas.ti software package. We conducted an initial open coding of 25% of the data by both researchers. Afterward, a coding tree was established through iterative meetings. We then coded the rest of the data with the agreed-upon set of codes. A final meeting was conducted where we grouped codes to establish the four emergent themes: WORK ORGANIZATION, COMMUNITY, SUPERVISION, and PRACTICALITIES.

3.3 Findings

In this section, we present four themes that emerged from the analysis of our data set. We present each theme and our understanding of the constraints, tradeoffs, and opportunities involved. The presented quotes have been transcribed from their original German into English.

3.3.1 Work Organization

Cooking was seen as a key activity in the learning process required for gaining a higher degree of independence. In our observations, both tenants and supervisors gave heavy at-

tention and concern to how tasks were divided within the cooking group. Much of this task division within the supervised cooking session was done by the supervisors in advance, and tasks were created both for the tenants cooking that day and for the supervisor on duty. The supervisors communicated clearly which tasks were intended to be performed by tenants and which were reserved for supervisors:

> *"Tasks are usually divided between tenants and supervisors. The parts were carried out until completed; nobody stops suddenly within their task." (P1)*

One supervisor showed us how a weekly cooking plan helped organize the communal activities in the group. The cooking duties were distributed every week within a weekly meeting where supervisors and tenants took part. The cooking responsibilities were discussed verbally. Tenants were informed about their responsibilities by the assigned supervisors.

These plans served as a valuable organization and accountability tools on tenants' journey to greater domestic independence. One officer explained that managing schedules and making sure all tenants were involved in cooking served not only practical purposes but also worked as a means of assuring participation in this important activity of daily living on days when tenants were not intrinsically motivated:

> *"Many tenants do not feel like they want to cook. . . A cooking plan is created, which forces everyone to cook regularly. Usually, we find out whether there are any important appointments before a cooking schedule is created." (P1)*

The supervisors assumed responsibility for developing these plans, leveraging their knowledge of tenants' current ability levels and other commitments to make the plans maximally effective in the limited time available.

Coordination in the kitchen was necessary, as the kitchens constituted the only common rooms in the facilities we visited. We observed tenants constantly entering and exiting the kitchens. Some tenants who were not involved in cooking tasks that day came to socialize or observe the cooking tenants. Other tenants who were assigned to specific cooking duties might leave their task briefly, and return to them after a short walk. The diverse patterns of the tenants within the kitchen at any time meant that supervisors had to engage in constant monitoring of task completion to ensure that coordination was occurring.

The division of labor among stakeholders was emphasized not only in terms of which tenants performed which kitchen duties, but also concerning assigning paid personnel and volunteers to supervision:

> *"There is a duty roster for interns and employees. This plan specifies, how many supervisors have to be there during cooking." (P2)*

External factors also affected the way kitchen work was organized. For example, drinks provided by an external supplier arrived on a fixed schedule that also needed to be included in the kitchen plans:

> *"Drinks are delivered every week on Wednesday. The group buys groceries by themselves every Tuesday and Friday. On Wednesday, we buy groceries together. Usually, one supervisor is associated with each tenant." (P1)*

These considerations show a considerable amount of coordination that is required for the supervisors and tenants in the supervised cooking environments. While the meta-level scheduling work was performed solely by the supervisors, individual cooking sessions and supportive tasks were more collaborative efforts between volunteers and tenants.

Overall, we observed the multiple aspects of the division of labor, time management, and logistics, all of which significantly affected the way the cooking was enacted.

3.3.2 Community

The social dynamics of the sheltered living community were important to both tenants and to the supervisors, who saw these social relationships as valuable learning experiences. Aspects of communication and collaboration between tenants and supervisors were often observed and mentioned in our data set. Tenants interacted with each other while cooking and the supervisors reported that due to social expectations, tenants did not eat alone and instead joined communal activities:

> *"Nobody would cook together and then decide to eat alone in his room." (P1)*

Tenants who were not assigned cooking tasks enjoyed observing the cooking tenants performing their duties and would choose to join the kitchen and monitor the cooks as they worked through their tasks.

While the tenants did not initiate social activities outside of the facility, the supervisors often planned these events as an important training experience for gaining independence:

> *"We go to the cinema or do other activities on weekends. But this depends on the initiative of the supervisors." (P1)*

These independence-building activities helped to foster community among the tenants and supervisors, and could be useful in preparing tenants to navigate social relationships in the more independent housing options.

Despite this emergent community among the tenants and supervisors, there was potential for conflicts to occur connected with the execution of the kitchen work. Supervisors, in particular, were concerned about these potential conflicts and helping the tenants resolve these conflicts amicably. One supervisor reflected on how some tenants were particularly concerned about the alignment of cutlery on the kitchen table, indicating that tenants with different levels of ability might experience frustration with each other:

"How exactly cutlery is placed depends on the person. Some execute their tasks very accurately, while others do their tasks in a rudimentary way. We often experience conflicts between inhabitants because of that." (P1)

Similarly, another supervisor was concerned about how critique could be communicated gently, and how it affected tenants differently:

"Some have problems accepting critique and answer with statements like "This has been always been done this way." (P2)

These interpersonal disagreements are likely to arise in more independent kitchens as well, so supervisors desired more effective ways to resolve these conflicts.

Within the community of tenants, different tenants assumed unique roles in the cooking process. This often was at odds with the goals of the supervisors, who wanted tenants to participate in cooking equally as part of their rehabilitation. Instead, more motivated tenants often tried to monopolize the cooking positions:

"Our tenants are very motivated to shop for the ingredients. However, the motivation regarding cooking is different per person. Surprisingly, it is always the same persons who volunteer for kitchen service." (P3)

Some tenants were not intrinsically motivated to do the complex work required in the kitchen. In these instances, social factors were often mentioned by supervisors as key to increasing their motivation. Enabling tenants to express approval of each other's actions and their role in the community was seen as extremely valuable to the group dynamics:

"Approval of other tenants is important; much more important than approval from the supervisors." (P2)

3.3.3   Supervision

The nature of the supervision provided to the tenants was a unique aspect of work in a sheltered housing facility. Supervisors explained that maintaining engagement and limiting frustration for tenants was one of the most important purposes of their work and one of the primary goals of the more structured cooking sessions. They conducted this supervision work both pre-emptively (while doing WORK ORGANIZATION, as described above) and by being available to help or intervene during the tasks themselves:

"We sort from the beginning inhabitants for specific tasks; for example if it is clear that a particular person should not cut hard things instead of vegetables and salad. They call for help if something is too difficult for them. We have to avoid frustration and confusion if inhabitants are not able to do their task." (P1)

Tenants were able to request help if needed, and the amount of supervision needed varied from tenant to tenant. In general, there was a high emphasis on tenant independence, and the details in the instructions provided to the tenants were left sparse to allow them to develop skills on their own. In their role as supervisors, the staff was most concerned that a given individual would be able to generally complete kitchen actions and that they pay attention to learning precise techniques to a sufficient degree:

> *"Most people can cut. We do not complain if something is cut too thick or thin. The most important thing is that it is cut and can be cooked." (P1)*

Guidance from supervisors was primarily required when tenants made mistakes or grew confused. When this occurred, both supervisors and other tenants stepped into a supervisory role to help the tenants understand the issues with their work. Feedback from all members of the group was often provided. Tenants who intervened were often concerned about the process implications of an error (*e.g.*, if a mistake would impact the overall meal) or highlighted safety concerns when they noticed unsafe behavior. For supervisors, helping the tenant understand what went wrong in the task execution was the key educational aspect:

> *"This [intervention] is very individual. It always depends on the mistake they make. Some do what we tell them and they do not get confused while doing their kitchen tasks and accept the critique." (P2)*

Certain complex aspects of cooking required constant verification from the supervisors and fewer opportunities for independence. For these tasks, the supervisors often overstepped their role as supervisors to join the cooking process directly. For example, the spice cupboard required careful instruction and often the supervisors would administer spices without any help:

> *"Managing spices is an interesting question: Who dispenses the spices? Our experience shows that adding spices is a fine-motor task requiring experience to not, for instance, over salt a meal. Spicing is mostly done by supervisors." (P1)*

However, some tenants had stronger fine motor skills than others and performed spicing in the supervised session with direct instructor supervision.

This intervention was in part based on tenants' skill limitations and was an integrated part of the supervision. Supervisors needed to bear in mind the specific tenants' perception of the cooking experience in each session, and be aware of stimuli that the tenants could not always process:

> *"I do not think that persons with cognitive impairments are able to determine when a meal is cooked or not." (P1)*

Finally, supervisors recognized that more instructions could be provided to the tenants and they could benefit from more attention to improve their long-term educational outcomes within the kitchens. However, the number of personnel available for communal cooking was limited:

> *"It is a communal kitchen. They have their room, but in principle, they should be able to cook for themselves. The process of learning how to cook independently in one's own house should be learned here." (P1)*

3.3.4 Practicalities

Our last theme addresses the practical concerns involved in cooking activities. A core concern is maintaining the safety of the tenants, who may enter the facility with little knowledge about safe cooking procedures. As many kitchen tools are potentially dangerous and electrical appliances were in use, supervisors were especially careful to monitor tenants as they performed potentially dangerous actions, but had to make crucial tradeoffs between safety concerns and independence:

> *"Knives, blender, everything that can cause potential injuries is dangerous. We try to avoid every chance where injuries could happen. But this is not right since they will not learn how to cut or blend... however, they* have *to learn how to cut or blend [to cook independently]. Therefore, negative experiences are necessary to learn what causes injuries. But our primary goal is to avoid these injuries as much as we can." (P1)*

Thus, the staff made risk assessments to determine if the use of particular skills outweighed their danger. Further, some supervisors mentioned that they prioritized the use of devices that offered a greater degree of safety, such as the microwave oven:

> *"The microwave offers comfort and security. Food is just put in and the microwave tells them when the food is ready." (P1)*

Some tenants were aware of the safety risks of certain tools, and would monitor each other and repeatedly provide feedback when they observed safety violations.

Another facet of the cooking experience that we observed was making sure that the process was hygienic. Hygiene rules were imposed on the facilities by legal regulations, and needed to be followed and enforced strictly by the staff:

> *"We provide a hygienic plan. The housekeeping management organizes hygiene schooling. Furthermore, hygiene is taught to the tenants over time, for instance, how to wash hands, disinfect them, and how to wear gloves. Additionally, supervisors know the hygiene-related habits of tenants. This means that the individual hygiene is very diverse." (P1)*

The medical condition of the tenants also affected the cooking experience. Distributing medicine was an integrated part of the communal cooking process, and the supervisors distributed pre-mixed dosages of pills in color-coded containers. The choice of the menu was also affected by health concerns:

> *"We serve salad as a side dish because of health reasons to ensure enough vitamins in meals. If we ask what they want to eat, they would prefer sausages, fried chicken, or fries all the time." (P1)*

Lastly, participants were aware that cooking was connected with other activities, such as setting the table, and different tenants would perform these tasks in parallel with the cooking tasks. While these activities required less help, supervisors still monitored if they were completed and coordinated:

> *"The tenants doing kitchen service needs to set up the table accordingly; including drinks and cutlery." (P1)*

3.4 Implications for Design

We present the lessons learned in the form of implications for design. These implications present an overview of the design space of assistive systems for cooking in sheltered living facilities. The implications can be used by future designers to assure that possible solutions for the cooking benefit the users with cognitive impairments. These implications are derived from the themes presented above. Although these implications are derived for persons with cognitive deficiencies, we believe that these could be generalized to implement smart kitchens for elderly populations, who often begin to experience cognitive decline through the aging process [34, 66, 82], as well as assistance systems integrated into other household environments.

3.4.1 Support Clear Task Division

From our inquiries, we learned that tasks have to be intelligently distributed and communicated among tenants. The goal of the tasks was communicated verbally since most of the tenants were not able to read or write. During our studies, we found that continuous support was needed. Supervisors were reminding tenants from time to time about their current task and kept them motivated. Furthermore, the quality of their work was evaluated and criticized by both supervisors and other tenants when necessary, as we observed in the SU-PERVISION theme. As a consequence, future smart kitchens should communicate recipes and tasks; *explicitly providing feedback on the component tasks that make up the cooking process*. Furthermore, attention and motivation should be fostered by highlighting the current

task to be performed repeatedly. Division of labor should be explicitly supported, giving the supervisor freedom in orchestrating the cooking experience based on their awareness of tenants' diverse capabilities, and should help them maintain an overview of all the activities in the kitchen.

3.4.2 Embrace the Group Experience

Our analysis revealed that maintaining social ties between supervisors and cohabitants was an important factor on each tenant's route to greater independence. In the COMMUNITY theme, we observed how cooking represented an important social activity, where tenants worked together to achieve a communal goal. Managing social activities and minimizing opportunities for conflicts within the group was regarded as a key role for supervisors. Furthermore, some tenants were not motivated to cook, since it required extra effort and paying attention to additional constraints as shown in the PRACTICALITIES theme, but were more motivated by their role in the larger community of the kitchen. Our research suggests that adaptive motivating elements should be *integrated to maintain and augment the social experience of cooking together*. For example, showing visible achievement metrics or cheering cartoon figures as a token of appreciation for completing a task may foster social interaction. Further, assistance systems should support conflict resolution by accounting for the tenants' unique abilities and personalities.

3.4.3 Prioritize High-Safety Instructions

Safety during cooking was the main concern for supervisors. As persons with cognitive impairments perceive pain differently, the risk of severe injuries is higher than among neurotypical populations [50]. In the PRACTICALITIES and SUPERVISION themes, we observed how instruction always prioritized safety features for tasks that posed a possible danger. Consequently, smart kitchen systems for tenants should *only feature displays in safe areas*, where the probability of getting injured is minimized. As dangerous tasks, such as operating the stove, cannot be fully avoided, the system should communicate safety hazards to avoid severe injuries. As we observed, some of the tenants had issues retaining focus; drifting in and out of the kitchen space during the completion of a single task. Future systems should *clearly show dangers that have a temporal aspect*, for instance, communicating that a hot plate cannot be touched until it has cooled down to a safe temperature.

3.4.4 Enable Customization for Different Abilities

In our work, we observed users with cognitive impairments with a diverse set of abilities and impairments. In the PRACTICALITIES theme, we observed that some users were able to occupy themselves with dispensing spices, measuring water, or performing tasks that required fine motor skills; while these tasks were inaccessible to others. While supervisors

strived to find an optimal division of the tasks, managing multiple parallel activities and the complex learning curves of multiple tenants was a very complex task as shown in the SUPERVISION theme. Future systems can not only aid in finding optimal ways to divide the tasks involved in preparing a meal but also *adjust the difficulty of the task to support the learning process* of a given tenant. It is important to note that the information provided should foster individual abilities, therefore the system should monitor the learning process and adaptively increase or decrease assistance. Furthermore, once tasks are complete, instant individualized rewards should be available to maintain management and support communication with the supervisor.

3.4.5 Provide Opportunities for Explicit Communication

The COMMUNITY theme showed how users often required confirmation from their peer group or superiors when performing cooking duties. Further, in the WORK ORGANIZATION and PRACTICALITIES themes, we observed how the logistics of food preparation and constant movement in the kitchen affected attention and communication between the supervisors and tenants. Consequently, we see opportunities for fostering reporting to supervisors and increased awareness of when tasks are finished or in progress. Future assistance systems for kitchens for persons with cognitive impairments should *explicitly encourage users to communicate with the supervisor* and show the group when tasks are completed. This, however, needs to be done in ways that do not provide distractions that may affect cooking performance.

3.5 Discussion

The considerations we present above focus on designing for assistance in kitchens within sheltered living facilities. In contrast, here we brainstorm future directions for assistive technologies that assist persons with cognitive impairments in unsupervised, independent, or group-based cooking sessions.

3.5.1 Accessible Assignment of Complex Tasks

One of the primary roles of the instructors in the kitchens was dividing tasks among the tenants, monitoring ongoing preparations, and coordinating their work to create a full meal. We see this coordination work as an important space where intelligent technologies could supplement or replace the trained instructors in the future, both to address the staffing shortages that limit the use of sheltered living facilities and to enable individuals with cognitive impairments to train in these skills from their own homes or the homes of family members.

The instructors created long-term plans for cooking tasks which pushed tenants to continue building their cooking skills and to complete cooking tasks even when they were not intrinsically motivated to cook at that particular time. Intelligent systems which assign cooking

tasks must balance these complex needs by generating tasks that are incrementally more difficult over time, and by sensing and responding to users' current level of ability and motivation in a given cooking session. These ability levels may change dramatically between sessions, especially among individuals with cognitive impairments caused by brain injury or aging, and an automated system must be able to flexibly adapt after observing a users' competencies or difficulties in the kitchen.

Intelligent systems may also be useful for home environments, where a person with cognitive impairment may wish to cook collaboratively with neurotypical family members. Family members of people with cognitive impairments can be inadvertent perpetuates of disability stigma against their loved ones, viewing them as less capable and giving them simplistic tasks and responsibilities [99]. These systems could engage in task division between the family members by identifying tasks that are appropriately complex for both the persons with cognitive impairments and neurotypical participants, reducing biases around cognitive impairment and allowing the person with cognitive impairment to maintain their role within the family structure.

3.5.2 Examining Collaborative Accessibility

Prior work has found that norms in environments where people with disabilities are the majority (*e.g.*, the National Federation of the Blind's annual conference) differ from norms in environments where people with disabilities are the minority [107]. In the independent housing areas of the sheltered living facilities we studied, all members of the cooking process have cognitive impairments, raising critical questions about how technologies can be designed to facilitate collaborative work without impacting the disability-specific norms of the tenants.

In the supervised observations, tenants turned to the instructing caregiver to ask questions or get feedback. This reliance on the instructor may make it difficult for tenants to develop truly independent skills in a supervised setting. As a result, the instructors gradually reduce their involvement until the tenants can demonstrate a level of independence that qualifies them for independent housing. However, the 'independent' housing is anything but independent – tenants are surrounded by other residents with cognitive impairments who they can turn to for support or instruction. Designing for these *collaborative* domestic settings is rarely studied in accessibility literature or is looked at between disabled and able-bodied partners rather than among members of a single disabled community. Future study of cooking practices among individuals with cognitive impairments can contribute greatly to our understanding of the co-construction of accessibility.

We recognize the importance of including representative users in accessibility research, as argued by [343], and strove to accurately represent the cooking practices of the tenants with cognitive impairments we observed. However, due to the level of cognitive impairment among the tenants, and the facility's desire to minimize researchers' direct intervention

into the cooking process, we could not perform a typical contextual inquiry with the tenants. Instead, we combined our observations with interview data from the facility staff, who may have their own biases in their interpretations of the tenants' actions and interactions [99]. While other HCI work studying users with cognitive impairments also relies on caregiver stakeholders as informants (*e.g.*, [90]), we see this as a major limitation of our current work. As we continue this thread of research, we intend to develop new methods to facilitate working directly with tenants with cognitive impairments, perhaps drawing from prior work which leverages participatory design methods [122].

3.5.3 Study Limitations

Our contextual observations and evaluation were affected by other constraints. The study and data collection were conducted focusing on West European living standards. The sheltered living facility model is unique to this context. Compared to other locations, different living habits regarding persons with cognitive impairments may be present. Additionally, both of our observations were conducted in different sections of the same sheltered living organization, whereas cooking procedures in other facilities may be executed differently. Communication of complex or novel information can confuse, while simplified visualizations may neglect to foster cognitive abilities and should be considered in a follow-up study.

3.6 Study Conclusion

We investigated the design space of smart kitchens for persons with cognitive impairments living in sheltered housing facilities. Within qualitative contextual inquiries, we obtained data by observing persons with cognitive impairments during a cooking session. Furthermore, interviews were conducted to explore current design gaps of assistive technologies in-depth. We derived the four relevant themes WORK ORGANIZATION, COMMUNITY, SUPERVISION, and PRACTICALITIES, which emerge as important factors, through a thematic analysis of the interviews. We conclude with five design implications which should be considered when designing smart kitchens for persons with cognitive impairments: clear communication of tasks, fostering the group experience, prioritizing safety, providing rewards, and enabling contextual adaptivity.

3.7 Chapter Summary

In this chapter, we conducted a qualitative inquiry to reveal opportune moments and potential interventions for cognitive assistance through workload-aware interfaces. Through the combination of observations and interviews with caretakers of a sheltered living facility, we

derived four themes as well as five design implications for the design of assistive technologies that augment cognitive capabilities. Persons with cognitive impairments were studied as a target group since they express the need for cognitive assistance explicitly to their supervisors. In a first step to understanding how tenants with cognitive impairments process and act using the available information, we conducted an observation to gather information during a communal cooking session. Using the results from the observations, we created a question catalog that addressed aspects of cognitive augmentation through assistive technologies.

Our findings show how persons with cognitive impairments collaborate during cooking with other tenants as well as the supervisor. We find that social collaborations, especially the social ties between tenant and supervisor, play an important role. For example, caretakers repeatedly reminded the tenants to be careful during dangerous or mentally demanding tasks. Situations that may cause injuries are therefore one particular moment that has the potential to benefit from cognitive assistance. We further found that communal settings require a clear task division. It is difficult for persons with cognitive impairments to manage task division by themselves without clear instructions from a caretaker. Assistive technologies can use this circumstance to deliver in-situ cognitive assistance. Our interviews revealed that the caretakers were compelled by the vision of constant cognitive augmentation. Although delivered through ambient visualizations, assistive technologies should be attentive during the whole cooking procedure and deliver assistance whenever needed. Hence, **RQ1** is answered by the presentation of the associated design implications and the aforementioned situations that benefit from cognitive assistance.

Chapter 4

Feedback Modalities

For every user action, there should be an interface feedback. For frequent and minor actions, the response can be modest, whereas, for infrequent and major actions, the response should be more substantial.

Ben Shneiderman

So far we have investigated user-centric characteristics when encountering situations in a cooking scenario that require cognitive assistance. This resulted in five design implications for workload-aware assistive systems. While these results deepen our understanding to recognize opportune moments for cognitive assistance, the feedback modalities of such systems have to be designed carefully. Cognitively exerting tasks tend to demand the full attention, alertness, and vigilance of a person. Unsuitable feedback modalities may break the attention span and effort that was put into the task. Hence, it can take between ten to 15 minutes before users can return to their original task after being disrupted [177]. Contrary, feedback modalities that demand the attention span frequently have a detrimental effect on the user's task performance. In this chapter, we compare modalities that utilize the visual, auditory, or tactile sensory channel to deliver feedback. For example, visual and auditory feedback can be employed in collaborative work settings while tactile feedback can be used to notify users while preserving privacy. Hence, feedback modalities may be perceived differently by each individual and for each environmental situation. The previous chapter investigated tenants from sheltered living facilities as a target user group for systems that provide cognitive assistance. The roles, as well as social ties between tenants with cognitive impairments and the supervising caretakers, have been investigated to receive insights for the future design of workload-aware interfaces. Thereby, the tenants have shown affinities with technologies that foster communication (*e.g.*, instant messaging or phone calls) throughout their tasks (*e.g.*, cooking).

<center>(a)　　　　　　　　　　　　　　　　(b)</center>
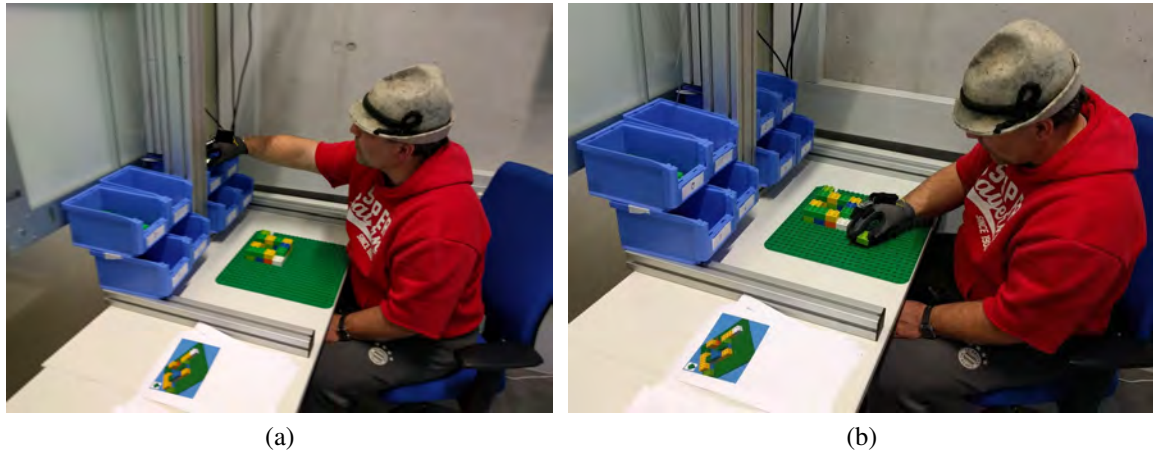
Figure 4.1: A participant is wearing an augmented glove providing tactile error feedback during assembly tasks.

The "United Nations Convention on the Rights of Persons with Disabilities" [1] describes how to ensure and protect fundamental rights of people with disabilities. This comprises the inclusion in daily life tasks, such as work and public leisure activities. Sheltered working has emerged to include persons with cognitive impairments in working life. Sheltered work organizations employ workers with cognitive impairments in a supervised work setting. This establishes social ties with peers and provides an income for personal use. Procedural assembly tasks, such as manual assembly at production lines, are common tasks conducted by persons with cognitive impairments. The assembly instructions can be broken down into a small task complexity to fit the mental capacities of workers with cognitive impairments. Traditionally, workers with cognitive impairments rely on a human instructor who supports them during assembly tasks. The instructor is often responsible for multiple workers which makes it difficult to supervise and control each assembly step of each worker for errors.

The use of assistive systems to provide assembly instructions has proved to reduce the number of assembly errors while decreasing the overall assembly time [136]. Since errors can not be completely avoided, assistive systems can detect assembly errors and provide appropriate feedback. However, error feedback is perceived different on an individual level. The type of intrusiveness, demand for attention, and privacy aspects are factors that have to be considered [131]. Previous approaches conclude that a combination of haptic and visual feedback is a suitable way of presenting error messages through assistive technologies. As these results were collected with participants that were not affected by cognitive impairments, it is unclear if which error feedback is suitable for workers with cognitive impairments. In this chapter, we close this gap by presenting a study with 16 workers with cognitive impairments comparing the three most common modalities for providing error feedback, including tactile, auditory, and visual feedback (see Figure 4.1). We present the results of a user study that investigates the aforementioned feedback modalities and we provide implications when

designing feedback modalities in working environments. This chapter sets the boundaries of communication modalities for workload-aware interfaces:

→ **RQ2**: What are suitable in-situ feedback modalities for cognitive support?

This section is based on the following publication:

- Thomas Kosch, Romina Kettner, Markus Funk, and Albrecht Schmidt. Comparing Tactile, Auditory, and Visual Assembly Error-Feedback for Workers with Cognitive Impairments. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '16, pages 53–60, New York, NY, USA, 2016. ACM

4.1 Related Work

The related work presented in this section includes past evaluations of feedback modalities and developments in assistive technologies for workplaces.

4.1.1 Feedback Modalities

Different feedback modalities have been used for providing information in different scenarios. The most important categories are tactile, auditory, and visual feedback. Tactile feedback is for example used by Bial et al. [32] by using a glove that is equipped with vibrating motors. Results show that tactile feedback can be used for navigation tasks. They use the tactile feedback to provide information for motorcyclists during driving tasks. Considering auditory feedback, Rauterberg and Styger [314] proposed adding additional auditory feedback to traditional visual feedback for assembly tasks. Visual and auditory feedback has been provided at the same time while managing computer-numeric-controlled centers. Their study suggests that combining auditory and visual feedback leads to a more positive mood and improves the participant's performance. The multimodal representation of feedback could lead to a high mental demand for participants with cognitive impairments. Plain visual feedback is for example used by Funk et al. [128] when comparing different visual approaches for providing feedback for workers with cognitive impairments at manual assembly workplaces. In their study, they compare video-based, pictorial and contour instructions. The results of the study suggest that visual contour instructions are perceived well among all Performance Index (PI) groups of workers with cognitive impairments. Moreover, Cuvo et al. [87] use textual feedback while instructing persons with mild cognitive impairments. In their study, they found that performance feedback is crucial.

Other fields already experimented with combinations of haptic, auditory, and visual feedback. Akamatsu et al. [4] compared the three feedback types in a mouse pointing task. Their study revealed interesting design implications although no difference in Task Completion Time (TCT) was found. Furthermore, Richard et al. [316] and Petzold et al. [301] compared the three feedback types when manipulating objects and assembling in virtual environments. Visual feedback was delivered on a screen, auditory feedback was provided by headphones and tactile feedback was triggered using pneumatic micro-cylinders, which applied pressure on the fingertips in a glove. Richard et al. found that both haptic and auditory feedback improve the workers' performance in manipulating virtual objects while Petzold et al. found that the performance is increased using additional haptic feedback.

4.1.2 Assistive Systems for Workplaces

In 1991, Pierre Wellner [383] suggested using a camera-projector system to provide additional digital information for regular physical objects. In his prototype, Wellner combined a digital and a physical workspace for enabling to use the best features of both spaces in one physical workspace. Later Pinhanez [306] was using a camera-projector system combined with a mirror for turning arbitrary surfaces into digital displays. Using Pinhanez's system, nearly every surface can become a display that shows information. Since then, systems using camera projector systems were deployed in different scenarios to provide cognitive assistance. For example, Rüther et al. [323] provide assistance in sterile environments, Löchtefeld et al. [249] augment a shopping scenario, and Butz et al. [62] used it to support searching tasks. In 2008, Bannat et al. [16] used a similar setup for providing assembly instructions at manual assembly workplaces. Their system uses a camera to detect the position of picking bins and a projector, to provide pictorial instruction for the assembly process. Similarly, Büttner et al. [60] uses in-situ projection at manual assembly workplaces by providing pictorial instructions. They found that using in-situ projection is faster and leads to fewer errors compared to displaying instructions on smart glasses [58]. Further, Korn et al. [210, 215] used in-situ projection in combination with gamification approaches for motivating and instructing workers with cognitive impairments during assembly tasks. Their study with workers with cognitive impairments revealed great potential for using gamification and in-situ instructions for instructing and motivating workers at assembly workplaces. However, they did not find a statistically significant difference. Further, a comprehensive summary of assistive systems for supporting workers at the workplace is provided by Korn et al. [211].

Recently, Funk et al. [136] used in-situ projection to provide instructions during assembly tasks at the workplace. Their results show that with increasing complexity, using in-situ instructions is leading to significantly fewer assembly errors and a significantly faster assembly time compared to pictorial instructions. In their system, they were providing red error feedback that illuminates the current picking bin when picking the wrong part. However, no scientific analysis of the error feedback was conducted. Therefore, Funk et al. [131] further investigated the effects of haptic error feedback, auditory error feedback, and visual

(a)　　　　　　　　　(b)　　　　　　　　　(c)
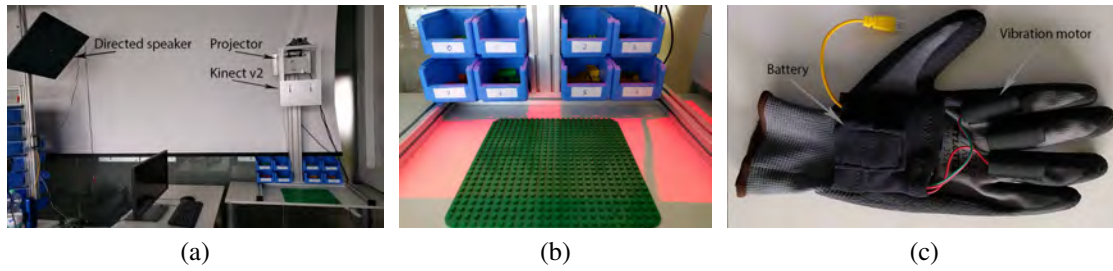
Figure 4.2: (a): The system uses a projector and a directed speaker for providing visual and auditory error feedback. Further, a Kinect v2 observes the picking of items from bins. **(b):** Red light providing visual error feedback. In case of an error, the whole working area is highlighted. **(c):** A glove equipped with vibration motors is further used to provide tactile error feedback.

error feedback by conducting a study with students. Their results reveal that a combination of haptic and visual feedback might be the best way to communicate errors while performing assembly tasks. However, this was only tested with students in a lab study.

Previous work suggested the use of haptic feedback for communicating errors during assembly tasks. This is considered a privacy-presuming way of communicating errors at workplaces [131, 301]. We argue that assistive systems will have a great impact on manual production lines and envision the inclusion of people with cognitive impairments in the work life [26, 329]. Thus, we are interested in how concepts of error feedback influence the efficiency of workers with cognitive impairments. Further, we investigate which error feedback modalities are preferred by the workers themselves.

4.2　System

To incorporate the use of tactile, auditory, and visual error feedback in an assistive system for workplaces, we extended the system presented by Funk et al. [136]. The system consists of modular components that are designed for providing different modalities of error feedback. The main system, which tracks assembly steps and provides feedback to the assembly area, is constructed out of multiple aluminum profiles. The profiles can be used to mount different hardware on top of the workplace. The system uses a top-mounted Kinect v2 [1] to detect picking steps. Therefore, the system can distinguish between correct and incorrect picks.

We placed the system on a table and placed a height-adjustable chair in front of it. Therefore the workers can work while sitting at a comfortable height. The system is constructed in a way that the assembly area is $70\,cm$ wide and $49\,cm$ high. This is enough space for placing instructions and assembling Lego Duplo constructions. The system further features 2×4

---

[1] `https://developer.microsoft.com/en-us/windows/kinect` - last access 2020-01-28

picking bins which are filled with Lego Duplo bricks. Each picking bin is filled with one type of Lego Duplo bricks that is unique in either color or shape. In the middle of the system, there is a Lego Duplo plate firmly taped to the work area (see Figure 4.2b). The system triggers an error when either a picking error was made, or an assembly error was made by the participants. Picking errors are detected automatically by the Kinect v2 if the user places his or her hand in a wrong bin. For detecting assembly errors, the system uses a wizard of oz approach where a study assistant observes the assembly process and uses a wireless presenter for triggering an error. In the case of an assembly error or picking error, an error message is triggered. Depending on the condition, our system is capable of presenting error feedback using the following three modalities:

Visual error feedback is provided using a projector (see Figure 4.2a), which is mounted at the top of the assistive system. If an error occurs, a red light illuminates the whole work area (see Figure 4.2b). While previous approaches only highlight the incorrect picking bin or the incorrect assembly position if an error was made, we decided to illuminate the entire work area. Therefore, the error message is harder to miss.

Auditory error feedback is provided using the Holosonics Audio Spotlight 24i[2] (AS24i) speaker. The speaker uses ultrasonic waves to prevent the sound from diffusing. Therefore, it is only noticeable by the person sitting at the workplace and therefore retains the user's privacy. As an error sound, we are using a deep error tone exactly as used by Funk et al. [131]. In case an error is made, the error sound is played by the speaker (see Figure 4.2a).

Tactile error feedback is provided using a glove equipped with two vibration motors (see Figure 4.2c). We decided to choose standard safety gloves, which are mostly used by workers while performing assembly tasks. The motors are placed on the index finger and the ring finger. Our glove uses an ESP8266 microcontroller and a battery to receive the error trigger messages via WiFi. In comparison to Funk et al. [131], our glove can, therefore, be used without using a wire connected to a computer. This wireless feature makes the gloveless obstructive while performing assembly tasks. In case an error is made, the glove uses an alternating vibration pattern that activates each motor twice directly after each other for 0.3 seconds each. This results in a vibration time of 1.2 seconds per error. A summary of the design and duration of every error feedback modality can be found in Table 4.1.

Stimulus	Feedback Design	Duration in ms
Visual	Projecting red light	2500
Auditory	Playing deep error sound	2000
Haptic	Vibration in worker gloves	1200

Table 4.1: Feedback design and duration of each stimulus used in the study. The duration is depicted in milliseconds (ms).

---

[2] `www.holosonics.com/15-products` - last access 2020-01-28

(a)   (b)   (c)
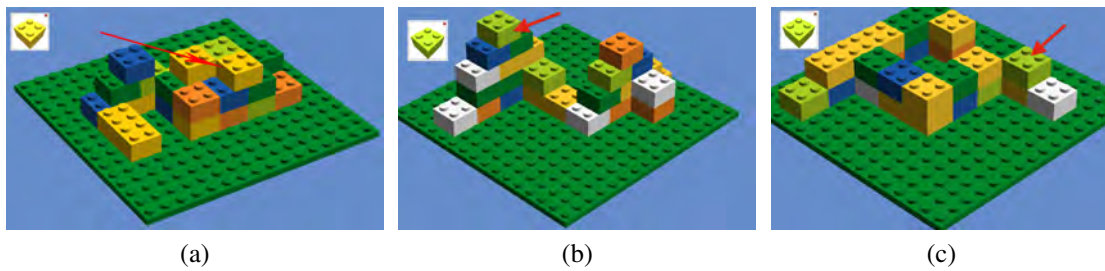
Figure 4.3: The assembly tasks used in the study. We used three different assembly tasks with equal complexity. The images depict the final step of the pictorial instructions.

4.3 Evaluation

We conducted a user study with workers with cognitive impairments to evaluate the usefulness and impact of different error feedback modalities. This section describes the user setup for the study, explains the procedure, and reports the results of the quantitative measures. Ethical approval for this study was given by the employee organization of our partner sheltered work organization and by the German Federal Ministry for Economic Affairs and Energy.

4.3.1 Design

To find the most suitable error feedback modality for communicating errors to workers with cognitive impairments, we designed our experiment following a repeated measures design with the used error feedback modality as the only independent variable. As dependent variables, we measure the TCT, the number of assembly errors, and the number of picking errors. We counterbalanced the order of the feedback modalities according to the Balanced Latin Square. Additionally, we collect qualitative feedback through semi-structured interviews and observations. We decided not to use a baseline condition that measures the time and errors without any error feedback as related approaches showed that error feedback is beneficial for workers with cognitive impairments. They usually ask their socio-educational instructors for feedback about the assembly [128, 136].

4.3.2 Apparatus

For our experiment, we use the system providing tactile, auditory, and visual error feedback described in the previous section. We configured it that only one feedback modality is active per condition. As previous work suggested, using Lego Duplo tasks provides a good abstraction of assembly tasks, which enables changing the complexity of tasks without changing the task itself [132, 136, 357]. Therefore we decided to use a Lego Duplo assembly task. As

we designed the experiment to have three conditions, we created three unique Lego Duplo assembly tasks consisting of 24 bricks per task[3]. The tasks are mainly inspired by the tasks that were used by Funk et al. [136]. Therefore, we used their 24 bricks task, increased their twelve bricks task by twelve more bricks, and used their 48 brick task and stopped at 24 bricks. The instructions in their final assembly state are depicted in Figure 4.3. We printed the instructions on a single-sided A4 sheet of paper in a way that one assembly step is printed on one sheet of paper. The upper left corner of the instruction depicts the brick that has to be picked from one of the eight picking bins. The assembly position is depicted in a way that the brick to assemble is shown in its final position. Further, a red arrow highlights the assembly position that it can be found immediately. Further, the instructions were designed in a way that the same brick does not have to be picked twice directly after the first occurrence.

4.3.3 Procedure

In preparation for the study, we asked for written consent from either the participants or their legal guardians before the study. As the study was conducted in our laboratory, which was a new environment for every participant, we initially made all participants familiar with our laboratory environment. Accompanied by their regular socio-educational instructors, we initially explained the assistive system and the three modalities that are used in the study. At first, we informed the participants that their participation in this study is voluntary. We told them that they should inform us whenever they felt unwell or uncomfortable, as we would immediately abort the study in this case. Afterward, we explained the intention of the study and why the tasks they perform are relevant. After explaining the course of the study, we made the participants familiar with the paper assembly instructions, *i.e.* which part to pick and where to assemble it. Once the participants felt confident using the paper assembly instructions, we introduced the error feedback modality for the current condition and explained what we count as an error and what the system will do if an error occurs. As the participants felt that they understood both error feedback and the paper assembly instructions, we began with the study and started measuring the TCT. During the study, three researchers were present at the scene. The first researcher triggered the assembly error feedback in case an assembly error was made as a Wizard of Oz (WoZ). Picking errors were detected using a Kinect v2, which triggered the error feedback automatically. The second researcher was measuring the TCT and counted the picking errors and assembly errors that were made for each condition. The third researcher was observing the worker reacting to the error feedback and taking subjective notes based on the observation. The error feedback was displayed immediately while performing the assembly when an assembly or picking error was triggered. After the assembly was completed, we asked the participant for their opinion about the used error feedback. Then we repeated the procedure for the other two remaining conditions. At the end of the third condition, we asked each participant for a subjective

---

[3] We provide the created assembly instructions to other researchers for reproducing the study:
`www.github.com/hcum/comparing-tactile-auditory-and-visual-assembly-error-feedback`
- last access 2020-01-28

rating which error feedback was perceived the best by her or him and we asked them why they preferred or disliked the feedback.

4.3.4 Participants

We invited 16 participants for our user study. The participants were aged from 34 to 53 years (M=40.33, SD=6.36) and were employees of a sheltered work organization working in manual manufacturing. All participants worked on manual assembly tasks daily. None of the participants were familiar with our system, the used task, or the error feedback modalities used in our study. We invited the participants according to their PI in a way that they represent the population of the sheltered work organization. Therefore, we used the PI of the sheltered work organization to categorize their capabilities. The PI is a percentage ranging from 0% to 100%, which indicates how capable a worker with cognitive impairments is of performing a work task. Inspired by previous work [128, 136], we divided the population to belong to one of three PI groups: 5-15%, 20-35% and above 40%. Accordingly, we invited 5 participants belonging to each of the three PI groups. The study took approximately 40 minutes per participant. Participants were monetary compensated through funding from a joint project.

4.3.5 Results

During our study, four participants belonging to the 5-15% PI group aborted the study as they did not want to wear the glove anymore. Therefore, we excluded these 4 participants from the quantitative evaluation.

We statistically compared the TCT, the number of assembly errors, and the number of picking errors, between the error feedback modalities using one-way repeated measures Analysis of Variance (ANOVA). Mauchly's test showed that the sphericity assumption was violated for the number of assembly errors ($\chi^2(2) = 6.852$, p = .033) and the number of picking errors ($\chi^2(2) = 9.201$, p = .010). Therefore, we used the Greenhouse-Geisser correction to adjust the degrees of freedom ($\epsilon = .668$ for the number of assembly errors and $\epsilon = .624$ for the number of picking errors). Further, we used a Bonferroni correction for all the posthoc tests.

Considering the TCT, the participants assembled the Lego Duplo construction fastest using the visual error feedback (M = 434.42 sec, SD =185.27 sec), followed by the auditory error feedback (M = 445.08 sec, SD = 197.40 sec), and the tactile error feedback (M = 575.42 sec, SD =276.71 sec). As the Shapiro-Wilk test did not show a non-normal distribution, we use a parametric one-way ANOVA. The one-way repeated measures ANOVA showed a statistically significant difference in the TCT between the feedback modalities $F(2,22) = 4.634$, $p = .021$. The posthoc tests reveal a significant difference between the visual and tactile feedback. The effect size estimate shows a large effect ($\eta^2 = .296$). Figure 4.4a shows a graphical overview of the results.
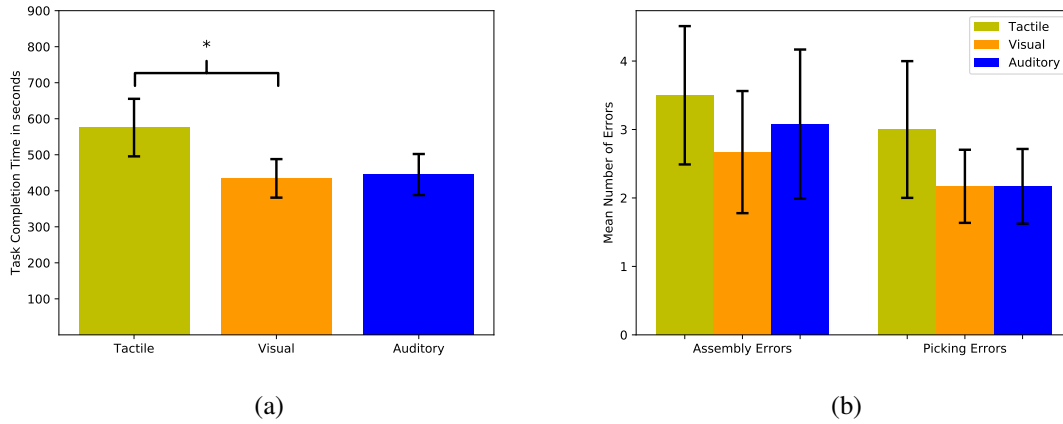
Figure 4.4: **(a):** An overview of the average Task Completion Time to assemble the Lego Duplo construction using the different error feedback modalities. **(b):** The average number of errors that were made using the different error feedback modalities for both assembly errors and picking errors. The error bars depict the standard error. The brackets indicate significant results.

Analyzing the number of assembly errors that were made by the participants using the different error feedback modalities, the visual error feedback resulted in the fewest assembly errors (M = 2.67, SD = 3.09), followed by the auditory error feedback (M = 3.08, SD = 3.77), and the tactile error feedback (M = 3.5, SD = 3.5). As a Shapiro-Wilk test showed a non-normal distribution for all three modalities (all $p < .05$), we use a non-parametric Friedman test. Accordingly, the Friedman test did not reveal a significant difference between the error feedback modalities ($p > .05$). The results are depicted in Figure 4.4b.

For the number of picking errors that were made using the different error feedback modalities, the visual error feedback (M = 2.17, SD = 1.85) and the auditory error feedback (M = 2.17, SD = 1.89) lead to the same number of picking errors. Using the tactile error feedback, the participants made the most picking errors (M = 3.0, SD = 3.46). As a Shapiro-Wilk test showed a non-normal distribution for all three modalities (all $p < .05$), we use a non-parametric Friedman test. However, the Friedman test could not reveal a significant difference ($p > .05$). The results are depicted in Figure 4.4b.

4.4 Qualitative Observation

Considering the qualitative observational feedback that was collected during the user study, two researchers continuously observed the participants during the study. The first researcher observed the interaction of the participants with different error feedback modalities and asked questions after the condition was finished in a semi-structured way. The second re-

searcher was performing an observational study for subjectively analyzing the effect of the different error feedback modalities on the participants.

When asked about the Lego Duplo task itself, a participant stated that *"with an increasing number of steps the task gets more complicated"* (P1). Another participant pointed out that the used task has several error sources as *"[he] had difficulties in distinguishing between the two different types of green bricks"* (P2). During the study, we subjectively observed that ten of our twelve participants had difficulties to recognize bricks, whereas most of them stated that they did not have problems differentiating between the different colors.

We also asked the participants about how easy it was to perceive the different feedback modalities. One participant liked the visual feedback as it is *"directly in the field of view"* (P2), others said that the visual feedback is *"easy to see"* (P4). The participants were unsure about the tactile feedback as *"the vibration is rather unpleasant during the work task"* (P2) and *"[it] distracted me in my workflow"* (P4). Some participants (P3, P8) also did not feel or pay attention to the tactile stimulus. Further, four participants had to abort the study after the tactile feedback condition. Two of these participants stated that they did not feel well wearing a glove, the other two were uncomfortable or scared of the tactile vibration. Considering the auditory feedback a participant stated that the *"sound was very easy to perceive"* (P2). On the other hand, three participants (P3, P8, P14) stated that they were distracted by the auditory error feedback as it *"scared [them] when it triggered"*.

Considering the privacy implications of the different error feedback modalities, some participants stated that auditory feedback would be *"not that good because others can hear when [I] made an error"* (P2, P13). On the other hand considering the visual error feedback many participants told us that *"[they] don't care if others could see that they made an error"* (P2, P4, P5, P12, P14). When we asked why the participants responded that *"usually the supervisors are watching the work steps as quality control and are telling us when an error occurred"*. (P4) One participant was concerned that his supervisor would not be able to see anymore when an error was made as only he could perceive the tactile feedback (P16).

One participant (P7) stated after the study that he perceived working with the system using any error feedback as very easy as he could completely relax and rely on the system because it would tell him when he makes an error. Two other participants stated that they *"could imagine using the visual feedback daily"* (P4, P5).

At the end of the semi-structured interview, we asked the participants to rank the feedback modalities according to which modality they liked best and to also consider which error feedback they would use daily. One of our 12 participants who finished the study did not want to rank the error feedback modalities. The results reveal that participants' subjective impression of the visual error feedback was best (5× first, 4× second, and 2× third) followed by the auditory feedback (4× first, 4× second, and 3× third). The tactile feedback was perceived as the worst (2× first, 3× second, and 6× third). The results are also depicted in Figure 4.5.

In addition to the interviews, five participants made comments on their performance or formulated their thoughts during the assembly task. In most cases, these participants showed a
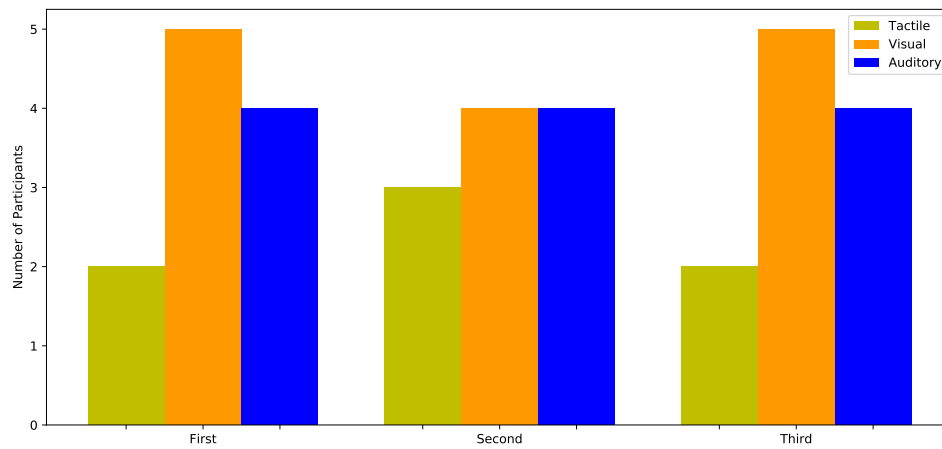
Figure 4.5: The subjective ranking of the feedback modalities ranked by the participants.

higher TCT than other participants. Another four participants needed to be guided verbally during their tasks and reacted positively to the verbal help.

4.5 Discussion and Implications

The results of our study reveal that both the number of assembly errors and the number of picking errors are not significantly different between the error feedback modalities. We argue, that the number of errors is not different between the error feedback modalities as the error feedback occurs after the participant made the error. Thus, we could not measure that the used feedback modality is influencing the participant in the way they are working, *e.g.* paying more attention to a correct picking or a correct assembly of the Lego Duplo construction or being insecure and making more errors due to an error feedback modality. However, the TCT is statistically different between tactile error feedback and visual error feedback. We assume that this difference in the TCT is caused by a faster error recovery after an error was made. The visual error feedback was perceived significantly faster than the error feedback provided by the tactile glove.

The qualitative observations and the answers we got from the participants through semi-structured interviews after using the different error feedback modalities tend towards using visual error feedback. Although some participants did not care about using error feedback that preserves their privacy at the workplace, auditory error feedback was perceived as distracting. Considering the tactile feedback using the vibrating glove, some participants were able to perceive the tactile feedback while others did not react to the tactile feedback at all.

While using privacy presuming error feedback mechanisms might be a good decision in general [131], the implications for using these error feedback mechanisms for workers are different. Through interviews, we found that errors that are made at the assembly workplace are considered non-private information, which is acceptable to communicate to supervisors. Compared to related work [131], which analyzed the error feedback for workers with no disabilities, the design implications are different as error privacy has a higher priority.

4.6   Study Conclusion

We compared visual, auditory, and tactile error-feedback for workers with cognitive impairments at manual assembly workplaces. In a user study, we obtained data from twelve participants to estimate the best feedback modality in terms of task completion time, measured errors, and qualitative feedback. We found a significant difference between visual and tactile error feedback regarding the task completion time, but could not find a significant effect between picking errors and assembly errors between each error feedback modality. We did not observe that altering the error feedback modality had an impact on the way the participants worked on assembly tasks. Therefore, the number of assembly errors throughout all error feedback modalities were not significantly different. However, we measured a significantly faster task completion time using visual error feedback compared to using tactile error feedback. Using visual error feedback, the participants were able to recover faster from errors than using tactile error feedback.

4.7   Chapter Summary

In this chapter, we investigated the efficiency of visual, auditory, and tactile feedback modalities during a Lego Duplo assembly task. Awareness about the current assembly and work context enables us to tailor feedback modalities for collaborative (*i.e.*, using visual or auditory feedback) or privacy-protective settings (*i.e.*, using tactile feedback). The presented study investigated the task completion time, number of errors, and user acceptance by using the three feedback modalities. We find a strong preference for visual feedback regarding acceptance and assembly efficiency followed by auditory and tactile feedback (**RQ2**). The study shows that the employed feedback modality plays an important role that has to be adjusted to the contextual and individual needs. While tactile feedback showed the lowest preference and performance, it was more preferred when focusing on privacy-related aspects. This trade-off between efficiency and privacy has to be considered during the development of workload-aware systems and interfaces.

III

SENSING COGNITIVE WORKLOAD

Chapter 5

Mobile Electroencephalography

*Humans read minds. We constantly try
to understand what our fellow humans
are thinking and feeling – and how they
are going to act.*

Hank Greely

In Chapter 3 and Chapter 4 we outlined the design requirements of workload-aware interfaces. Our results show that, regardless of the current chore, people seek assistance when their short-term memory is demanded frequently. The qualitative inquiry presented in Chapter 3 revealed that cognitive offloading is desired during repetitive tasks that are subject to small changes. This requires us to update the learned task while avoiding careless mistakes from the old execution pattern, and incorporate respective changes in task execution routines. Manual assembly at production lines is an example of such routine work. Modern manufacturing processes are increasingly defined by smaller lot sizes of bespoke designs. Gone are the days that require workers to act purely on rote learning. Instead, novel assembly instructions must frequently be committed to memory as soon as new designs and components enter into the production pipeline [127]. Thus, assembly is increasingly defined by its cognitive instead of physical demands. Manual assembly places workload on cognitive processes that underlie executive working memory [15, 144]. Given that cognitive workload impacts the individual task performance [113, 169], it follows that profit margins may suffer as a result of reduced production throughput and error-prone manufacturing.

Assembly instruction systems have been introduced to ameliorate this. Such assistance consists of instructions on printed paper, external displays, or *in-situ* instructions projected directly on the assembly workplace. Such systems are designed to simultaneously reduce cognitive demand while optimizing performance. Indeed, previous research has suggested that design manipulations of the visual representation and modality of assembly instructions can have a significant influence on self-reported measures of cognitive workload [127, 217].

Figure 5.1: (a) Paper-based and **(b)** projected *in-situ* instructions at a manual assembly workplace. An EEG headset measures the level of working memory. This provides objective insights about the working memory load that is placed during manual assembly.

Strong arguments have been made for the use of *in-situ* instructions that can be integrated "just-in-time" [61, 134]. Furthermore, visual *in-situ* instructions are attributed to a significant increase in production efficiency [138, 228].

Novel user interfaces – for instance, assistive instruction systems – are often evaluated for their improvements regarding the performance or subjective workload to justify their implementation over the *status quo*. Performance is typically assessed in terms of error rate and assembly completion time, while cognitive workload is often assessed using self-rated measures or semi-structured interviews [129]. Commonly used self-rated measures include the NASA-TLX [161] or the simpler RSME [45, 402]. However, these metrics are susceptible to individual differences in subjective reporting. For instance, extroverted confident individuals might be less likely to indicate workload to the same degree as introspective modest individuals [187]. Furthermore, novel assistive systems and visualizations might introduce a novelty effect. This means, that participants provide self-reported scores that are influenced by the excitement over a new technology [384]. Additionally, subjective reports can only be performed after test completion and rely on cognitive processes, namely working memory. Such methods are limited in their objective and real-time assessment of cognitive workload when evaluating assistive technologies.

Therefore, physiological methods are increasingly employed as a means to estimate mental states, such as cognitive workload [156, 223, 226]. For example, BCIs [7, 367] have been used successfully to assess the complexity of presented information in a variety of scenarios [7, 225] (see Figure 5.1). A BCI records brain activity in real-time and computationally derives estimates of a targeted mental state to either provide feedback about the physiological state of the user or enable device control through neuronal activity. Increasingly, advances in brain-recording technology and neuroscience findings contribute towards a vision of ubiquitous BCI deployment in everyday places of work and play [53, 198, 371]. In the current

context, a BCI for cognitive workload sensing could be used in collaborative work settings by allocating tasks to workers who are less fatigued [121, 129] or provide adaptive assistance whenever necessary [130]. However, for this future to be realized, it is necessary to ensure that recorded estimates are robust and valid for its given setting.

BCIs may record brain activity via EEG, which has been preferred for its lightweight and mobile hardware, high temporal resolution, and non-invasive nature [152]. EEG measures neuronal signals from scalp electrodes, relative to a reference electrode [118, 148]. Previous research has utilized EEG as a measure for cognitive workload to evaluate task difficulties in real-time [145, 146, 350]. To some extent, it is possible to rely on EEG to classify cognitive workload in real-time, particularly those that involve working memory processes [152]. This makes EEG to a viable alternative or complementary metric for mental workload measures.

In this chapter, we present a method to assess cognitive workload using mobile electroencephalography. EEG facilitates the measurement of cortical activity which represents markers for cognitive effort [145, 146]. We present a study that compares two assistive systems for manual assembly which have shown significant differences in subjective perception. This is complemented by the presentation of a design pipeline that suggests the use of EEG as a complementary measure alongside subjective assessments. Henceforth, the evaluation of EEG cognitive workload measures yields the potential for real-time adjustments of user interfaces. We seek to answer the following research question:

→ **RQ3**: Does electroencephalography provide measures of cognitive workload for user interface evaluation?

This section is based on the following publications:

- Thomas Kosch, Markus Funk, Albrecht Schmidt, and Lewis L. Chuang. Identifying Cognitive Assistance with Mobile Electroencephalography: A Case Study with In-Situ Projections for Manual Assembly. *Proc. ACM Hum.-Comput. Interact.*, 2(EICS):11:1–11:20, June 2018

- Thomas Kosch and Lewis L. Chuang. Investigating the Impact of Assistive Technologies on Working Memory Load in Manual Assembly through Electroencephalography. *Human Neuroscience Archive*, (114)

5.1 Background

Previous research has investigated effort in examining EEG data while subjects perform specific tasks. This section summarizes relevant research regarding the use of EEG as a non-invasive method to quantify cognition and presents assistive systems that adapt to the measured brain activity.

5.1.1 Quantifying Cognition using EEG

Much research has been performed to define and quantify cognitive workload. Sweller et al. [353, 354] defined three key components of cognitive workload comprising intrinsic, germane, and extraneous cognitive workload. Intrinsic workload describes the inherent complexity of the task itself, and can therefore not be manipulated by external sources. Germane workload describes the cognitive effort subjects need to comprehend and process new information. Extraneous workload describes the cognitive demand to understand and process the visual representation of the underlying information. Experiments often evaluate extraneous workload by influencing the visualization of information. Intrinsic and germane workload is not easy to manipulate since they are task-related or depend on individual mental capabilities.

More precisely, cognitive workload defines the mental effort being used in working memory, which is responsible for fast information processing and limited within its capacities [15]. Gevins et al. [146] invested effort to quantify the occurrence and amount of working memory using EEG. They found a decrease in alpha frequencies and an increase in theta frequencies during cognitively demanding tasks. Jensen et al. [181] replicated their results using a memory demanding task. Scharinger et al. [330] also showed a variety of other tasks that correlate with increased working memory. However, working memory strongly depends on individual attributes, such as age or neuronal health state [199].

Machine learning became popular to classify mental states using EEG data [253, 251, 252] and has been used to estimate workload in different contexts [273]. Lee et al. [238] collected EEG data of participants to classify cognitive demanding tasks according to their difficulty. Participants were asked to do different cognitive tasks, including mental rotation and mental arithmetic. Instead of using previously recorded data associated with similar tasks [6, 192], data was collected from scratch. Using machine learning, a classification accuracy of 92.4% was achieved. However, the experiment has not focused on classifying working memory itself. An approach to classifying working memory was undertaken by Grimes et al. [152]. Participants performed a *N*-back task, which is a popular task to demand short-term memory [266, 291]. Their study included four different difficulties, where their results yielded a classification accuracy of 99% for binary classification and an accuracy of 88% for classification between four classes of difficulty. Nonetheless, a medical EEG headset was utilized for their experiment, which is hard to deploy on the fly. Furthermore, an artificial task was used to measure the level of working memory instead of using a real-world scenario.

5.1.2 Evaluating and Adapting User Interfaces

Besides using EEG for quantifying and classifying working memory researchers have employed this measurement to enable the development of cognition-aware systems. Zander et al. [397] investigated how passive brain activity measurements can be leveraged to handle take-over tasks during autonomous driving. Prinzel et al. [237] researched how task alloca-
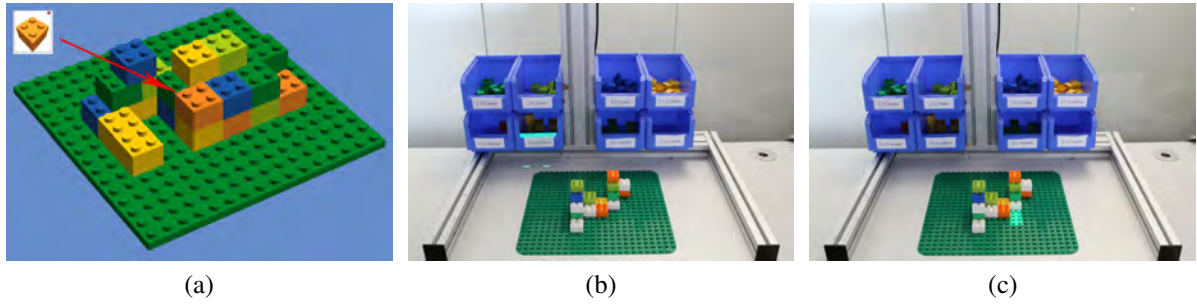
Figure 5.2: **(a):** Snippet from the used Lego Duplo paper instruction. The next brick can be seen in the upper left corner. A red arrow denotes the final assembly position of the brick. **(b):** Highlighted box using *in-situ* projections. **(c):** Projection of the target position of a selected brick.

tion can be carried out efficiently using EEG-related metrics. Finally, El-Komy et al. [109] used physiological sensing devices in an assembly environment comprising additional artificial tasks to measure the workers' stress levels. However, only results from the emotion and arousal scores retrieved by the integrated BCI interface were reported.

To the best of our knowledge, no prior work concerning the evaluation of EEG to derive cognitive workload during manual assembly and with different instruction systems has been done. In this work, we close this gap by utilizing standardized manual assembly tasks with two different instruction systems. We record and compare EEG data in a repeated measure experimental design to evaluate EEG as a valid assessment tool for cognitive workload induced by assistive technologies during manual assembly.

5.2 Assembly Instruction Systems

We identified two different instruction systems from related research, which are different regarding subjective perception [134]. Within both instruction systems, Lego Duplo bricks are assembled. This task resembles a full replacement for a real assembly task and enables to change the complexity of the task without changing the task itself [134, 228].

5.2.1 Assembly Instruction Visualizations

Informed by related work, two instruction visualizations are identified which differ in overall interpretation complexity [134]. We have chosen paper instructions, as they represent the current state of the art when it comes to transferring assembly instructions in manual assembly lines [132]. This is compared to projected *in-situ* instructions, where assembly instructions are projected in the workplace. We have chosen these two assembly instruction modalities as subjective measures suggest an alleviation of workload for *in-situ* projected

instructions compared to printed paper instructions [134, 138]. In the following paragraphs, we describe both instruction systems used in the study in detail.

Paper Instructions

We printed single-sided instructions on an A4 sheet of paper. Each work step was printed on a single page, such that the position and size of every step were the same. The paper instructions were put together in the correct order using a folder and positioned to the left or right of the users, depending on their handedness. The folded assembly instruction remained the same position relative to the user's position. The instruction shows a brick on the upper left corner which is required to select for completing the current work step. Marked by an arrow, the assembly instruction shows the final position of the brick (see Figure 5.2a).

In-Situ Instructions

We compare the presented paper instructions with *in-situ* projections displaying assembly instructions. We use a similar system as shown by Funk et al. [137]. A projector mounted above the work table displays the next assembly step on the workspace. A Kinect v2[1] validates each work step of the Lego Duplo construction. This includes the verification of correct item selections from bins by observing the hand movements of the participant (see Figure 5.2b) and assembly steps by comparing the (see Figure 5.2c). The next work step is displayed when the current work step is performed correctly. The system is waiting until the current work step is carried out correctly and does not proceed if the user makes an error.

5.3 Manipulating Cognitive Workload

We manipulate working memory to assess the validity of our setup. We use a visual N-back task [266] with two levels of task difficulty ($N=0$ and $N=2$) to induce cognitive workload, namely executive working memory. In the N-back task a series of numbers is presented (*i.e.*, numbers). Each symbol appears in a fixed position. Upon symbol representation, participants have to decide if the current symbol is equal to the symbol shown N steps ago. Participants have to keep a sequence of N symbols constantly in their memory, decide if a match occurred, and then update the sequence in their memory.

For example, the *0*-back task requires participants to compare each displayed symbol with the first one seen in the series. Since the currently displayed number matches with the first one in the series during the *0*-back task, no memory updates are required. However, the task difficulty can be manipulated by changing N [145, 152, 291]. For $N=2$, participants have to memorize the last two symbols in the series while paying attention to matches when the currently displayed symbol is the same as shown two items ago. Table 5.1 shows an example of the N-back.

---

[1] `https://developer.microsoft.com/en-us/windows/kinect` - last access 2020-01-28

Displayed number	5	8	3	4	3	9	1
Press button (*0*-back task)	5	8	3	4	3	9	1
Press button (*1*-back task)		5	8	3	4	3	9
Press button (*2*-back task)			5	8	3	4	3
Press button (*3*-back task)				5	8	3	4

Table 5.1: Example of the *N*-back task task. Participants have to confirm by a button press, that the currently displayed number matches with the number seen *N* numbers ago.

The motoric requirements remain constant across difficulty levels during the *N*-back task. This enables comparisons between different difficulties as working memory load is measured while excluding reactions from external stimuli.

A smartphone app[2] is used to display a single matching *N*-back task. Throughout the experiment, we use a Nexus 5X[3] with a screen size of 5.2 inches to run the *N*-back trials. The displayed symbols ranged between 0 and 9, which appeared in random order. Each number is displayed in the center of the smartphone screen for one second. Afterward, the screen remains blank for 2.5 seconds before the next number appears. Participants have to press a button on the screen if a match occurs.

5.4 EEG as Indicator for Workload

In the following study, we assess the validity of our setup by conducting two *N*-back tasks with different complexities as described before. Afterwards, we conduct two assembly tasks per participant, each with the two previously mentioned assembly instruction systems. The overall hypothesis is that *in-situ* projections induce less cognitive workload than printed paper instructions. This leads to the following hypotheses:

- **H1:** Projected *in-situ* instructions will induce *higher alpha power*, relative to printed instructions.

- **H2:** Projected *in-situ* instructions will produce *lower scores of subjective workload*, relative to printed instructions.

- **H3:** Projected *in-situ* instructions will produce *fewer item selection errors*, relative to printed instructions.

- **H4:** Projected *in-situ* instructions will produce *fewer assembly errors*, relative to printed instructions.

---

[2] `www.play.google.com/store/apps/details?id=cz.wie.p.nback` - last access 2020-01-28

[3] `www.gsmarena.com/lg_nexus_5x-7556.php` - last access 2020-01-28

Figure 5.3: (a): The Emotiv Epoc wireless EEG headset featuring 16 electrodes including two reference electrodes. **(b):** Electrode placement layout of the 14 measurement electrodes [386].

- **H5:** Projected *in-situ* instructions will produce *faster completion times*, relative to printed instructions.

5.4.1 Methodology and Measures

We used the Emotiv Epoc as brain-sensing device during the whole experiment (see Figure 5.3a and Figure 5.3b). Since the alpha band varies between participants, we conducted an eyes opened and eyes closed task to estimate the individual alpha band [20]. The overall duration of this trial was one minute. The participants started with their eyes opened and were verbally instructed to close their eyes after 30 seconds to provoke a sudden increase in alpha power [20]. The participants kept their eyes closed for another 30 seconds. We use this peak as a reference point for extracting alpha power for later analysis. Furthermore, we took ±2 Hz around the peak frequency as a measure for the alpha band.

We continue then with the *N*-back task to verify the validity of our setup and accuracy of our EEG measures. Participants start with a *0*-back task to be induced with low workload, followed by a *2*-back task to be induced with high workload. The 2-back induces sufficient complexity to make the differences in working memory between resting and *N*-back task visible in EEG data [47]. A NASA-TLX questionnaire is filled out after each *N*-back condition to collect subjectively perceived workload.

We begin with the assembly task afterward. Inspired by several reference tasks proposed by previous research [132], we use a Lego Duplo task to evaluate the paper and projected

in-situ instructions in terms of measured workload. We prepared two different assembly instructions, where each of them is modeled as paper instruction and projection as shown in Figure 5.2. For the assembly tasks, we used a repeated-measures experimental design with the instruction visualization as a single factor including the levels paper and *in-situ*. We counterbalanced the order of conditions according to the balanced Latin square. As identified in the previous work [35, 59, 134, 328], we measure the number of errors and the task completion time per trial. The number of errors was divided into item selection and assembly errors. An item selection error is counted whenever participants put their hands into a box where incorrect bricks reside. An assembly error is counted when a brick is assembled in a wrong position. To reduce the number of head movements, we seated the participants before the assembly experiment at a comfortable distance to the assembly setup, so that boxes and the assembly plate can be reached with minimal effort. Participants filled out a NASA-TLX questionnaire after each trial to provide their subjectively perceived workload during the last assembly condition. Ultimately, we asked the participants verbally about their preference between both assembly instruction systems and noted their answers for later analysis.

5.4.2 Procedure

Participants signed a consent form and provided their demographics after we explained the course of research to them. We put the BCI on the participant's head and ensured good connectivity between the scalp and all electrodes. We explained to the participants what EEG signals and noisy artifacts are. We asked participants to keep their head as still as possible and to avoid unnecessary eye blinks.

Participants started with a one-minute baseline resting task. Participants kept their eyes open for 30 seconds. After the first 30 seconds elapsed, participants were instructed to close their eyes for another 30 seconds. The study continued with two *N*-back tasks, each comprising 20 numbers. The total runtime of each *N*-back was one minute and ten seconds. The experiment started with the *0*-back task. Participants had to press the match button every time they saw a number since the currently displayed numbers refer to the same number. Participants continued with the *2*-back task. The participants had to press the match button whenever the currently displayed number was equal to the number shown two numbers back. We recorded EEG data during all tasks. Participants were asked to fill out a NASA-TLX questionnaire after each *N*-back task.

After the verification procedure, participants started with the assembly procedure. They began with either paper instructions or projected *in-situ* instructions based on the order of the balanced Latin square. Additionally, we shuffled the order of the Lego Duplo instructions itself between the conditions. During the assembly, we recorded raw EEG data, counted the number of errors separated into item selection or assembly errors, and the time used for assembly. During all conditions, participants were instructed to keep their head still and avoid unnecessary eye blinks. After every assembly, participants were asked to fill out

a NASA-TLX questionnaire to assess the perceived the workload during the usage of the given instruction system. Qualitative comments about the preference of users regarding the instruction systems were collected in the end.

5.4.3 Data Processing

We apply the following data processing procedure: Data is filtered using a spatio-spectral decomposition method [279] with filter thresholds between 0.5 Hz and 20 Hz to remove unwanted frequencies caused by eye blinks or head movements. We average all 14 channels by calculating the element-wise mean of the signals. We remove the first and last four seconds of the signal to avoid unwanted artifacts caused by the beginning and end of the trial [238]. We divide the signal into one-second slices with an overlap of half a second. Instead of extracting the alpha band between 8 Hz and 12 Hz, we determine the maximum peak during the eyes opened and eyes closed task for each individual [117, 199]. The power spectra around the maximum peak (± 2 Hz) is averaged and used as individual alpha power.

5.4.4 Results

We recruited twelve participants over our university mailing lists (8 male, 4 female). All participants were students and had a normal or corrected-to-normal vision. None of the participants were affected by neurological disorders. The mean age was 23 years (SD = 2.22). Participants were compensated with 10 Euro for their participation.

Executive Working Memory Load and EEG

We identified alpha desynchronization as features that corresponded with an increased load in executive working memory. The alpha band is known to vary across individuals [100]. Therefore, we determined an individual alpha frequency (IAF) bandwidth for each participant based on their peak frequency value given their baseline EEG. We defined ± 2 Hz around the peak frequency as the individual alpha band. Overall, the mean IAF band ranged between 6.5 Hz and 10.5 Hz (SD = 2.14).

The mean power of the IAF power for each participant was submitted to a two-tailed paired-samples t-test for the factor *executive working memory load* consisting of the *0*-back and *2*-back task. A Shapiro-Wilk test confirmed that the data was normally distributed. The results reveal a significant difference, $t(11) = 5.90$, $p < 0.001$, $d = 1.70$, between the *0*-back and *2*-back task. Figure 5.4a shows the direct comparison of alpha activation between both conditions. To summarize, our chosen EEG feature was sensitive to load in an executive working memory task, namely, there was less power in the IAF band when there was a higher load in this task. Thus, we report a large effect size in support of H1.
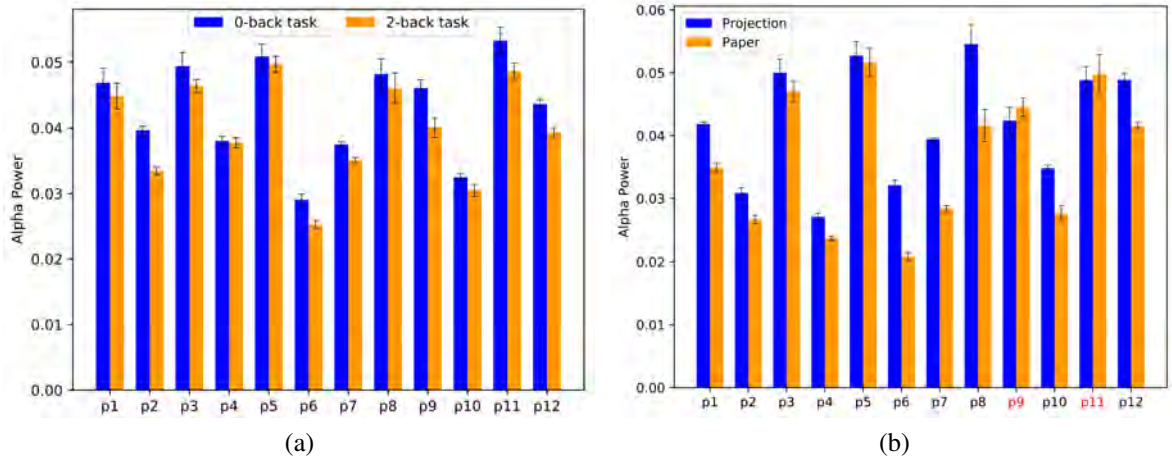
Figure 5.4: (a): IAF power for working memory load, *i.e.*, *0*-back, and *2*-back. The IAF power is higher for lower working memory load. The error bars depict the standard error of the sample mean. **(b):** Participant-wise comparison of IAF power when using *in-situ* projections and paper instructions. Except p9 and p11, the use of *in-situ* instructions results in higher IAF power compared to paper instructions. The error bars depict the standard error of the sample mean.

Assembly Performance and Alpha Power

The IAF power of each participant was submitted to a two-tailed paired-samples t-test for the factor *haptic assembly instructions* consisting of paper and *in-situ* instructions after a Shapiro-Wilk test confirmed that the data was normally distributed. The results reveal a significant difference, $t(11) = 3.86$, $p = 0.003$, $d = 1.12$. Figure 5.4b shows the mean alpha activation per condition and participant. Similar to our findings with the executive working memory load, we find that IAF power is generally lower for paper instructions compared to when participants experienced projected *in-situ* instructions instead—note that this was not true for p9 and p11 at the individual level. Thus, we report a large effect size in support of H2. More importantly, we show that the IAF power responds for lower executive working memory load as it does for our projected *in-situ* instructions.

Additionally, we statistically compare the number of errors made by the participants during assembly as well as the time they required to finish the assembly between the different conditions. We classify errors into item selection errors and assembly errors. Overall, participants did in average 2.25 (SD = 2.301) item selection errors during the paper instruction condition and 0.083 (SD = 0.289) errors during the *in-situ* condition. 0.917 (SD = 1.379) assembly errors were performed when using paper instructions and 0.25 (SD = 0.452) during the *in-situ* condition. A Shapiro-Wilk test did not show a normal distribution for the item selection and assembly errors. Thus, we have conducted a Wilcoxon signed-rank test. A significant difference was found for the number of item selection errors, $p = 0.009$. However, no significant difference was found for the number of assembly errors, $p = 0.188$. Figure 5.5a compares the number of item selection and assembly errors between both instruction systems.
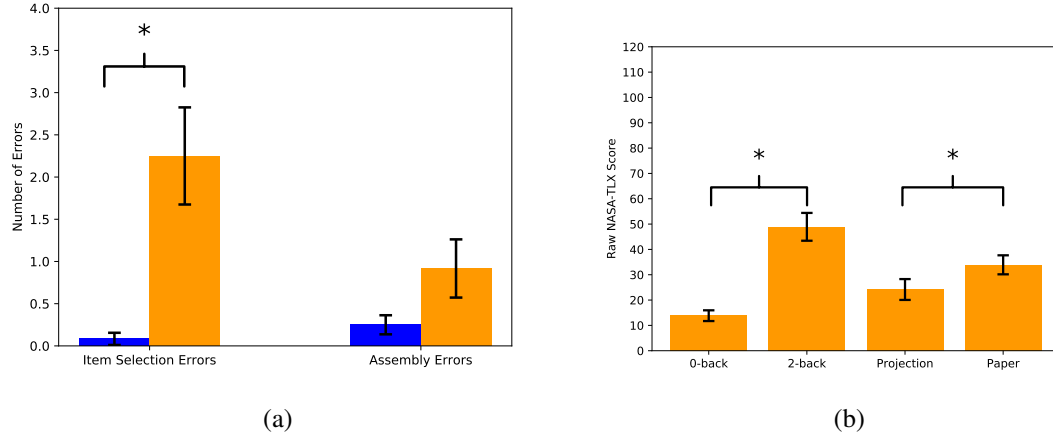
Figure 5.5: (a): Mean item selection and assembly errors. The error bars indicate the standard error. Brackets indicate significant differences. **(b):** Mean raw NASA-TLX scores per condition. The error bars depict the standard error. Brackets indicate significant differences.

The task completion time averages to 217.08 seconds (SD = 40.31) for paper instructions and 124 seconds (SD = 13) for projected *in-situ* instructions. The task completion time shows a significant effect between both conditions regarding task completion time, $t(11) = -8.82$, $p = 0.001$, $d = -2.55$.

Qualitative Assessment

We statistically analyze the subjectively perceived workload using the collected NASA-TLX questionnaires. We average the scores received from the six Likert scales per condition to calculate the mean raw NASA-TLX scores. The averaged NASA-TLX score amounts to 13.83 (SD = 7.34) for the *0*-back task, 48.92 (SD = 19.01) for the *2*-back task, 33.92 (SD = 13) for the assembly task using paper instructions, and 24.17 (SD = 14.26) during assembly using projected *in-situ* projections.

A Shapiro-Wilk test confirmed that the data is normally distributed. A t-test shows a statistical difference in NASA-TLX scores between the *0*-back and *2*-back condition, $t(11) = -6.41$, $p = 0.001$, $d = -1.85$. Therefore, more workload was subjectively perceived in the *2*-back task than in the *0*-back task. A significant difference between paper and *in-situ* instructions, $t(11) = -3.86$, $p = 0.003$, $d = -1.12$, was found. Therefore, lower alpha activation was measured during the *in-situ* conditions compared to the printed paper instruction conditions. Figure 5.5b shows the mean NASA-TLX scores.

Additionally, we collected comments from the participants by asking for their preference regarding the assembly in the instruction system. Most participants provided us with positive feedback. For example, one participant mentioned that *"[. . .] projected instructions were easier and faster to understand than paper instructions. The light was good guidance."* (P1, similar to P2, P3, P7). Another participant mentioned that *"[he] felt that I was faster*

using projections since I did not have to flip the paper. Having both hands free enhanced the overall assembly." (P5). Finally, one participant mentioned that *"It was easier to follow the light than to follow the paper instructions itself. However, I felt like a robot during assembly."* (P4).

However, some participants stated that *"[...] the higher assembly speed using projections was stressful. Paper instructions provided a short relieve in workload when flipping the page."* (P9 similar as P11, P12) or that they were *"[...] unused to it, but [I] could familiarize with it after some time."* (P3, similar to P6, P9).

It is interesting to note that P9 and P11 perceived the assembly as stressful due to its fast pace. Both participants also show less alpha activation during the projected *in-situ* condition than in the paper condition (see Figure 5.4b). Instructions were provided immediately by the *in-situ* instruction system, which was perceived as stressful two participants (P9, P11). Paper instructions provided cognitive alleviation when flipping the page to mentally prepare for the next work step. Previous research also supports a positive correlation between stress and cognitive workload [283, 342]. The discomfort of the provided assembly instruction system can also be elicited from our results.

5.4.5 Assessing Cognitive Workload in Real-Time

Our results show a significant difference in alpha activation between paper and projected *in-situ* instructions. Using the collected data, we analyze the real-time applicability of EEG for cognitive workload assessment. We achieved this by calculating the alpha fluctuation over time for both conditions.

Similar to the previous analysis steps, we preprocess the data using the spatio-spectral decomposition method [279] with filter thresholds between 0.5 Hz and 20 Hz and calculate the mean of all channels. We remove the first and last four seconds of the signal to remove unwanted response artifacts. We calculate the individual alpha band per participant and averaged the alpha power overall participants for both conditions. Figure 5.6 shows the alpha fluctuation for both conditions.

For the paper instruction, there is an increase followed by a decrease of alpha activation over time for the paper instructions. Workload starts to differentiate with time when information from paper instructions have to be held continuously in the short-term memory. This could be a reason for the decrease of alpha power with time.

Projected *in-situ* instructions show a decrease followed by an increase in alpha power. We assume, that the novelty of the system for the participants is responsible for the decrease in alpha power at the beginning of the condition. Alpha power increases when the participants become familiar with the system and less information has to be kept in the short-term memory. A Shapiro-Wilk test showed a non-normal distribution. A Wilcoxon signed-rank test resulted in a statistical difference between both conditions, $p = 0.001$.
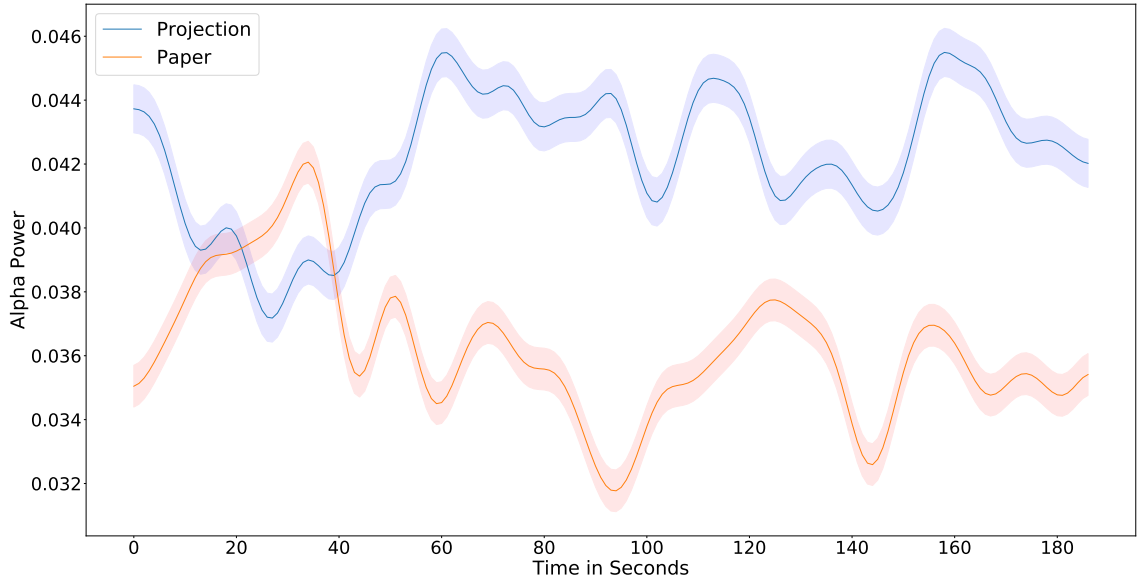
Figure 5.6: Mean IAF power fluctuation of all participants across time. Alpha power increases when using projected *in-situ* instructions and decreases for paper instructions. The shadowed area describes the standard deviation of each data point per participant.

5.5 Discussion

We evaluate the viability of a commercially available EEG device for estimating the mental workload of two different instruction systems to augment a manual assembly task. We discuss the validity of our hypotheses and present a framework to evaluate interactive assistive technologies based on EEG measures.

5.5.1 Validating the EEG Setup prior to the Experiment

To validate the correctness of our EEG setup and measures, we manipulated the difficulty of a *N*-back task to vary working memory load and investigated its impact on EEG measurements. Specifically, we focused on alpha power in an individually determined frequency bandwidth, proximal to 10 Hz. Previous research has consistently demonstrated lower alpha power in participants who experience high working memory load compared to low working memory load [146]. Thus, we expected lower alpha power during the *2*-back task compared to a *0*-back task. Indeed, we found lower alpha power for higher task difficulty and higher alpha power for lower task difficulty during the *N*-back conditions. In other words, participants experienced less cognitive workload during the *0*-back task compared to the *2*-back task. This supports the applicability of using commercial devices to infer cognitive workload and converges with the results found in previous work [238, 152].

Before evaluating novel interactive assistive systems using EEG, our findings suggest to conduct an eyes opened and eyes closed task to elicit the individual's alpha peak. Afterward, a N-back task with $N = 0$ and $N > 1$ should be conducted to infer the validity of EEG measures in terms of lower alpha activation for higher task complexities. The individual alpha peak is elicited when participants close their eyes [199] enabling an individual IAF analysis for upcoming EEG trials.

5.5.2 Evaluating Assembly Instruction Systems

Statistical comparisons of alpha power confirmed that projected *in-situ* instructions generated significantly higher alpha power than the use of printed paper instructions. In other words, the projected *in-situ* instructions induced less cognitive workload. Thus, H1 is statistically supported. This finding that is based on EEG measurements agrees with the subjective self-reporting measures based on NASA-TLX questionnaires [129]. A significant difference in subjectively perceived workload was found between paper and projected *in-situ* instructions (H2). In general, our results support the idea that the alpha-band frequency power of EEG measurements is a valid metric for estimating the levels of cognitive workload induced by a specific instruction system or visualization. Altogether, this agrees with the motivation of *in-situ* projections, which is to reduce cognitive load by providing situated information at the appropriate times.

Regarding assembly performance, we can partially confirm the outcomes of previous research [134]. We found a significant difference in the number of item selection errors. We confirm that fewer item selection errors were observed using projected in-situ instructions compared to printed paper instructions (H3). However, no significant difference between both instructions systems was found on assembly errors. Therefore, we cannot confirm H4. Assembling with projected *in-situ* instructions takes significantly less time than printed paper instructions. This agrees with previous research [134] and supports our final hypothesis (H5).

Our results encourage to measure and evaluate EEG to assess the mental demand of assistive technologies in manual assembly processes. However, the experiment has taken place under controlled conditions in lab environments where participants were restricted to a reduced number of eye blinks and head movements. We recommend the evaluation of EEG to evaluate assistive technologies in controlled environments before they will be deployed in real-world scenarios or for further scientific research inferring the mental demand during usage. This way, novelty biases can be detected by comparing correlations between EEG measures and self-rated assessments. System architects benefit from these evaluation steps as the objectively measured workload can be considered into the design pipeline.

| Individual Features (e.g. individual alpha band via eyes opened/closed task) | EEG Verification (e.g. N-back task with $N = 0$ and $N = 2$) | Experiment Execution (e.g. measuring EEG with different instruction visualizations) | EEG Analysis (e.g. analyzing alpha band or event-related potentials) | System Comparison (e.g. comparing individual alpha bands by conditions) |
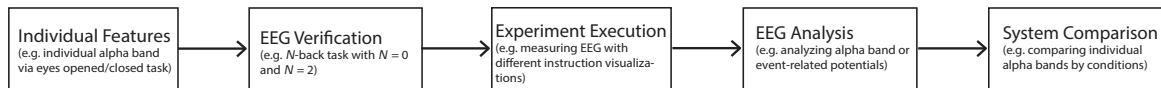
Figure 5.7: Experimental protocol which enables system designers to evaluate mental demanding elements in their user interface design on a cognitive level.

5.5.3 EEG as Real-Time Evaluation Tool

Additionally, we investigated the applicability of EEG as a real-time evaluation tool for detecting working memory. Previous research used medical EEG systems to derive the current level of executive working memory in real-time [29, 263, 275]. We show that differences in cognitive workload are measurable for two different instruction visualization systems using a mobile consumer EEG headset.

Figure 5.6 shows, how the level of alpha power changes with time during assembly. The alpha power for paper instructions decreases with time as the current bin for an item selection has to be recalled with every assembly step. Furthermore, the final position of the brick has to be recognized and placed correctly according to the paper instruction.

Projected *in-situ* instructions show the opposite effect. After a decrease in alpha power, which could be attributed to a lack of familiarity with the system, an increase in alpha power is observed. This can be due to the alleviation of working memory load since projected instructions eliminate the need to maintain a set of manual assembly instructions in working memory. Projected instructions are updated by each step of manual assembly and at the appropriate time steps. Thus, unlike with the use of paper instructions, it is no longer necessary to maintain and recall instructions from working memory.

The stability in alpha fluctuation for both instruction systems support the use of commercial EEG devices for real-time workload estimation. This agrees with the subjective perception of workload through NASA-TLX questionnaires and verbal feedback provided by the participants. A real-time system for estimating cognitive load could benefit use-case scenarios such as in evaluating user interfaces, assessment of workload in safety-critical tasks, or real-time adaptation of user interfaces suited to the current level of cognitive workload. By detecting constant low or high alpha fluctuations the provided assistance can be adjusted depending on the measured workload levels. Therefore, we present an experimental protocol that evaluates the cognitive demand assistive systems require during runtime (see Figure 5.7). Designers can use this protocol as an assessment template to find distracting or cognitive demanding user interface elements.

Novel engineered assistive technologies can benefit from real-time insights into the mental resources by an operator. This enables user interface designers to test visualization adaptations for different measures of cognitive workload. However, the question of how to provide user interface adaptation for different levels of cognitive workload remains and depends highly on the use case scenario in which assistive technologies are deployed.

5.5.4   Other EEG Metrics for Evaluating Mental Workload

We relied on an established EEG metric (*i.e.*, alpha power) to confirm previous claims that projected *in-situ* instructions can reduce working memory. Frequency domain measures of cognitive load are well-established and proved to be viable [144, 145]. However, frequency-based measures often lack the discrimination of functional interpretations, afforded by time-domain measures. Event-related potentials (ERPs) [305] refer to the time-domain waveform that is contingent upon the occurrence of a critical event such as the presentation of a stimulus.

Unlike frequency domain measures, the time-varying EEG activity of subsequent voltage deflections can be individually attributed to functional mechanisms that underlie information processing. For example, an early negative deflection in the ERP waveform around 100–200 ms is often associated with stimulus detection while a later positive deflection between 250–600 ms is associated with stimulus recognition and working memory updating [292]. This allows for stronger functional discriminability in the type of cognitive work that is experienced by the user, compared to the all-encompassing term 'mental workload'. Signal processing and classification approaches have been proposed to utilize ERP features for mental workload classification [47, 191].

The use of ERPs for estimating mental workload is limited by the fact that ERPs are event-related. This means that ERPs can only be extracted if the event that triggers them is known beforehand. These are typically the items presented in *N*-back tasks. Unfortunately, not all interfaces that are of interest will consist of a *N*-back component. One approach is to introduce task-irrelevant probes for ERPs, such as an environmentally sound. More recent research has shown that involuntary ERPs to task-irrelevant stimuli vary in their amplitudes depending on the cognitive demands of the primary task that they are engaged in [54, 271, 332]. Application cases include estimating the workload of users playing Tetris across different difficulty levels for estimating the immersion of users in high-fidelity driving simulators.

5.6   Study Limitations

The study was affected by certain limitations. The study was conducted in a controlled laboratory setting, whereby participants were instructed to blink as little as possible and to avoid unnecessary head movements to minimize artifacts in the EEG signal. Users are unlikely to abide by such instructions in mobile real-world settings. Nonetheless, we point out that the manual assembly task is one that involves substantial activity. Therefore, our EEG recordings were likely to have contained a significant degree of motoric noise and continued to be robust for our current estimation of cognitive load. Cortical activity related to the planning of motor actions can also result in lower frequency power in a bandwidth that overlaps with alpha (*i.e.*, mu-power: 8 Hz to 12 Hz [300]). Therefore, the current results could have been affected by motor planning during item selection and placement within the

assembly task, rather than cognitive load per se. Nonetheless, we assume that our results are valid for cognitive load and not a motoric activity, given that we validated our EEG measure with a corresponding N-back task. Furthermore, we extracted the individual alpha band, which is not necessarily located in the mu frequency range.

5.7 Study Conclusion

In this study, we investigated if EEG is a suitable measurement modality for cognitive workload. In a case study leveraging an assembly scenario, we find that projected *in-situ* instructions reduce cognitive workload compared to traditional paper instructions. We employed a commercial EEG headset to derive direct measurements of neural activity that varied by working memory load in a N-back task, namely alpha power [199]. This same measurement (*i.e.*, alpha power) was larger for projected *in-situ* instructions than the traditional approach of paper instructions, demonstrating that cognitive load was lower when a projected *in-situ* system was employed. To date, only subjective questionnaire estimates have been collected. The current work demonstrates the viability of using a commercial EEG device for evaluation purposes, even in a setting that involves a large amount of user activity (*i.e.*, manual assembly). The collected data set is hosted on Github[4].

5.8 Chapter Conclusion

In this chapter, we investigated if EEG represents a suitable real-time metric for cognitive workload. We analyzed alpha oscillations to quantify the current demand of working memory load, a cognitive component that describes the available capacities of the short-term memory. Contrary to post hoc assessments, EEG provides a real-time measure for cognitive effort. Using a manual assembly experiment as a case study, we found that cortical activity is a reliable indicator to measure the cognitive demand placed by user interfaces. Based on our results, we provide an evaluation pipeline that suggests the use of EEG for user interface evaluation. Reflecting on our results, we conclude that EEG is a suitable real-time measure for cognitive workload that answers according to the research question (**RQ3**). The presented pipeline (see Section 5.5.3) provides an initial cornerstone for the real-time evaluation and intervention of EEG-based workload-aware interfaces. Therefore, we envision the use and development of novel pipelines with existing or alternative physiological modalities that build on the presented approach.

---

[4] `www.github.com/hcum/identifying-cognitive-assistance` - last access 2020-01-28

Chapter 6

Smooth Pursuits

The eyes are the mirror of the soul and reflect everything that seems to be hidden; and like a mirror, they also reflect the person looking into them.

Paulo Coelho

The previous chapter investigated the feasibility of EEG signals as a measure for cognitive workload. The analysis of cortical activity has proved useful to detect changes in cognitive demand in real-time. Thereby, EEG provides the capability to derive the mental demand directly from the brain. However, EEG relies on a correct setup of the headset that has to be kept steady to avoid noise in the signal. Other factors, such as head movements and eye movements, may influence the signal as well. Therefore, alternative modalities have been juxtaposed to overcome the limitations that EEG measures pose.

As such, eye tracking has emerged as both inferences for cognitive states as well as an interaction modality. Eye gaze-based interactive systems hold a lot of promise for cognition-aware interaction [53]. The human eye represents the central organ which enables visual processing. Complemented by previous research, eye gaze as an input for interactive systems has been extensively explored [102]. One specific eye movements, namely smooth pursuit eye movements, have been utilized as an alternative eye movement input modality for interactive systems [176, 374]. In contrast to EEG, smooth pursuit has the advantage to be a contactless measure. Eye movements can be recorded using a remote eye tracker, neglecting the need for hardware that has to be attached to the user. Furthermore, smooth pursuit eye movements cannot be pretended since they require locking onto a moving target [372]. This robustness against false positives is another benefit of using smooth pursuit eye movement as an interaction technique.

Research has shown that the behavior of the eye is strongly affected by psychological [311, 335] and psychophysiological states of the human body [347, 400]. One such state

91
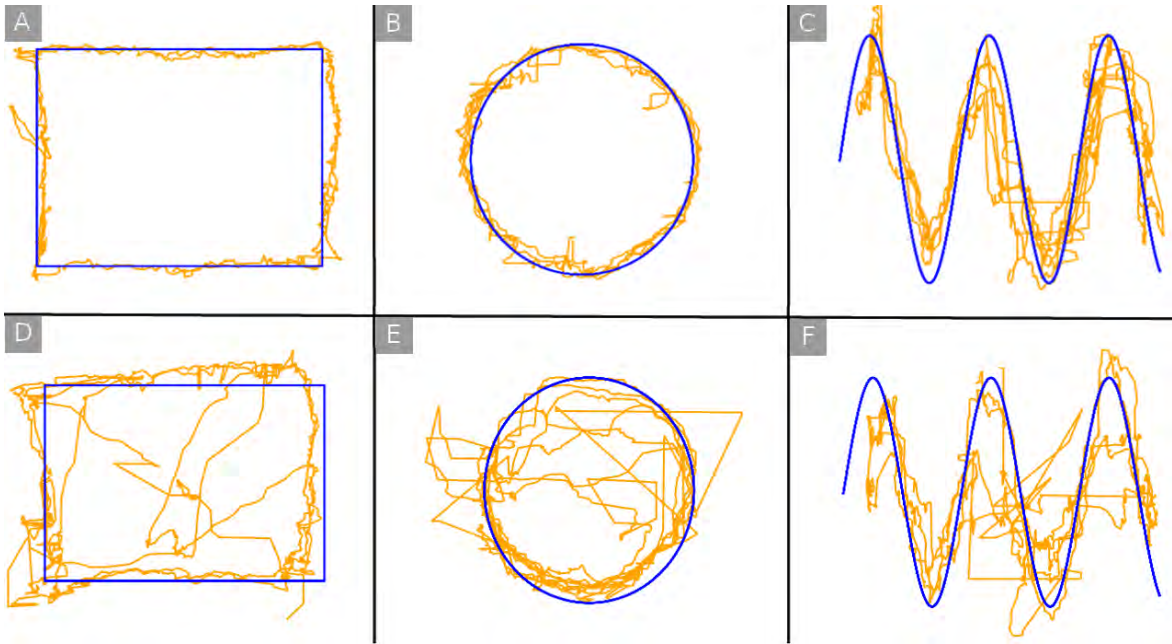
Figure 6.1: Smooth pursuit recordings under different levels of cognitive workload for three trajectories. The blue line shows the displayed trajectory, while the orange line visualizes the gaze path. (A), (B), and (C) show the gaze path during low cognitive workload phases, while (D), (E), and (F) shows the gaze path during states of high workload.

is cognitive workload, which has a remarkable impact on eye movements [18, 184, 351]. Figure 6.1 shows how cognitive workload affects smooth pursuit eye movements. Driving in a stressful context or performing multitasking during cognitively demanding tasks under time pressure are just two examples. Previous measures for cognitive workload facilitate questionnaires, such as the NASA-TLX [160, 161] or the Driver Activity Load Index (DALI) [297], which are often used in HCI to evaluate interfaces. Yet these questionnaires are prone to the interpretation of the questions by the user and only allow for an assessment at the end of the task, hence providing rather coarse-grained insights. At the same time, eye-tracking data can be used to assess cognitive workload [95]. In particular, previous work showed that under controlled conditions, pupil dilation provides an estimate of cognitive load [302, 304]. However, pupil dilation is highly sensitive to light conditions, which change in outdoor environments, and hence cannot easily be applied in many ubiquitous computing settings.

In this chapter, we propose an approach that exploits smooth pursuit eye movements to assess cognitive workload. We hypothesize that working memory, a component of cognitive workload, measurably affects smooth pursuit eye movements [14]. Henceforth, we define smooth pursuit describes eye movements that occur as the eyes closely follow a moving object. These movements are evaluated in several studies used for user interface element selection [114, 372] or intuitive eye tracker calibration [194, 374]. By conducting a user study, we found that smooth pursuit can be used for a contactless assessment of cognitive workload with an accuracy of up to 99.5%. Thereby, the need for body-worn sensors is obviated. This

chapter reports on a user study that investigates the impact of cognitive workload on smooth pursuit eye movements. Based on the insights from the user study we build a classifier to approximate the level of cognitive workload using gaze differences during smooth pursuit eye movements. We discuss how these findings can be used during the evaluations that require the assessment of cognitive workload as well as for the design of cognition-aware interactive systems. This chapter investigates smooth pursuit eye movements as part of the following research question:

→ **RQ4**: Do eye gaze metrics enable the classification of cognitive workload states?

This section is based on the following publication:

- Thomas Kosch, Mariam Hassib, Paweł W. Woźniak, Daniel Buschek, and Florian Alt. Your Eyes Tell: Leveraging Smooth Pursuit for Assessing Cognitive Workload. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 436:1–436:13, New York, NY, USA, 2018. ACM

6.1 Related Work

We present related work concerning existing research on using smooth pursuit eye movements for interaction and the influence of cognitive workload on eye movements.

6.1.1 Interacting with Smooth Pursuit

Using eye gaze as input for ubiquitous interactive systems has been extensively explored. Duchowski [103, 102] performed a literature survey about current eye-tracking applications and technologies. Focusing on usability, related research shows how eye tracking technologies evolved until they became usable in human-computer interaction research [179, 308]. Recent research has addressed the use of eye gaze for selection tasks [284, 345], which can be used by physiologically impaired individuals to perform input on computers [176]. Text input via eye input has been researched using dwell time [25, 256], off-screen targets [178], or eye movements along the y-axis on a display [365]. The usage of such methods was also evaluated in real-world scenarios [91, 399]. Furthermore, eye movements and eye gestures can be used to interact with devices.

Recently, smooth pursuit eye movement [176, 374] has been used as an alternative gaze-based input modality to interactive systems. This interaction technique overcomes the need for precise calibration and training of the user before interaction and further supports ubiquitous interaction. Several researchers have investigated how smooth pursuits can be used

intuitively. People cannot pretend to perform smooth pursuit movements since they require locking onto a moving target [372]. This reduces the likeliness of detecting false positives during an interaction.

Another approach to leverage smooth pursuit as an input modality has been researched by Esteves et al. [114]. They used a mobile eye tracker to enable hands-free interaction with a smartwatch by showing moving dots on the smartwatch display. Results show robustness against false positives regarding input, no need for calibrating an eye tracker, and efficient hands-free interaction. Using smooth pursuit to interact with physical real-world devices also showed promising results. Velloso et al. [372] developed and evaluated an object-driven system, which can be operated by only performing smooth pursuit. As soon as a user approaches an object such as a fan or windmill, it begins to present moving targets that trigger actions as soon as selected by smooth pursuit.

Researchers have also used smooth pursuit to calibrate eye trackers. Pfeuffer et al. [303] investigated this approach by using animations on a display to implicitly calibrate an eye tracker. Their results state that a 5-point calibration achieved a higher detection rate and required less calibration time. They concluded that smooth pursuits calibration provides better usability and flexibility for eye tracker calibration on small displays.

Vidal et al. [374] proposed different applications for smooth pursuit eye interaction, such as selection tasks or password authentication. The implemented applications showed a fast selection and completion time for different tasks. Since smooth pursuit can be performed with almost every kind of animated trajectory, more recently Khamis et al. [194] investigated which trajectories are most suitable for smooth pursuit interaction and calibration when showing text-based content on a display. While showing a question in the top left corner of a screen, they also showed several answers to the question on the same screen displayed in different trajectories. Schenk et al. [336] used smooth pursuit as an element selection mechanism in desktop settings to avoid the Midas touch problem [180]. Similarly, Lohr and Komogortsev [250] compared smooth pursuit-based input against dwell-based input approaches. A significant faster selection of elements was achieved when using smooth pursuit at the cost of likely unwanted selections.

6.1.2   Impact of Cognitive Workload on Eye-based Properties

Previous research has shown that the behavior of the eye is strongly affected by the psychological [311, 335] and psycho-physiological [347, 400] state of the human body. One such state is cognitive workload, which has a shown impact on eye movements [18, 351].

Researchers found a relation between pupillary responses and cognitive demand [2, 166]. During a task comprising different task complexities, the pupil diameter of participants was measured as an indicator for cognitive workload. Results show increasing pupil dilation with increasing task difficulty. Pfleging et al. [304] created a model to estimate cognitive workload under different lighting conditions based on pupil dilation. The feasibility of measuring

pupil dilation using a remote eye tracker has been investigated by Klingner et al. [203]. Their findings show a lower accuracy compared to mobile eye trackers due to noise and head movements. Kruger et al. [235] investigated eye behavior when perceiving a stimulus with and without subtitles using electroencephalography and pupillary measurements. In an experiment with two groups, lower cognitive workload was measured within the group perceiving subtitles than in the group lacking subtitles.

Liang and Lee [247] compared the efficiency of different machine learning algorithms to estimate distractions during driving tasks based on saccadic eye movements. Higher frequencies of saccadic eye movements have been used as an assessment for cognitive workload. Best results were achieved using a Support Vector Machine (SVM) for machine learning. However, their findings address saccadic eye movements only and do not examine the effects of smooth pursuit eye movements under cognitive workload.

Benedetto et al. [28] investigated the correlation between blink duration and visual and cognitive workload of a driver operating a car in a simulated environment. Participants had to perform a Lane Change Test [262] while doing an in-vehicle information system task [373] at the same time. Their findings show a lower blink duration compared to their baseline task than in the cognitively demanding task. Ahlstrom and Friedman-Berg [3] also found a significant correlation between shorter blink durations and cognitive workload.

Stuyven et al. [351] investigated the impact of cognitive workload on saccadic eye movements. Their findings show an increased occurrence of saccadic eye movements when inducing cognitive workload. Tsai et al. [363] investigated how eyes behave under cognitive workload while performing a paced auditory serial addition task [153], however, their work did not investigate the impact on smooth pursuit eye movements. Recently, Zagermann et al. [396] developed a model and showed concepts to derive cognitive workload from eye behavior, such as saccades, fixations, pupil dilation, and eye blinks. Cognitive workload influences microsaccadic eye movements [346]. They found that microsaccades occur more frequently with higher perceived workload during a non-visual task.

The voluntary involvement of smooth pursuit eye movements has been researched by Barnes [18]. He showed that cognitive processes in smooth pursuit eye movements are even involved without voluntary participation. Important factors for following moving stimuli were attention and awareness, which trigger the process of smooth pursuit on a neuroscientific level. Therefore, smooth pursuit eye movements are voluntary up to a certain degree. In contrast, Collewijn and Tamminga [78] investigated contexts in which smooth pursuit movements are voluntary. They used different targets and backgrounds to investigate smooth pursuit performances of humans. Contreras et al. [79] researched the eye-target synchronization performance of people with traumatic brain injury. People suffering from traumatic brain injuries show a worse performance when performing smooth pursuit in terms of deviation points apart from a shown moving object on a display.

Previous work has investigated how smooth pursuit can be leveraged for calibration and interaction. However, to our knowledge, no prior work has proposed to use work to propose using smooth pursuit eye movements as a measurement to derive cognitive workload. This
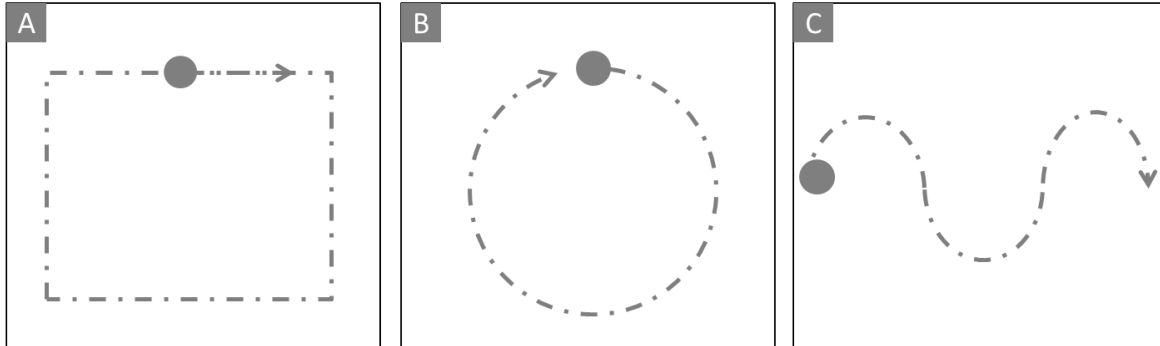
Figure 6.2: Three different trajectories chosen for evaluation: (A) Rectangular trajectory, (B) Circular trajectory, (C) Sinusoidal trajectory. The size of the moving object in the experiment was 10 pixels.

opens new opportunities for researchers and practitioners alike. In particular, using smooth pursuits as a real-time measurement for cognitive workload is valuable in the context of evaluating interactive systems as well as for developing cognitively adaptive systems.

6.2 Study

To understand the impact of cognitive workload on smooth pursuit eye movements, we designed a lab user study where we induced cognitive workload while participants performed smooth pursuit eye movements.

6.2.1 Independent Variables

In our experiment on investigating the impact of cognitive workload on eye movements, we explore the influence of three independent variables when interacting with smooth pursuit systems: (1) trajectory type [194], (2) speed of the stimulus [193], and (3) task difficulty. In the following, we describe these three independent variables in detail.

Trajectories

Since smooth pursuit eye movements can be triggered by showing a stimulus moving along a particular trajectory, we implemented a rectangular, a circular, and a sine wave animation to produce this effect. We have chosen these trajectories based on previous research [193, 194, 374]. Furthermore, the chosen trajectories may have a physiological effect paired with the current task difficulty, since the human eye has six muscles responsible for horizontal and vertical movements as well as eye rotations [241]. Rectangular trajectories demand muscles on the left and right side of the eye for horizontal movements, while vertical movements demand the upper and lower eye muscles. Circular and sinusoidal eye
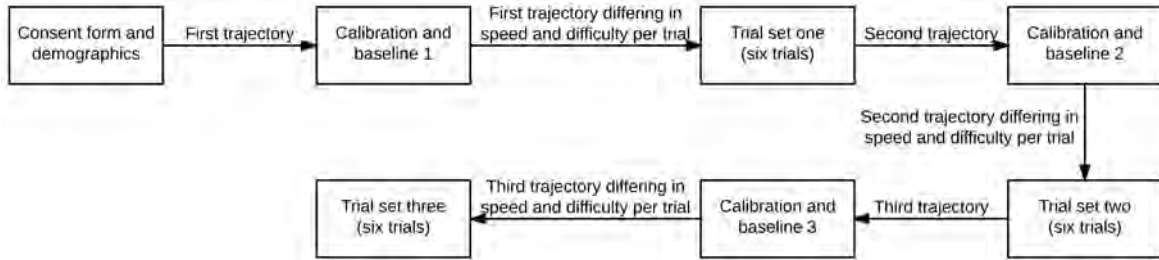
Figure 6.3: Study procedure. First, consent approval and demographic data were collected. Afterward, a baseline task was conducted followed by a set of six trials, where each trial differs in trajectory speed and task complexity. This procedure is repeated for every remaining trajectory type.

movements demand four muscles around the eye. Rectangular trajectories require horizontal or vertical movements only, while circular and sinusoidal include diagonal movements as well. In the context of leveraging smooth pursuit as interaction modality, the presented trajectories are used to distinguish between different user inputs. The chosen trajectories are depicted in Figure 6.2.

Speed of Stimulus

We compare two different speeds at which stimuli are moving based on previous research [193]: $450 \frac{px}{s}$ (slow[1]) and $650 \frac{px}{s}$ (fast[2]). Displaying slower or faster animations while experiencing cognitive workload can lead to different performances since eye muscles have to deal with different strains per speed and trajectory [241]. Furthermore, trajectory speeds can be used to differentiate user input, since eye movements adapt to different speeds.

Task Difficulty

To induce cognitive workload, we use an auditory delayed digit recall *N*-back task from Mehler et al. [266] with an English-spoken number set. The *N*-back task is a commonly used task to artificially elicit working memory resources [14], a component of cognitive workload which strains temporal memory capacities and affects secondary task performances negatively [30, 47, 366].

Throughout the study, we use the *N*-back task to manipulate cognitive workload by demanding working memory with different difficulty levels [186]. For each trial, participants hear randomized numbers consisting of ten digits between 0 and 9. Hereby, *N* corresponds to the *N*-last digit. After hearing the number, participants have to say out loud the digit they heard *N* digits ago. In our experiment, we use a *1*-back, *2*-back, and *3*-back tasks to induce cognitive workload. By increasing *N*, more digits have to be remembered, hence increasing task difficulty. To collect baseline measures, participants were asked to follow a trajectory

---

[1] This corresponds to 17.14° per second at a viewing distance of approximately 50 centimeters

[2] This corresponds to 24.76° per second at a viewing distance of approximately 50 centimeters

without performing an *N*-back task. Similar as in the study presented in Chapter 5, Table 6.1 shows an example of the *N*-back task.

Heard number	5	8	3	4	3	9	1
Number to say (*1*-back task)		5	8	3	4	3	9
Number to say (*2*-back task)			5	8	3	4	3
Number to say (*3*-back task)				5	8	3	4

Table 6.1: Example of the auditory delayed digit recall *N*-back task. Participants have to remember the *N*-th number back of a spoken number sequence and say the number out loud.

6.2.2 Apparatus

The study was conducted in a quiet lab with no windows, where lighting conditions were fixed. The setting was spatially divided into an experimenter area and a participant area. A separator divided both areas. While the experimenter controlled the experiment using a laptop, the participant saw an animated trajectory on a 22 inch screen with a resolution of 1680×1050 pixels and a refresh rate of 60 Hz. Eye gaze data was collected using a RED250 from SensoMotoric Instruments with a sample rate of 250 Hz. No filter was applied to the captured gaze data, thus raw gaze data was recorded only. We used a Holosonic Audio Spotlight 24i directed speaker to provide the auditory delayed digit recall *N*-back task.

6.2.3 Method and Measures

We used a repeated-measures design with three independent variables as described in the previous section; namely trajectory (rectangular, circular, sine wave), speed (slow, fast), and cognitive workload (no task, *1*-back task, *2*-back task, *3*-back task). Each experiment consisted of three sessions, where animated trajectories were changed for each session. The animated trajectory consisted of a white dot with a diameter of 10 pixels. The background was set to gray (RGB: [128,128,128]) to avoid eye exhaustion caused by screen brightness.

Before starting a new session, the eye tracker was calibrated to retrieve gaze points for later analysis of gaze deviations between baseline and smooth pursuit eye movements. Each session began with a 30 second baseline trial, where the animated trajectory used for the session was shown to participants without inducing cognitive workload. This allowed us to estimate eye movement differences from the displayed trajectory when no cognitive workload was present. We chose a slow speed ($450 \frac{px}{s}$) to make participants familiar with the displayed trajectory.

Cognitive workload was induced by providing a task difficulty using the auditory *N*-back task described in the prior section while showing the animated trajectory at a certain speed.

Task difficulty and speed were counterbalanced during a session according to the Latin square, while each session showed the same trajectory. The order of sessions was counterbalanced participant-wise according to the Latin square. This resulted in seven trials per session, including the baseline task. Running all three sessions, the experiment comprised 21 trials per participant. Overall, the duration of each trial was 25 to 30 seconds, dependent on the length of the spoken number for tasks including task difficulty. After completing a trial, participants were asked to fill out a NASA-TLX questionnaire to assess subjectively perceived cognitive workload. Participants took a 30-second long break afterward. Figure 6.3 shows an illustration of the study procedure.

6.2.4 Participants

We recruited 20 participants (9 female, 11 male), aged between 22 and 34 years (M = 27.5, SD = 3.13). Before the experiment, each participant signed a consent form and provided their demographic data. All participants were computer science, students or researchers. All participants had a normal or corrected-to-normal vision. Participants were recruited through university mailing lists. They received sweets and five Euro as compensation. The duration of the study was approximately 30 minutes. We explained the purpose of the study and tasks to the participants and informed them they could exit the study at any point. Participants signed informed consent forms and were seated in a comfortable chair, approximately 50 centimeters in front of the display before the experimental setup. Due to technical issues, two participants were excluded from the analysis, as their gaze data was not recorded properly.

6.3 Results

We analyze our data to compare the impact of trajectory type, speed, and task difficulty on smooth pursuit eye movements. We report on quantitative results by comparing measured eye gaze data with the showed trajectory. This is complemented by a subjective analysis through NASA-TLX questionnaires.

6.3.1 Smooth Pursuit Differences and Cognitive Workload

To evaluate the effect of different task difficulties on smooth pursuit performances, we based our analysis on pixel differences between coordinates of the displayed trajectory and measured eye gaze position at the screen. More formally, we calculated the difference between two coordinates p and q using the Euclidean distance with the formula

$$D = \sqrt{\left(p_{t,x} - q_{t,x}\right)^2 + \left(p_{t,y} - q_{t,y}\right)^2} \qquad (6.1)$$

Paired Wilcoxon Signed-Rank Test		Significance
Circle Baseline	Circle *1*-back fast	$p = .002$
Circle Baseline	Circle *2*-back fast	$p = .002$
Circle Baseline	Circle *3*-back fast	$p = .003$
Circle Baseline	Circle *2*-back slow	$p = .004$
Circle Baseline	Circle *3*-back slow	$p = .005$
Sine Baseline	Sine *1*-back fast	$p < .001$
Sine Baseline	Sine *2*-back fast	$p < .001$
Sine Baseline	Sine *3*-back fast	$p < .001$
Sine Baseline	Sine *1*-back slow	$p = .003$
Sine Baseline	Sine *2*-back slow	$p = .002$
Sine Baseline	Sine *3*-back slow	$p = .002$

Table 6.2: Summary of significant results. Comparisons between other conditions did not result in significant differences.

where p and q depict a two-dimensional vector comprising the baseline coordinates and measured gaze coordinates from participants. We introduce the variable t describing the temporal dependency between eye gaze and displayed stimulus as the used 250 Hz eye tracker might introduce a temporal offset of four milliseconds, which is below the perceptual threshold for interaction. Differences were normalized with respect to the maximum gaze deviation from the shown trajectory. To enable a descriptive analysis, normalized coordinates were averaged for all participants and conditions. Lowest mean eye gaze deviations (M_{pd}) were measured for all slow trajectories, where no cognitive workload was induced (Rectangle: $M_{pd} = 9.14$, SD = 6.75, Circle: $M_{pd} = 9.51$, SD = 8.03, Sine: $M_{pd} = 4.37$, SD = 1.60). Fast rectangular trajectories using a *3*-back task ($M_{pd} = 13.25$, $SD = 11.06$), fast circular trajectories using a *2*-back task ($M_{pd} = 14.54$, SD = 7.14) and fast sinusoidal trajectories using a *3*-back task ($M_{pd} = 9.85$, SD = 3.53) led to highest mean eye gaze deviations.

Applying a Shapiro-Wilk test on the mean data set showed a non-normal distribution for all conditions (all $p < .05$). A Friedman test showed no significant differences between various levels of cognitive workload and gaze deviations of smooth pursuit eye movements for slow rectangles ($\chi^2(3) = 3.000$, $p = .392$). However, a Friedman test found significant differences within various levels of cognitive workload for fast rectangles ($\chi^2(3) = 11.867$, $p = .008$), slow circles ($\chi^2(3) = 18.867$, $p < .001$), fast circles ($\chi^2(3) = 29.400$, $p < .001$), slow sine waves ($\chi^2(3) = 14.667$, $p = .002$) and fast sine waves ($\chi^2(3) = 30.667$, $p < .001$). We conducted a Wilcoxon signed-rank post hoc test to find significant differences between pairs of task difficulties including baseline trials and normalized gaze deviations after applying a Bonferroni correction (significance level set to $p < .0083$). A summary of significant results can be depicted from Table 6.2. Further, the Cohen's d effect size values of the significant statistical comparisons ranged between $d = 0.66$ and $d = 0.88$.
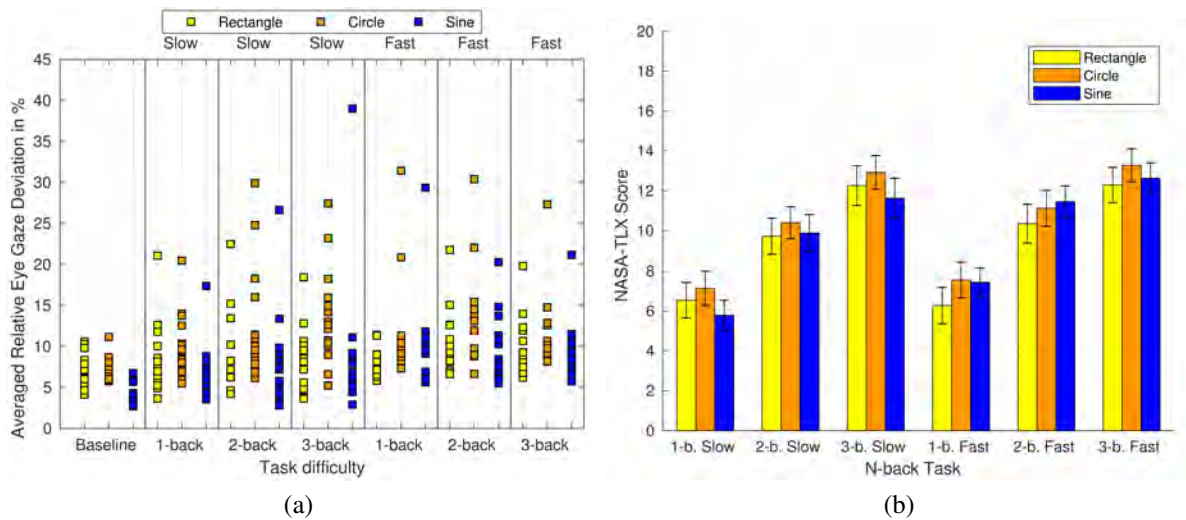
Figure 6.4: (a): Scatter plot of the mean gaze deviation per participant and condition. Each dot represents the mean gaze deviations between the recorded and showed the trajectory of a participant. The baseline measurements show a constant behavior and do not scatter apart from the displayed trajectory. Compared to the baseline tasks, circular and sinusoidal trajectories scatter along the y-axis with increasing task difficulty. (b): Mean NASA-TLX score for different trajectories, task difficulties, and speeds. Increasing task difficulties led to higher NASA-TLX scores. The error bars depict the standard error.

To visualize averaged relative differences when performing smooth pursuit eye movements, all averaged relative eye movement differences per participant and condition were plotted. This resulted in one data point per participant and per trial, or 21 data points per participant. The averaged plot is depicted in Figure 6.4a, where the y-axis depicts the gaze deviations from the shown trajectory in percent. The bottom x-axis is annotated with task difficulties and the top x-axis describes trajectory velocities. The trajectory type is color-coded.

Gaze deviations obtained from baseline trials do not scatter in contrast to gaze deviations measured from trials with task difficulty for circular and sinusoidal trajectories. Further-more, gaze deviations increase between slow and fast trajectories when raising the task diffi-culty. However, results show important constraints when assessing cognitive workload from smooth pursuit eye movements. Rectangular trajectories show less eye gaze deviations when increasing task difficulty compared to baseline trials. Thus, evaluating the presence of cog-nitive workload using rectangular smooth pursuit eye movements is less accurate compared to circular and sinusoidal trajectories. Depending on the setting of a smooth pursuit-driven user interface, rectangular trajectories should be favored when accurate input is required even when the user is impacted by cognitive workload. Circular and sinusoidal trajectories show higher gaze deviations under cognitive workload. Circular and sinusoidal trajectories show clear differences between trials with and without cognitive workload. Such trajectories may be used to determine the existence of cognitive workload.

6.3.2 Subjective Analysis of Cognitive Workload

The lowest NASA-TLX scores per trajectory were subjectively perceived by participants when showing fast rectangles during a *1*-back task (M = 6.27, SD = 4.03), slow circles during a *1*-back task (M = 7.13, SD = 3.87), and slow sine waves during a *1*-back task (M = 5.78, SD = 3.41). Per trajectory, the highest NASA-TLX scores were measured when displaying fast rectangles during a *3*-back task (M = 12.29, SD = 3.97), fast circles during a *3*-back task (M = 13.28, SD = 3.68), and fast sine waves during a *3*-back task (M = 12.63, SD = 3.42). A repeated measures ANOVA showed statistically significant differences between different levels of cognitive workload and NASA-TLX score ($F(2, 321) = 68.503$, $p < .001$). However, no statistically significant differences were found for different speeds ($F(1, 322) = 2.035$, $p = .211$) and displayed trajectories ($F(2, 321) = 21.461$, $p = .722$) compared to NASA-TLX scores. Figure 6.4b illustrates the averaged values from the obtained NASA-TLX scores.

The quantitative analysis shows significant outcomes regarding gaze deviations with different stimuli speeds and *N*-back complexities. To address the impact of different variables on the individual's subjective perception, we conducted a statistical analysis to investigate single NASA-TLX items in correlation to the stimulus speed and *N*-back difficulty. We compare the single NASA-TLX items grouped by the two different speeds. Our results show a significant difference for the *physical demand* ($p = .006$, $M_{slow} = 5.11$, $M_{fast} = 5.69$), *temporal demand* ($p = .003$, $M_{slow} = 7.29$, $M_{fast} = 8.27$), and *effort* ($p = .001$, $M_{slow} = 9.53$, $M_{fast} = 10.29$) scales. However, no significant difference was found in the mental load scale ($p > 0.05$).

To show that different *N*-back difficulties were responsible for higher perceived cognitive workload, we conducted a Wilcoxon signed-rank test to compare the mental NASA-TLX items between the different task complexities. We found significant differences between all three *N*-back difficulties (all $p < .001$) for the *mental demand* scales ($M_{N=1} = 5.84$, $M_{N=2} = 10.56$, $M_{N=3} = 13.84$). Further, we investigate whether the relationship between NASA-TLX scores and gaze deviations was linear using a Pearson correlation. No correlation was found ($.11 < r < .15$, $\forall N \in \{1, 2, 3\}$). Thus, it appears that the relationship between the variables is more complex *i.e.* non-linear. This shows the need to understand how smooth pursuit affects subjectively perceived workload.

6.4 Predicting Cognitive Workload

Results from the study showed significantly increased eye movements during cognitive workload for circular and sinusoidal trajectories. The perception of cognitive workload by the individual is supported by subjective ratings of participants using NASA-TLX questionnaires. We train a classifier that predicts cognitive workload from smooth pursuit eye movements. We investigate the performance of person-dependent and person-independent classification for the different experimental conditions.

Stimulus	Binary Pers.-Indep.				Multilabel Pers.-Indep.				Multilabel Pers.-Dep.	
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	F1
Rectangle Slow	75.0%	37.5%	50.0%	42.9%	40.3%	26.9%	40.0%	30.5%	48.6%	52.4%
Rectangle Fast	76.4%	40.1%	52.8%	46.0%	54.2%	44.0%	54.1%	47.2%	69.4%	74.2%
Circle Slow	95.8%	89.6%	91.7%	90.5%	40.3%	26.9%	40.2%	30.9%	69.5%	60.0%
Circle Fast	97.2%	93.0%	94.4%	93.7%	55.6%	45.8%	55.6%	48.5%	85.0%	74.4%
Sine Slow	80.1%	55.6%	64.8%	59.8%	43.1%	30.6%	43.0%	34.5%	79.3%	76.3%
Sine Fast	99.5%	98.9%	99.2%	99.0%	59.7%	51.9%	59.7%	54.4%	88.1%	84.5%

Table 6.3: Accuracy, precision, recall, and F1 scores of the binary and multi-label person-independent (pers.-indep.) as well as person-dependent (pers.-dep.) classifications. Fast circular and sinusoidal trajectories yield higher classification performances than slower linear trajectories.

6.4.1 Attributes, Instances, and Classes

Data preprocessing was necessary before training a predictive model. We removed the first two seconds per trial to avoid distortions in the signal caused by participants initially searching for the stimulus. We used the collected gaze data and normalized it concerning the coordinates of the shown trajectory. We then calculate the Euclidean distance (see Equation 6.1) between each coordinate of the normalized displayed trajectory and the normalized gaze points recorded from participants. These gaze deviations are defined as the only attribute for the instances we used for classifier training and evaluation later. We smoothed the data using an averaging running window (length 250 samples) and a hop size of one sample.

The instances used for classifier training and evaluation consisting of a one-dimensional vector, where the normalized gaze deviations represent the only attribute. We define two sets of classes with different class labels. The first class has two labels consisting of *low workload*, referring to trials where no *N*-back task was used, or *high workload*, thus referring to trials including any *N*-back task. We investigate binary classification performances using this class. In contrast, the second class contains the four labels *0-back*, *1-back*, *2-back*, and *3-back*, each referring to the measurements of trials using the corresponding *N*-back task. Overall, we constructed 378 instances, each containing 6000 gaze values, which were used for training and evaluation.

6.4.2 Classifier Performance

An SVM with a linear kernel was used to investigate the prediction performance [247, 364, 389]. The previously described instances were used for training and assigned to their appropriate class labels. Within our classifier learning process, we aimed to evaluate the classification efficiency for binary classification, such as detecting low cognitive workload and high cognitive workload. Furthermore, we investigated the classifier performance for deter-

mining different levels of cognitive workload concerning the N-back task difficulty. We used scikit-learn[3] to train different prediction models.

Person-Independent Binary Classification

We performed a person-independent leave-one-person-out classification using two labels; one assigned to low cognitive workload referring to baseline trials and one assigned to high cognitive workload referring to trials comprising a *1*-back, *2*-back, or *3*-back task. A leave-one-person-out classification uses all except one participant for training, while the final participant is used for validating the trained model. The leave-one-person-out classification was carried out for each trajectory and speed separately, where the instances contain the normalized gaze deviations from the shown trajectory per participant and difficulty. The leave-one-person-out classification procedure was repeated for every participant[4] and the results were averaged. Table 6.3 shows the accuracies, precisions, recalls, and F1 scores of the binary classification per trajectory and speed. The classification results favor fast trajectories in combination with changing directions, such as circular or sinusoidal stimuli.

Person-Independent Multilabel Classification

To investigate the performance of classifying different levels of cognitive workload, the same leave-one-out validation procedure was conducted for four labels, with each label assigned to instances with their respective task difficulty. Again, the validation was conducted for every participant after using the other participants for training. The results are shown in Table 6.3.

The overall classification efficiency is lower compared to the binary classification accuracy. Different reasons can be responsible for this. First, we are aware that multiple labels may lead to a lower efficiency when the difference between values within the N-back conditions is low. Second, we combine multiple participants who may differ in their smooth pursuit behavior individually. The generalized data could, therefore, be biased by individual gaze behavior, which leads to a distorted result. Therefore, we investigate person-dependent training for cognitive workload classification purposes.

Person-Dependent Multilabel Classification

We analyzed instances within each person to investigate if higher accuracies could be achieved due to person-dependent properties. Instead of a leave-one-person-out classification, we conducted a leave-one-repetition-out classification for each participant, speed, and trajectory. However, the number of folds had to be set differently, since the animations iterated a different number of times depending on the trajectory and speed[5]. Therefore, we

---

[3] `www.scikit-learn.org` - last access 2020-01-28

[4] Overall 18 runs to use every participant for validation

[5] Each trajectory started and ended at the same position

adjusted the number of folds of the leave-one-repetition-out classification per trajectory and speed.

The number of folds was set to $k = 2$ for slow rectangles, $k = 3$ for fast rectangles, $k = 5$ for slow circles, $k = 7$ for fast circles, $k = 2$ for slow sine waves, $k = 3$ for fast sine waves. The data per trajectory and speed and participant was randomly partitioned into k folds, where $k - 1$ folds were used for training while the last fold was used for evaluation. This procedure was conducted k times per participant and trajectory with the different N-back difficulties assigned to the training set. The results were first averaged per participant and then over all participants. Table 6.3 shows the results of the cross-validation. Person-dependent classification of multiple cognitive workload levels shows a higher accuracy for fast smooth pursuit movements. Especially fast circular and fast sinusoidal trajectories show better performances compared to rectangular trajectories.

6.5   Discussion

Results from our study show how cognitive workload leads to increased gaze differences of smooth pursuit eye movements during the presence of cognitive workload. We found that faster circular and sinusoidal trajectories led to higher gaze deviations, while rectangular trajectories did not show this effect. We believe, that circular and sinusoidal trajectories required more effort from users since they had to focus more on their task completion due to constantly changing directions. This may be a reason for improved classification results for this kind of trajectories. In contrast, rectangular trajectories can be deployed within smooth pursuit-based interfaces whenever reliable input, independent from the perceived cognitive workload, is required. Depending on the use case, a cognition-aware system designer can decide if reliable input through rectangular trajectories or mental workload estimation by circular and sinusoidal trajectories is desired.

The speed of the trajectories had a quantitative measurable effect on the overall smooth pursuit performances. However, as the statistical investigations of single NASA-TLX scales showed, *mental demand* was not affected compared to *physical* and *temporal demand* as well as *effort* by different speeds. Faster speeds cause therefore a different type of physical workload than cognitive workload which significantly impacts smooth pursuit eye movements. Different trajectory speeds are thus not necessarily responsible for inducing cognitive workload. By comparing different N-back difficulties about the NASA-TLX scales, we found a significant difference between all difficulties. This supports, that subjectively perceived cognitive workload was altered by different N-back complexities and that both variables manipulate measurable smooth pursuit performances.

The classification results yield higher accuracies for distinguishing between low and high workload levels than for detailed levels of cognitive workload. Furthermore, binary classification can be achieved without the need for person-dependent calibration, while separating

different levels of cognitive workload requires a person-dependent calibration regarding cognitive workload.

In a real deployment scenario, binary classification can be used to provide additional help for users when high cognitive workload is classified. This refers to very simple scenarios, where only the estimation of low and high workload is desired. Such places could be public places when, for example, interacting with public displays. In contrast to short interactions in the public, a classifier can be trained for multilevel classifications in private spaces where long-term interaction is conceivable.

Calibrating a classification model with multiple workload levels in the public can result in a cumbersome procedure due to external factors impacting the individual cognitive capacities, such as distracting pedestrians walking by. Our results indicate binary workload classification in the public using a pre-trained classification model, while a classification model with multiple workload levels can be used in private spaces where calibration can be done without any disruptions.

To employ sensing of cognitive workload, smooth pursuit must be elicited. In contrast to smooth pursuit-based user interfaces, other environments require explicit or implicit integration of moving elements. For example, as short waiting times occur as a result of a database query or as a new task is loaded during a user study, feedback on the system status could be presented in a way that fosters smooth pursuit movements. For example, this can be elicited by an animated progress bar. In this way, traditional methods such as the NASA-TLX or DALI questionnaires can be complemented.

Finally, a suitable workload level must be found. Looking back at the classification results of the multi-label model, we achieved reasonable accuracies for person-dependent classification. A cognition-aware system must be able to find the right difficulty for each user, as permanent support by a system may lead to boredom or no support leads to frustration for the user. Keeping the task difficulty at its highest or lowest level might not be favored. However, for deploying binary classification in public use cases minor support might be helpful compared to private settings, where user expectations on cognition-aware computing systems are higher.

6.5.1 Limitations

Our study is prone to certain limitations. We collected gaze data under controlled lab conditions and hence, do not know how our results generalize to other situations, where participants may be distracted for example, by the presence of other people. Still, despite the controlled conditions, participants' gaze behavior may have been influenced by physiological wellbeing, such as lack of sleep. Our calculation of gaze difference is based on 30 seconds of recording. However, there may be situations in which assessing workload with finer granularity is desirable. Additionally, blink frequency and blink duration were not evaluated during our studies. Another limiting factor was the study execution in a calm lab. Natural

distractions in real-world environments could alter our results. Furthermore, we have not investigated the effect of different eye tracking frame rates in our study. Consequently, before assessing cognitive workload through smooth pursuit, eye trackers must be tested for their suitability.

6.6 Use Cases for Workload-Aware Systems

In the following, we present several different use cases to be supported within smooth pursuit scenarios.

6.6.1 Support in Safety-Critical Environments

Smooth pursuit can be utilized to assess cognitive workload during a monitoring task, such as air traffic controller surveying airplane flight processes, to support or warn operators for cognitive exhaustion. Alternatively, workload can be dispatched to a colleague who does not have to cope with high workload during work. Since objects of interest can be visualized using a small moving circle, smooth pursuit eye movements can be triggered this way. As a result, accidents, which occur due to mental overload, distractions, or fatigue, can be avoided. For example, the user interface can be adapted by simplifying displayed content or dispatching a part of the observation task to another colleague. This use case is transferable to other applications, where moving objects occur naturally and require permanent attention of the user. Figure 6.5b shows an example of how smooth pursuit can be used in such situations.

6.6.2 Adaptation of Pursuit-based Interactive Systems

Prior research has introduced many applications that use smooth pursuit primarily for interaction. As we illustrated in our related work section, smooth pursuit interfaces have been used for smartwatch interaction [114] and interaction with smartphones [248]. Using mobile devices enables ubiquitous sensing of cognitive workload in outdoor settings, bypassing the disadvantages of using pupillary measures being prone to lighting conditions. Interacting with large distant displays [374], where touch and gesture interaction have emerged as state of the art to communicate input [298, 376] can use smooth pursuit to implicitly measure mental states.

Our approach enables implicit contactless assessment of cognitive workload while interacting with these devices. Pursuit-based interactive systems benefit from our classifier to dynamically predict the current cognitive workload level of the user during interaction and adapt to it accordingly. If high task load is identified during an interaction, the user interface or task objective can be modified by an easier one. Figure 6.5a shows how the existence of cognitive workload can be estimated to adapt a public display app.

(a)   (b)
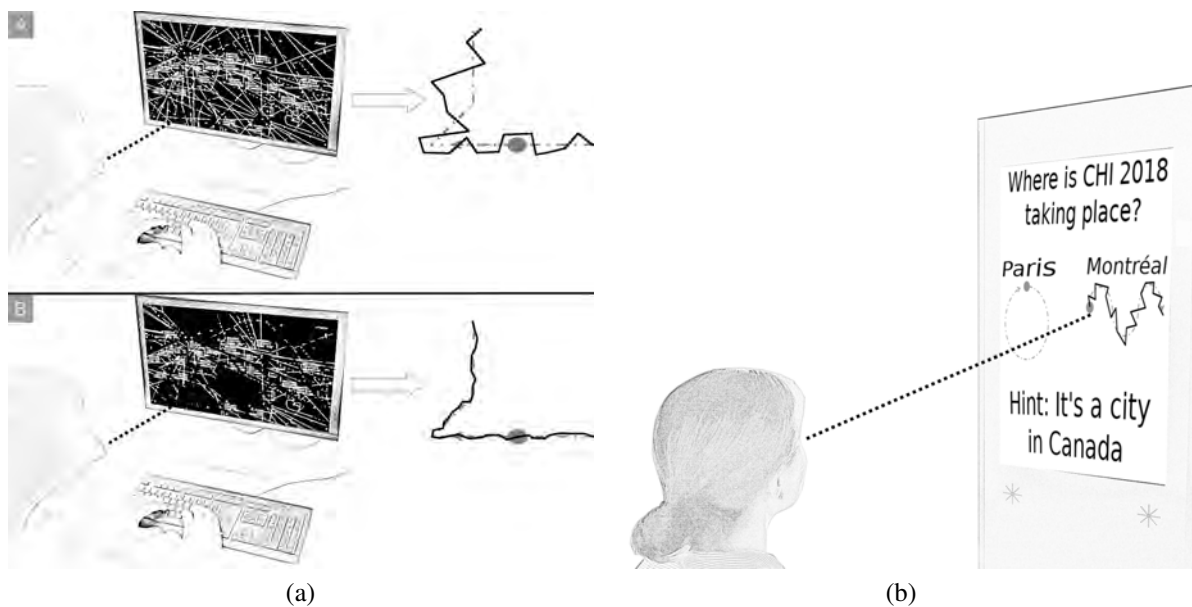
Figure 6.5: (a): User working in an air control tower. Moving dots representing airplanes on the screen can cause smooth pursuit eye movements. (A) The system detects high cognitive workload from the user and dispatches some observation tasks to a colleague. (B) Alleviated cognitive workload measured after user interface adaption. **(b):** User playing a quiz game on a public display. The system infers that the question is inducing high cognitive workload. The system is, therefore, observing if this behavior persists and provides a hint to avoid frustration.

6.7   Study Conclusion

The presented study investigated the influence of cognitive workload on smooth pursuit eye movements using three different trajectories with two different velocities. Using an auditory delayed digit recall *N*-back task to induce cognitive workload, a higher deviation of gaze points from shown trajectories is measured compared to measurements when not inducing cognitive workload. Based on our results, we create a person-independent classifier for estimating binary workload and a person-dependent classifier for distinguishing different levels of cognitive workload. While binary cognitive workload classification can be elicited in the public using smooth pursuit interfaces, private spaces benefit from person-dependent classifier calibration to determine different levels of cognitive workload. The collected data set is available on Github[6].

---

[6] `www.github.com/hcum/your-eyes-tell` - last access 2020-01-28

6.8   Chapter Conclusion

In this chapter, we investigated the characteristics of eye movements during different levels of cognitive workload. Eye movements can be recorded using a remote eye tracker which obviates direct body contact with the user. We show that cognitive workload levels can be classified using specific smooth pursuit eye movements using machine learning techniques. Contrary to EEG, no calibration procedure or tedious setup is required. Thus, the real-time analysis of eye movement holds promise as an additional measurement channel for workload-aware computing interfaces.

Having such a measurement modality without the need of body-worn devices goes a step towards real-time mental state estimation in ubiquitous computing environments. User interfaces can then provide intervention mechanisms to relax or help users based on their current context. Our classifier depends on eye gaze only and fits into several application scenarios. It can be deployed in real-world scenarios to estimate the presence of cognitive workload in real-time. Thereby, the assessment can be done without the need for contact or additional bodyworn sensors, which comes closer to Mark Weiser's vision of ubiquitous computing. However, the classification performance varied throughout the displayed trajectories, and for both person-dependent and person-independent cases. Whereas accuracies of up to 99% can be achieved for binary person-independent classification, person-dependent classification excels at the classification for different N-back difficulties. Workload-aware interfaces that are deployed in the public may benefit from cognitive workload estimation using a binary classification whereas it may be acceptable to calibrate for fine-grained workload determination in private spaces. This chapter answers **RQ4** with the focus of measurable changes in smooth pursuit eye movements, which can be leveraged by workload-aware systems for cognitive workload classification.

Pupil Dilation

Sitting in the far horizon
The evergreens across the border
Can you see a different story
With the failure and the glory
The truth is in the eyes of the beholder
The truth is in her eyes when you hold her

Andrew Stockdale

The previous chapter investigated changes in eye movements during different levels of cognitive workload. An *N*-back task was employed to elicit different levels of cognitive workload. Specifically smooth pursuit eye movements were affected and could be successfully classified between and within persons. However, moving elements are required to elicit smooth pursuit movements since they cannot be pretended or simulated by humans [372]. Alternative metrics for measuring cognitive workload using eye-based properties have found relevance in previous research [102, 104] including pupil dilation.

Pupil dilation has attracted attention in the HCI community as a measure for cognitive workload (see Figure 7.1) [105]. Demanding the short-term memory frequently over a timespan causes cognitive effort [10, 11, 164, 166, 185], a mental process which extends the pupil diameter (see Figure 7.1b). Eye trackers are used to retrieve eye gaze positions which can be used at the same time to evaluate pupillary data in real-time. Using such a metric as an assessment of cognitive workload can be utilized by applications to provide help and assistance for users when high cognitive workload is detected. For example, workload-aware systems can assist in real-time during a cognitively demanding task. Furthermore, the usability of interfaces can be evaluated by conjunctively assessing pupil diameter and eye gaze during user interaction. Although being sensitive to lighting conditions, the evaluation of pupil dilation to infer cognitive workload has been examined by various researchers. Benedetto et al. [28] explored the impact of cognitive workload on pupil dilation and eye blink duration. Since

<center>(a)</center>

<center>(b)</center>
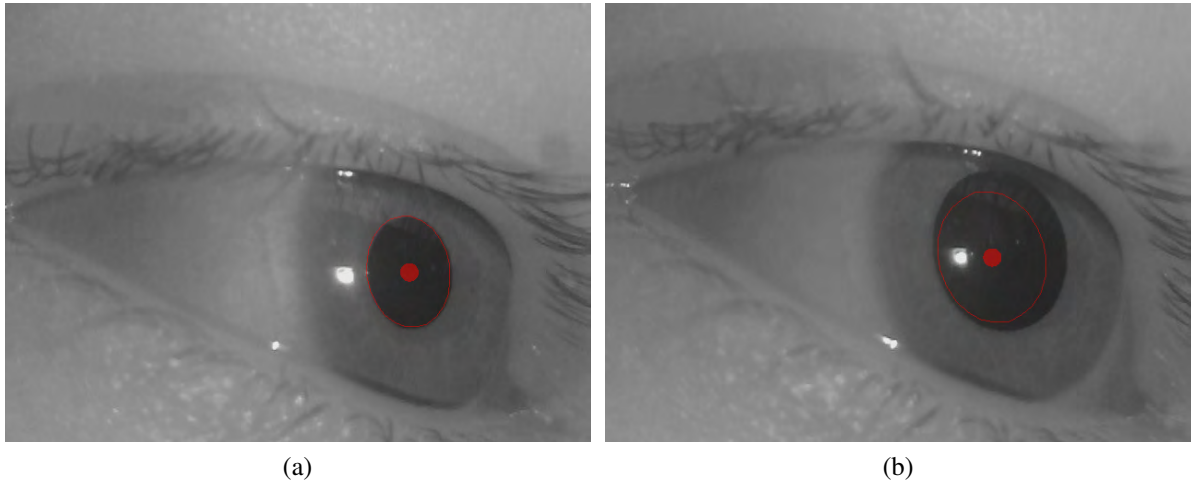
Figure 7.1: Exemplary impact of cognitive workload on pupil dilation. **(a):** Measured pupil diameter under low cognitive workload. **(b):** Measured pupil diameter under high cognitive workload.

pupillary measurements are error-prone to lighting conditions, Pfleging et al. [304] proposed a model for classifying cognitive workload of pupil dilation under different lighting conditions. Kiefer et al. [197] evaluated the pupil diameter under different task difficulties to assess the perceived task complexity. Gollan et al. [149] examined the assessment of pupil dilation under cognitive workload in real-time. Hess et al. [166] investigated the impact of simple mental processes, such as solving one and two-digit multiplications, on the pupil diameter. Zbrodoff and Logan [398] found a correlation between higher pupil extensions and increasing math exercise complexity. By using multiplication tasks, a higher pupil diameter was measured when multiplying two-digit numbers than one-digit numbers.

This chapter investigates the performance of a classification model that distinguishes between two different levels of workload. This includes the assessment of pupil dilation to create adaptive workload-aware interfaces that adjust the difficulty in real-time. Thereby, task complexity is set to a suitable difficulty to keep users engaged, thus becoming more difficult or easier depending on current mental workload measurements. This is complemented by presenting and evaluating a proof-of-concept application, which sets its task complexity based on pupillary data to prevent mental underload or overload. Hence, this chapter complements the second part of the research question:

→ **RQ4**: Do eye gaze metrics enable the classification of cognitive workload states?

This section is based on the following publication:

- Thomas Kosch, Mariam Hassib, Daniel Buschek, and Albrecht Schmidt. Look into My Eyes: Using Pupil Dilation to Estimate Mental Workload for Task Complexity Adaptation. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages LBW617:1–LBW617:6, New York, NY, USA, 2018. ACM

7.1 Application

To examine the usefulness of adapting task complexity based on pupillary data, we constructed a prototype capable of evaluating pupil dilation measurements in real-time.

7.1.1 Recording and Feedback Setup

A mobile eye tracker from Pupil Labs[1] is used to receive pupillary data. Data was updated at 30 Hertz and processed on the attached computer. An external screen is used to display stimuli. To avoid distractions during the experiment, the setup was divided into two areas using separators, respectively for the experimenter and participant. The experiment was conducted in a room without windows and constant lighting conditions. The stimulus monitor had a resolution of 1920×1080 and a screen size of 23 inches.

7.1.2 Classifier

Two baseline trials are conducted in the beginning to obtain ground truth data about pupil dilation during an *easy* and a *difficult* math task, each lasting three minutes. A single complexity was assigned to every trial (*easy*/*difficult*). The trained classifier is then used to adaptively control the complexity of the displayed multiplications in a third trial. The classifier aims to estimate low and high cognitive workload based on the data retrieved from the two baseline tasks. The prototype simulates a cognition-aware system by setting task complexity to *easy* if high mental workload is measured. In contrast, task complexity is set to *difficult* when low cognitive workload is classified from the previous baseline measurements.

To assess cognitive workload from the users' pupil dilation, we trained individual support vector machines (SVMs) with a linear kernel to infer required changes of task difficulty in real-time. We used the individual pupil dilation as the only feature. We defined two classes, *easy* and *difficult*. The classifier was trained individually for every participant after the two

---

[1] `www.pupil-labs.com` - last access 2020-01-28

baseline trials. In the last trial, the previously trained classifier is used to predict cognitive workload. The task difficulty is set to the opposite task complexity to force cognitive alleviation or effort. The person-dependent classifier accuracies ranged between 64% and 99% (M = 79%, SD = 0.16%) resulting from a k-fold cross-validation with $k = 5$.

7.2 Study

We conducted a pilot study to evaluate the feasibility of changing task complexity based on pupil diameter measurements as an indicator for cognitive workload. The study configuration conforms to the previously described prototype.

7.2.1 Cognitive Task

To induce cognitive workload, we use a multiplication math task with two different complexity levels including one-digit and two-digit multiplications. Both complexities lead to different pupil dilations [10, 11, 398]. To add a time constraint, the multiplication math tasks were moving centered from the top to the bottom of the screen within five seconds. During the time limit, participants were asked to type the correct solution on a keyboard number pad. When the multiplication reached the bottom of the screen, it disappeared and a new multiplication is displayed at the top of the screen. If the user fails to enter a solution during the given time frame, an error is counted. If the solution entered during the time is correct, the multiplication disappears and a new multiplication task is displayed at the top of the screen. The complexity is divided into *easy* and *difficult*. The *easy* condition uses numbers ranging from $0 - 9$, randomly selecting two numbers which have to be multiplied together. In the *difficult* condition, two numbers are randomly chosen, where the first number ranges from $10 - 19$ and the second from $0 - 9$ (see Figure 7.2). Numbers are constantly displayed in black font on a gray background (RGB: $[150, 150, 150]$).

7.2.2 Participants

Six participants (4 male, 2 female) took part in the pilot study, ranging from an age between 22 and 34 years (M = 29.17, SD = 4.41). All participants were computer science, students or researchers. Participants had a normal or corrected-to-normal vision. The overall duration of the study was approximately 15 minutes. Before the start of the experiment, participants signed an informed consent form and provided their demographic data.

Figure 7.2: Multiplication math task with equations moving from top to bottom comprising two different complexities. **(a):** *Easy* multiplication containing one-digit numbers. **(b):** *Difficult* multiplication comprising a two-digit number and a one-digit number.

7.2.3 Procedure

First, we explained to the participants the study procedure and familiarized them with the setting. The mobile eye tracker was then calibrated. The study consisted of three trials; two baseline multiplication trials (*easy/difficult*) and one *adaptive* multiplication trial. The duration of each task was three minutes. The order of the first two baseline trials was counterbalanced according to the balanced Latin square. The pupillary data from the baseline tasks was used to derive an SVM classifier capable to distinguish between cognitive workload induced by a *easy* or *difficult* task. During the *adaptive* trial, the trained user-dependent classifier is used to adaptively set the complexity of the multiplication task. The complexity is set by classifying measured pupil diameter data in five-second intervals. If the classifier estimates a high diameter difference, the task complexity is set to *easy* to avoid mental overload. In contrast, the task complexity is set to *difficult* if the classifier estimates a low pupil diameter to avoid mental underload. If the classifier determines a complexity change during a multiplication task, the new task complexity is adjusted to the next appearing multiplication. After every trial, participants filled a NASA-TLX [161] questionnaire to obtain subjectively perceived workload during the trials.

7.3 Exploratory Results

Overall, the *easy* multiplication condition consisted of 618 displayed calculations, the *difficult* condition 420 displayed multiplications, and 466 displayed calculations for the *adaptive* condition. Least errors were measured during *easy* multiplication, comprising 571 correct (92%) and 47 wrong (8%) answers. Most errors were made during *difficult* trials with 255
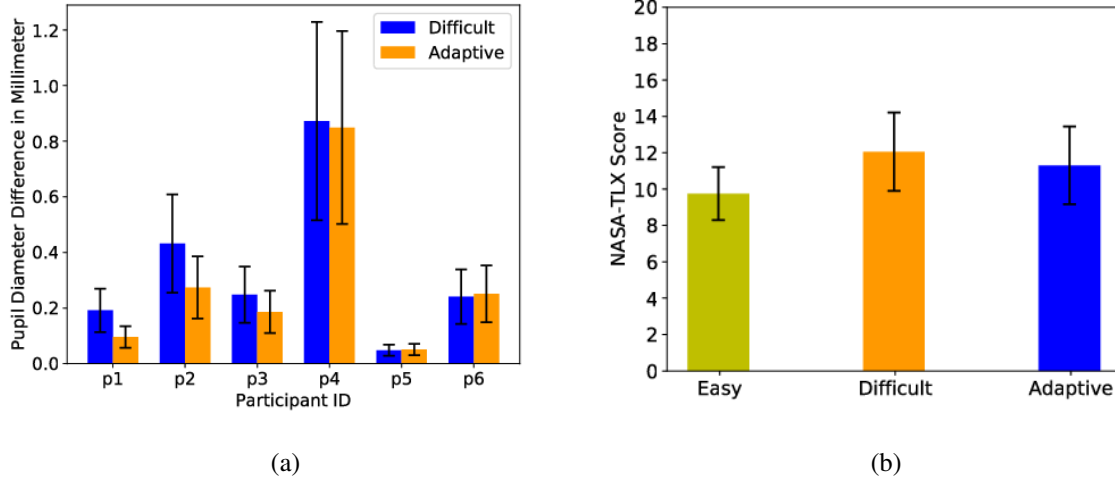
Figure 7.3: (a): Averaged differences of pupil diameter per participant between *easy* and *difficult* task complexities as well as *easy* and *adaptive* task complexities. The difference between the *easy* and *difficult* task complexity is higher compared to the *adaptive* task complexity except for two participants (p5, p6). The whisker lines depict the standard error. **(b):** Averaged NASA-TLX scores for the different task difficulties. The whisker lines depict the standard error.

correct (61%) and 165 wrong (39%) answers. The *adaptive* condition placed itself in between with 313 correct (67%) and 153 wrong (33%) answers.

We investigate individual differences in the pupil dilation measures between the different task complexities. The differences in the *easy* and *difficult* task complexity ranged between 0.05 and 0.87 millimeters. The difference between the *easy* and *adaptive* complexity ranged between 0.05 and 0.85 millimeters. Figure 7.3a shows the difference between the averaged pupillary measurements per participant. The mean number of complexity switches during the *adaptive* condition results in 2.6 complexity changes for a total of 466 multiplication tasks during the *adaptive* condition. The mean NASA-TLX score (see Figure 7.3b) reveals lowest subjectively perceived workload during the *easy* condition (M = 9.75, SD = 1.45) and highest workload measurements during the *difficult* complexity (M = 12.05, SD = 2.15). The *adaptive* condition was rated between both baseline trials (M = 11.30, SD = 2.14).

7.4 Discussion

In this study, we explored the impact of adapting the complexity of a cognitive workload inducing task based on the measurement of pupillary changes.

7.4.1 Adaptive User Interfaces

Our findings support the use of pupil dilation to detect mental effort. Assessing mental work-load in real-time to change the task complexity has shown its feasibility when participants performed adapted math tasks. The adaptive trial shows how computing systems can make use of the pupil diameter to create engaging user interfaces while avoiding the perception of high or low mental workload.

7.4.2 Creating User Specific Models

The study shows the viability of individually trained classifiers. Since physiological measurements differ among users, person-dependent measurements cannot be avoided. Limited to pupillary data, encouraging results were achieved by allocating short training times to create a user-specific model. This is supported by averaged pupil diameter measurements and subjective feedback provided through NASA-TLX questionnaires.

7.5 Study and Chapter Conclusion

This chapter investigated the assessment of cognitive workload based on pupillary data to change the complexity of a task in real-time. In a preliminary study, we trained person-dependent classifiers through math tasks with different complexities. During an *adaptive* trial, task complexity was changed depending on the classification of pupillary data. The results show the practicability of using pupillary data in controlled environments to evaluate user interface adaptation mechanics and user interface assessment.

Eye gaze-based properties reflect cognitive states. In this chapter, we investigated the suitability of pupil dilation as an alternative workload measure. A simple classification scheme revealed reliable discrimination between low and high workload states for difficulty alternations, hence assessing pupil dilation as a suitable measure for cognitive workload (**RQ4**). Performance and subjective reports imply that more engagement and less frustration is fostered when alternating between task difficulties. In contrast to smooth pursuit, dilation-based workload assessments do not require a moving object to be detected. However, while not being affected by moving objects, the pupil size is sensitive to light conditions. The presented experiment leveraged a controlled setup to avoid bias through lighting conditions. A potential solution is the normalization of pupil dilation through in-situ light measurements. The combination of lighting conditions and pupil dilation can provide a decent estimate for the currently perceived workload.

IV

APPLICATIONS

Chapter 8

Design Pipeline Evaluation

> *For people without disabilities,*
> *technology makes things easier. For*
> *people with disabilities, technology*
> *makes things possible.*
>
> IBM Training Manual

So far, we have investigated communication guidelines for workload-aware interfaces and measurement modalities that assess cognitive workload in real-time using physiological measures. This provides an implicit measure to evaluate interactive systems regarding the mental demand placed on the user and adaptive systems that benefit from this insight. In this chapter, we incorporate the results from Chapter 3, 4, and 5 in a field study. To recapitulate, the previous chapters investigated design implications for assistive technologies and presented an EEG-based approach to evaluate these. Persons with cognitive impairments were investigated as a target group as they provide explicit cues when cognitive assistance is needed.

In Chapter 3, cooking was identified as an important communal activity in the context of sheltered living [233]. The preparation of a meal represents a common task that is achieved by working together. Hence, social skills are conveyed and the awareness of fulfilling duties is invigorated. Cooking transfers a beneficial skill set for future independent living. However, uncommon tasks (*e.g.*, operation of novel kitchen devices) are still perceived as challenging by persons with cognitive impairments. Expert labor collaboratively supervises the cooking process but is scarce due to a worker shortage in the field. Assistive computing systems are therefore more commonly employed in sheltered living facilities.

Our previous research in Chapter 4 showed that visual feedback of assistive systems was favored over auditory and tactile feedback by persons with cognitive impairments [228]. In an assembly scenario, we evaluated the working memory capacities using EEG and found that visual in-situ projections decrease working memory strains. In this chapter, we combine

the concepts of (a) cooking, (b) visual in-situ feedback, and (c) assistive computing for persons with cognitive impairments within a field study in a realistic context. Specifically, we deployed a projection-based assistive system over two weeks that aims to support a cooking process.

The chapter reports on a user study that investigates cooking performances between *caretaker assistance* and *in-situ assistance*. We then investigate differences in subjective feedback about the individual cooking experience with each feedback modality. Finally, we discuss how a combination of *caretaker assistance* and *in-situ assistance* enables independent cooking for people with cognitive impairments. We believe that the presented concepts translate on other chores and address, therefore, the following research questions:

→ **RQ5**: Does in-situ feedback provide cognitive support during a cooking task?

This section is based on the following publication:

- Thomas Kosch, Kevin Wennrich, Daniel Topp, Marcel Muntzinger, and Albrecht Schmidt. The Digital Cooking Coach: Using Visual and Auditory In-situ Instructions to Assist Cognitively Impaired During Cooking. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '19, pages 156–163, New York, NY, USA, 2019. ACM

8.1 Related Work

The impact of deploying assistive technologies in real-world scenarios has been investigated by past research. We summarize relevant work in the following.

8.1.1 Supporting Persons with Cognitive Impairments

Assistive technologies that focus on support for the elderly and persons with cognitive impairments were the subject of previous research. Assistive systems require design implications and guidelines to coalesce kitchen environments with home environments. For this purpose, Pollack et al. [307] explored how assistive technologies in home environments can be designed to support people with cognitive impairments. This includes a taxonomy for assistive computing systems. Bouchard et al. [40] developed and evaluated a plan recognition framework for smart homes using microsensors. These sensors made contextual data available [245] of which people with dementia and caretaker personnel benefited from activity recognition provided by the framework. Furthermore, Arcelus et al. [8] used a variety

of sensors, such as microphone arrays and accelerometers, to collect contextual information. The collected data was used to train the artificial intelligence that provided individual context-aware assistance. Mihailidis et al. [267] found that the integration of training activities into daily life tasks, such as washing hands, reduced the development of dementia. Serious games that simulate regular daily kitchen tasks have shown positive effects on people with cognitive impairments [258]. Ethical aspects need to be considered before using assistive technologies in real-world environments since, depending on the level of impairment, individuals are limited in their ability to give consent [349]. However, the integration of assistive technologies raises ethical considerations regarding individual autonomy in home environments. This includes privacy concerns [81, 82] and user acceptance [48, 92].

Providing *in-situ assistance* at industrial workplaces through spatial augmented reality has shown mental alleviation for people with cognitive impairments [67, 128]. Projecting visual in-situ information enhanced the overall assembly efficiency regarding the number of errors and task completion time [136]. Furthermore, fewer cognitive resources were utilized [127]. Motivation can be increased or maintained by incorporating gamification into the working environment of persons with cognitive impairments [212, 213]. They found that visual and auditory notifications outperform tactile error alerts concerning usability and understanding.

8.1.2 Assistive Technologies in Smart Kitchens

Integrating assistive technologies in kitchen environments has been the focus of various researchers before. Scheible et al. [334] presented how smart kitchens can be designed to enhance the overall cooking experience. With the availability of microsensors, a huge number of contextual data in kitchen environments can be collected. Thereby, Hashimoto et al. [163] analyzed behavioral data to recognize the current cooking action and prepared food materials. Blasco et al. [34] developed and evaluated a smart kitchen system for older adults within real-world scenarios. They found that their participants benefited from cognitive assistance during the cooking sessions. Besides providing support during meal preparation, smart kitchens can provide contextual nutrition-aware information. Based on this, suited recipes can be mediated to ensure the intake of important nutritions [73, 72]. Cooking has become a social activity in which recipes are shared and prepared together. Therefore, Schneider et al. [341] present the "Semantic Cookbook" which can capture, share, and exploit cooking experiences semantically. This way recipes can be recorded shared among other smart kitchens and relatives. Since most recipes are passed down to the next generations, Terrenghi et al. [360] developed the "Living Cookbook". Cooking experiences were recorded to practice meal preparation techniques and to teach people who are unfamiliar with cooking. Previous research has incorporated language learning tasks into smart kitchens [171]. Design recommendations were presented for integrating task-based learning within the meal preparation process. Miyawaki et al. [272] developed cooking support for people with higher brain dysfunction. The use of their system during rehabilitation training has shown positive effects. Prototyping new smart kitchen solutions require a robust testing environment. Olivier et al. [288] presented a prototype setting in which new smart

kitchen solutions can be evaluated. Several augmented reality-based applications in smart kitchen environments were evaluated by Bonanni et al. [38]. This included the assessment of usability, attention, cognitive workload.

Previous work has invested effort into the construction, evaluation, and integration of assistive technologies in home environments. However, the influence of in-situ cooking assistance for persons with cognitive impairments has not been considered yet. To close this gap, we conducted a field study in a sheltered living facility where people with cognitive impairments cooked either with (a) traditional *caretaker assistance* or (b) *in-situ assistance*. We report on differences in cooking performance and subjective feedback between both cooking modalities.

8.2 User Study

Related research has informed how accessibility can be ubiquitously integrated into daily life situations to compensate cognitive impairments. At the same time, assistive technologies have proliferated into home environments that are available for the wider public. However, how assistive technologies impact the behavior of persons with cognitive impairments in sheltered living facilities during cooking has not been explored yet. In the following, we present a study that compares *in-situ assistance* and *caretaker assistance* in terms of cooking performance as well as subjectively perceived feedback. This chapter and the subsequent study aims to answer the following two research questions:

- **RQ 1:** How does *in-situ assistance* changes the cooking performance, measured by the overall meal preparation time, compared to *caretaker assistance*?

- **RQ 2:** How is *in-situ assistance* subjectively perceived compared to *caretaker assistance*?

8.2.1 Evaluation

We extended the system of Funk et al. [136, 129] by functional modules to provide additional visual and auditory feedback. The main system that tracks cooking steps and provides feedback to the cooking area consists of multiple aluminum profiles assembled. The profiles enable to mount different hardware on top of the cooking area by placing the construction on the cooking area. The system uses a top-mounted Kinect v2[1] which is mounted 1.35 meters above the cooking area to detect finished cooking steps. A projector that is mounted 1.50 meters above the working area presents visual cooking instructions [128]. Dedicated audio speakers, placed next to the cooking area, provided auditory feedback by playing verbally

---

[1] `https://developer.microsoft.com/en-us/windows/kinect` - last access 2020-01-28
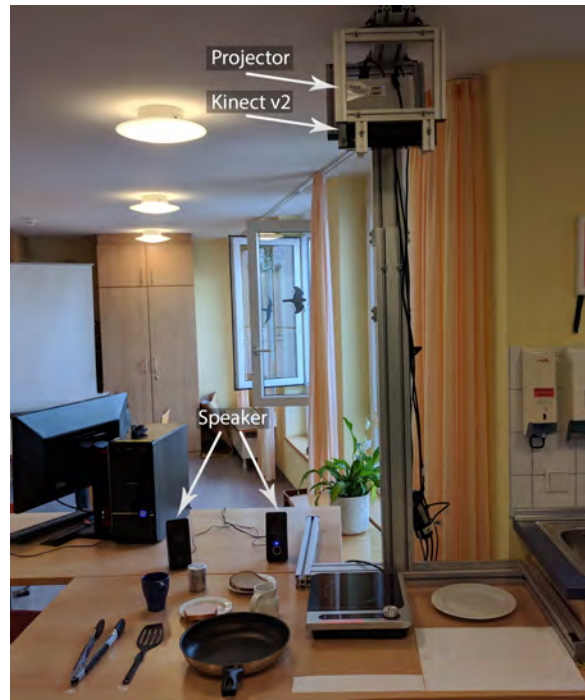
Figure 8.1: Study setup with pre-portioned ingredients. Auditory feedback is provided via external speakers while visual feedback is delivered by projections.

recorded cooking instructions (see Figure 8.1). In the following, we describe the cooking feedback modalities which were facilitated in the study.

Auditory Instructions

Auditory instructions are the current communication standard for single cooking steps between tenants and caretakers in sheltered living facilities. Most tenants are not able to read and rely on the verbal exchange of instructions. To provide a similar experience through *in-situ assistance*, we recorded single cooking steps with the voice of a caretaker. External audio speakers, set to a suitable volume, were integrated into the cooking environment (see Figure 8.1). Cooking instructions were played back with each start of a cooking step. We used a Trust 2.0 speaker setup[2].

Video Instructions

Video instructions were provided by a projector mounted above the cooking environment. The videos were short looped clips, where caretakers showed how to accomplish the next cooking step. The video instructions were projected on a dedicated field right to the cooking

---

[2] `www.trust.com/en/product/17595-remo-2-0-speaker-set` - last access 2020-01-28

(a)   (b)

Figure 8.2: Visual output modalities of the cooking instruction system. **(a):** In-situ projections of video instructions and progress bars were displayed inside the cooking area. The projections were placed on white sheets to enhance the visibility. **(b):** In-situ projection of video instructions and contours which provide visual cues for grasping or placing objects and food.

area (see Figure 8.2a). Upon step completion and validation through the Kinect v2, the next instruction video was played back. An Acer K330 was used as projector[3].

Contour Projections

Contour projections were displayed on or around objects which are of interest during the current cooking step. This is accomplished by projecting a green light on different cooking utilities (see Figure 8.2b). Furthermore, a progress bar below the hotplate is projected during waiting times (*e.g.*, when frying ingredients). Figure 8.2a shows the position of the progress bar.

8.2.2 Methodology and Measures

We employ a within-subject design study including the independent variable *assistance modality* consisting of cooking instruction provided by *in-situ assistance* and instructions provided by *caretaker assistance*. The prepared meal was meatloaf with fried eggs and bread. We have chosen this recipe because it was unknown among the participants. Thus, the overall cooking process was unknown to the participants. The runtime of the experiment was ten days (*i.e.*, two weeks without weekends). Two cooking sessions with two different participants were conducted on each day, where one participant cooked with *caretaker assistance* and another participant with *in-situ assistance*. After the first week, the same participants from the last week were invited on the same weekday again with the contrary assistance modality, *i.e.*, participants who used to cook with *caretaker assistance* were either assisted by *in-situ assistance* and participants who were assisted by *in-situ assistance* used *caretaker assistance* instead. By introducing the break of seven days, we reduce the probability that participants remember the cooking procedure when they cook the same meal again using the contrary instruction modality. In other words, five participants cooked with

---

[3] `www.projectorcentral.com/Acer-K330.htm` - last access 2020-01-28

in-situ assistance and another five with *caretaker assistance* during the first week. In the second week, participants that started with *in-situ assistance* were invited to cook with *caretaker assistance*. In contrast, participants who cooked with *caretaker assistance* in the first week used *in-situ assistance* in the second week. Note, that all participants in the second week were already aware of the recipe and the cooking procedure from the first week. The required cooking utilities were placed on predefined positions before the experiment. We ensured a similar positioning of cooking utensils for each session.

We measure the meal preparation time for both assistance modalities to investigate temporal differences between both assistance modalities. We subtracted constant waiting times, such as cooking steps that require frying, from the overall task completion time (see Table 8.1). By this, individual waiting times between *caretaker assistance* and *in-situ assistance* are removed from the analysis. Overall, four cooking steps facilitated waiting times (*i.e.*, heating oil, frying meatloaf from both sides, and cooking the egg) during one trial. Table 8.1 shows each cooking step with their accompanied instruction system for *in-situ assistance*. Finally, we conducted semi-structured interviews with the participants about personal preferences in cooking assistance after the experiment.

8.2.3 Procedure

Before the study, we asked for written consent from either the participants or their legal guardians. We conducted the study in a kitchen within a sheltered living facility. We carefully explained the intention of the study to the participants to avoid misunderstandings. Since the system was unfamiliar to the tenants, we made them familiar with the visual and auditory instructions. Furthermore, we showed where the feedback cues were generated to avoid confusion for participants during the study. After being familiar with the system, participants started with the cooking instruction modality according to the balanced Latin square. Before the experiment, ingredients were pre-portioned since tenants affected by motoric disorders could not handle the doses by themselves. The experiment started after ensuring all safety arrangements.

Visual instructions were presented alongside auditory feedback for each step. A voice recorded by a caretaker gave instructions on how to perform the current cooking step. Additionally, a video was projected to the right of the cooking area. These videos had a length between two and four seconds that demonstrated how the current cooking step has to be performed. Furthermore, objects of interest were highlighted by contour projections. A Kinect v2 detected whether cooking steps were performed successfully. If a cooking step was conducted correctly, the visual, auditory, and contour instructions proceeded to the next step. The whole cooking procedure comprised 25 cooking steps including four waiting steps to fry the ingredients (see Table 8.1). Since waiting times may occur during frying steps, short cartoons were projected into the video area. This ensured engagement during waiting times since the participants' concentration may be affected and reducing the likeliness of returning to the cooking task [362]. At the same time, a declining progress bar indicates the

remaining waiting time (see Figure 8.2a). Notes about the interaction between participants and systems were recorded during the experiment. A caretaker was always present during the cooking sessions to ensure the safety of participants. Participants were lauded at the end of the cooking session.

Afterward, participants participated in a semi-structured interview. The participants were asked about their experience when cooking with *caretaker assistance* or *in-situ assistance*. Furthermore, we asked about suggestions for improvements. The answers were recorded and noted by the experimenters. Overall, the study took approximately 30 minutes including meal preparations and post hoc interviews.

8.2.4 Cooking Equipment

A hotplate and a pan are used to heat the ingredients. We use an induction burner as a hotplate for safety reasons. A spatula and a reacher were used to put and turn ingredients in the pan. Finally, fried ingredients were placed on a plate right to the cooking area (see Figure 8.2). The whole cooking procedure uses pre-portioned ingredients including a piece of meatloaf, salt, eggs, oil, and a piece of bread.

8.3 Results

We analyzed the measures as well as qualitative data collected throughout the study. The results are presented in the following.

8.3.1 Participants

We recruited ten tenants of a sheltered living facility (6 female, 4 male) aged between 24 and 56 years (M = 40.9, SD = 9.93). The participants had cognitive impairments that limit their ability to process and understand information. None of the participants were affected by dementia. Five caretakers (3 female, 2 male) aged between 19 and 34 years (M = 26.8, SD = 7.05) were involved for the *caretaker assistance* condition. Their individual work experience ranged between several months to twelve years. We constantly assigned each caretaker to two participants. We ensured that the same caretaker was cooking with similar participants during the *caretaker assistance* and observed them during *in-situ assistance*. Participants were monetary compensated through funding from a joint project.

8.3.2 Cooking Performance

We compare the cooking performance between the conditions *caretaker assistance* and *in-situ assistance* regarding the time participants required to prepare a meal. We process the

Step No.	Cooking Step	Audio	Video	Contour
1	Take pan	X	X	
2	Pan on hotplate	X	X	
3	Start hotplate	X	X	X
4	Set heat	X	X	X
5	Take oil	X	X	
6	Put oil into pan	X	X	
7	Wait until pan is hot	X		
8	Take reacher	X	X	
9	Take meatloaf	X	X	
10	Put meatloaf into pan	X	X	
11	Wait until meatloaf is fried	X		
12	Turn meatloaf	X	X	
13	Wait until meatloaf is fried	X		
14	Put meatloaf on plate	X	X	X
15	Put reacher on table	X	X	
16	Take cup with egg	X	X	
17	Put egg into pan	X	X	
18	Wait until egg is fried	X		
19	Take spatula	X	X	
20	Put egg on plate	X	X	X
21	Turn hotplate off	X	X	
22	Take bread	X	X	
23	Put bread on plate	X	X	X
24	Put salt on meal	X	X	
25	Completion notice	X	X	

Table 8.1: Single cooking steps with their assigned instruction systems. The steps 7, 11, 13, and 18 (gray rows) consist of waiting times.

data by subtracting the fixed waiting times from the overall cooking times of *in-situ assistance* and *caretaker assistance* (see Table 8.1). *In-situ assistance* comprised a mean waiting time of seven minutes and eight seconds while *caretaker assistance* had a mean waiting time of eight minutes.

First, we statistically analyze the cooking times between *caretaker assistance* and *in-situ assistance* which were collected during both weeks. A Shapiro-Wilk test did not reveal a deviation from normality, $p > .05$. Thus, we submitted the data to a paired samples t-test and found a significant effect between both assistance modalities, $t(9) = 3.628$, $p = .006$, $d =$

1.147. Investigating the mean cooking times between both conditions shows longer cooking times for *in-situ assistance* (M = 236.3, SD = 46.24) compared to *caretaker assistance* (M = 228.9, SD = 77.03). Figure 8.3a shows the aggregated mean times of the first and second week between both assistance modalities.

Since either five participants cooked with *in-situ assistance* or *caretaker assistance* per week, we separately investigate the cooking times between both conditions for the first and second week. Note that the participants in the second week were aware of the cooking procedure since they already prepared the meal in the first week with the contrary assistance modality. A one-way ANOVA revealed a significant main effect between the cooking times for each week and instruction modality, $F(3, 16) = 3.504$, $p = .04$. A Tukey post hoc test revealed a significant effect between the use of *in-situ assistance* in the first week and *caretaker assistance* in the second week, $p = .024$, $d = 2.511$. However, we could not find a significant effect when using *caretaker assistance* in week one and *in-situ assistance* in week two. Furthermore, there is no significant effect when using *in-situ assistance* or *caretaker assistance* in the same week. Participants required more time when using *in-situ assistance* in the first week ($M = 262.8, SD = 34.77$) compared to the second week using *caretaker assistance* ($M = 159.4, SD = 46.7$). *Caretaker assistance* required less time in the first week (M = 206.8, SD = 70.42) compared to *in-situ assistance* in the second week (M = 209.8, SD = 42.96). Figure 8.3b shows the separated cooking times of *in-situ assistance* and *caretaker assistance* per week.

(a) (b)
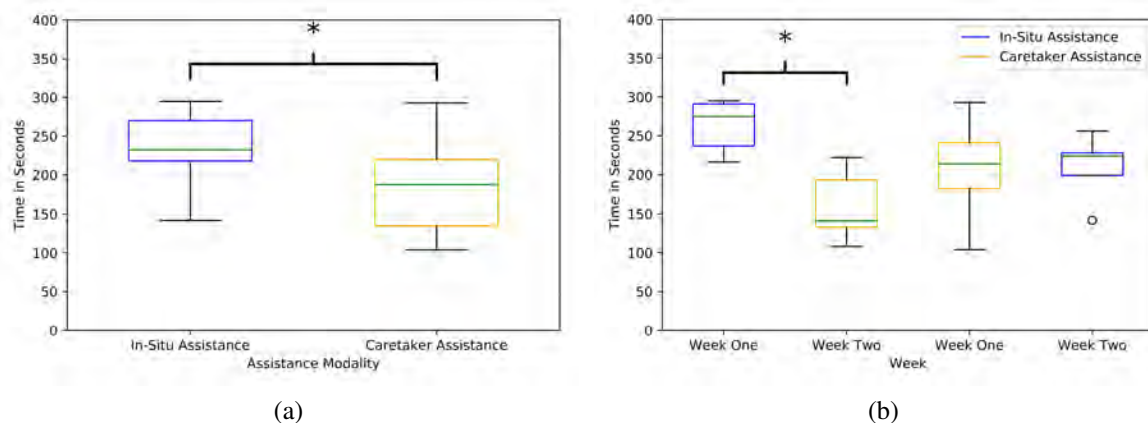
Figure 8.3: Results of the study. **(a):** Mean cooking times between *in-situ assistance* and *caretaker assistance* for both weeks. Brackets indicate significant differences. **(b):** Mean cooking times for *in-situ assistance* and *caretaker assistance* for each week. Participants that started with *in-situ assistance* had higher cooking times in the second week using *caretaker assistance*. Brackets indicate significant differences.

8.3.3 Subjective Feedback

We collected qualitative comments from the participants about the experience with traditional *caretaker assistance* and *in-situ assistance*. In the presence of the caretakers, we asked participants if the cooking procedure was pleasant for them. Participants who cooked with *caretaker assistance* and *in-situ assistance* enjoyed cooking. However, one participant stated that "[. . .] *the videos were not helpful*" (P4) while others found that the "[. . .] *videos were funny*" (P9) and that "[. . .] *the contour projections were engaging*" (P7).

Furthermore, we asked participants if they found the visual and auditory feedback helpful. All participants except one (P4) found the projected videos helpful. Four participants stated that they noticed them when they were stuck in the cooking process (P2, P4, P5, P10). The sound was perceived by all participants positively except one participant who stated that "*The voice was not helpful, the videos helped me more.* " (P9).

Finally, we asked if the participants could imagine integrating such a system into their daily communal cooking procedure. All participants except one (P8) agreed with an integration of the system into their daily life. However, one participant mentioned the importance of social interaction during cooking activities (P3). One of the participants stated, that the system "[. . .] *provided more safety when cooking alone*" (P2). One participant preferred to cook with a caretaker instead of the system (P8).

8.3.4 Qualitative Observation

We observed the interaction between the tenant and the assistive system in the kitchen. We noted, that all participants were able to perform most of the cooking steps by listening to the voice. Video instructions, which were displayed in every cooking step, were perceived when the voice instruction was not understood. The tenants are used to voice-driven instructions since cooking instructions are communicated verbally during the cooking process. A similar observation was found in the perception of the progress bar, which was noticed after they were hinted by the experimenter or caretaker. Four participants required help to understand the system before conducting the first cooking steps. Afterward, the cooking process was guided by the system without interruptions.

8.4 Discussion

We evaluated an assistive cooking system in a sheltered living facility. In the following, we discuss the results of our study.

8.4.1 Cooking Performance and Learning Effect

Overall, we measured longer meal preparation times using *in-situ assistance* compared to *caretaker assistance*. However, we find that the cooking times with *caretaker assistance* in the second week improves significantly when *in-situ assistance* was employed in the first week. Since cooking instructions are usually conveyed verbally, the visual feedback generated by in-situ assistance might complement the cognitive processing of instructions. Therefore, a combination of visual and auditory cooking instructions could result initially in a longer cooking procedure, but can also contribute to an improvement in understanding the cooking process. Furthermore, a learning effect can eventuate as a result of using *in-situ instructions*. We could not observe this effect when starting with *caretaker assistance* in the first week and continuing with *in-situ assistance* in the second week. This assures the hypothesis that visual feedback can be used as a complementary feedback modality to help tenants to remember or understand complex cooking instructions.

8.4.2 Feedback Modalities

From our observations, we find that most participants could handle simple cooking instructions using auditory feedback. Video and contour projections were considered during waiting times or more complex steps, such as turning ingredients in the pan. These results conform with the results of the Chapter 3, where auditory instructions are preferred to convey and process information. However, most participants considered video and contour instructions useful to understand details although not always needed. We believe that video and contour instructions supplement auditory feedback for rather complex cooking instructions.

8.4.3 Subjective Perception

We asked caretakers and the tenants about their perception of cooking with assistive technologies. The responses were positive by both caretakers and tenants. Participants stated that visual and auditory feedback complement each other in terms of understanding the current cooking step. Since tenants are usually not able to cook without *caretaker assistance*, most of the participants enjoyed their autonomy during the cooking process using *in-situ assistance*. This indicates that suitable feedback content is a critical factor to consider when conducting further research on assistive technologies in private and sheltered living facilities.

8.5 Study Conclusion

We presented a study which compares *caretaker assistance* and *in-situ assistance* during several cooking sessions for tenants with cognitive impairments. We found that *in-situ assistance* requires more meal preparation time in the first week compared to *caretaker assis-*

tance. However, the cooking performance of *caretaker assistance* improves fundamentally when the cooking procedure was conducted with *in-situ assistance* beforehand. Through semi-structured interviews, we find that the visual and auditory feedback provided by *in-situ assistance* is well accepted among tenants. We believe that such assistive technologies can be used in the future to compensate for cognitive impairments. Learning how to accomplish daily life tasks by assistive technologies goes a step towards independent living for tenants with cognitive impairments.

We are aware that our research is prone to several limitations. We did not consider the individual cognitive impairment per participant. Before the study, we ensured that the level of cognitive impairment was similar for each participant. However, assessments to estimate the individual level of cognitive impairment are conducted irregularly. As mental abilities develop further, the assessment may become irrelevant regarding individual impairment. Furthermore, we did not consider the individual level and type of cognitive impairment in our study design. Thus, the results may not be generalizable to people that are affected by other cognitive impairments, such as dementia or motoric disorders. For safety reasons, a caretaker was always present during the study regardless of the cooking assistance modality. Thus, we did not investigate how individual behavior may change when a caretaker is not present. Therefore, it is not clear how people with cognitive impairment interact with assistive technologies in their home environments when caretakers are not nearby. To foster research in this area, we host a reference implementation of the presented workload-aware assistive system on Github[4].

8.6 Chapter Conclusion

This chapter presented a case study in which visual in-situ projections were used to guide persons with cognitive impairments through a cooking process. The design of the assistive system followed a user-centric development approach: after recognizing opportune moments for in-situ assistance through qualitative inquiries, we evaluated suitable feedback modalities that were considered during the implementation of an in-situ assistance system. To answer **RQ5**, we conducted a long-term study over two weeks in a sheltered living facility. The study revealed that no improvements in task completion time and error rate were immediately visible. A potential reason could be the unfamiliarity with the system itself, whereas tenants are used to cooking with their respective caretakers. In contrast, in-situ support was positively perceived and well rated by the participants. Since all participants completed the cooking sessions, we believe that user-tailored assistive technologies will foster autonomy for people with cognitive impairments.

---

[4] `www.github.com/hcum/kobelu` - last access 2020-01-28

Chapter 9

Reflection

In order to improve the mind, we ought less to learn than to contemplate.

René Descartes

EEG has shown to be a feasible tool for workload quantification. In chapter 5, we have conducted a study to develop a pipeline that uses EEG as a physiological real-time measure that evaluates the workload placed on users by the respective interface. Such a pipeline enables the development of neurofeedback loops that utilize cognitive workload for feedback mechanisms. However, the brain is a sensitive organ that is highly unexplored. The raw EEG signals are difficult for novice users to interpret and process.

In contrast, BCIs are a category of physiological sensors that are currently coping with these challenges to make their way into the consumer market. Traditionally, they have been used in the medical field to help patients with severe disabilities to gain more control over their lives. EEG signals from the brain, gathered through electrodes placed on or inside the scalp, can be interpreted to allow patients to move a wheelchair, a cursor [356, 391] or provide feedback to the user in what is called a neurofeedback loop. Neurofeedback training involves interpreting brain signals in real-time and providing the user with feedback about their state through different stimuli, for example, using VR environments. VR in combination with BCIs have been researched and used in several applications ranging from treating brain damage [318], substance abuse [239] to mindfulness training [5]. Hence, such BCI applications abstract the complexity of EEG and enable users to utilize brain signals for implicit and explicit interaction purposes. Using new, light-weight, and less complex consumer BCI devices coupled with a VR system allows us to create adaptive virtual environments that can change according to the measured brain signals to increase or decrease activation in certain parts of the brain. This can especially help people with neuropsychiatric diseases (*e.g.* depression) and their doctors to investigate brain activation in real-time during VR-based neurofeedback sessions. The desire to investigate the impact of VR scenes on the brain is high but becomes complicated because of the linearly increasing output of an EEG over time. Whereas

the measured brain activity at certain electrodes is what drives the neurofeedback system, the origin of the electrical sources of activation inside the brain is also required to analyze which parts of the brain are affected by the VR scene. The low-resolution electromagnetic tomography algorithm (LORETA) solves this problem for a sufficient number of electrodes in a three-dimensional space [294].

In this chapter, we present a simple and understandable real-time visualization of the mentioned LORETA algorithm. A simulation is executed to show the correctness of the implementation. A study is conducted to show that changes regarding brain activation can be observed in real-time. This is complemented by an evaluation using a VR-based neurofeedback loop that reveals the affected brain areas. This chapter showcases the potential that EEG holds for neurofeedback platforms for future UI adaption systems. Therefore, we investigate the following research question:

> → **RQ6**: How can complex physiological signals be visualized to be utilized by non-expert users?

This section is based on the following publication:

- Thomas Kosch, Mariam Hassib, and Albrecht Schmidt. The Brain Matters: A 3D Real-Time Visualization to Examine Brain Source Activation Leveraging Neurofeedback. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 1570–1576, New York, NY, USA, 2016. ACM

9.1 Related Work

Visualizations of brain data to interpret neuronal activity is an important tool for researchers. In this context, research has been conducted that can be broadly summarized into two categories: (1) the impact of virtual reality on the brain in medical treatments and (2) currently existing brain data visualizations.

Baumgartner et al. [23] have shown that neurophysiological changes can be found when using VR. Increased brain activity at the parietal and frontal lobe could be found. The EEG analysis of the displayed VR scene was carried out using LORETA. However, the output cannot be inspected in real-time and the implementation is not interactive. Grealy et al. [151] performed a 4-week intervention program for treating patients with traumatic brain injuries using VR. Significant cognitive increases could be noticed in terms of speech, visual learning and reaction times. Improvements were quantitatively measurable, but the affected brain parts by this training could not be observed at all. Hoffman [168] used functional magnetic

resonance imaging (fMRI) to investigate perceived pain when treating burn damages in combination with and without VR. The fMRI has shown that distraction delivered by VR helps to reduce the amount of brain activity in pain-related brain areas. The brain scans presented in the study still lack real-time capabilities to investigate the augmentation and impact of brain activation throughout a VR session. A three-dimensional visualization is presented by the project Glass Brain from the company Neuroscape[1]. It enables the visualization of source localization in real-time with interactive features. However, the inner of the visualized brain is hard to investigate since it is filled with a fixed solid color.

Overall, the presented brain visualization solutions lack interactive features, real-time capabilities or suffer performance issues. In contrast, our work shows a simple visual representation of the activities in the brain which allows a human observer to understand the current brain activation in real-time. Currently, we investigate this in two contexts: (1) Providing information to improve medical treatment of brain-related issues, (2) assessment of reactions to visual stimuli in the context of human-computer interaction. In our work, we propose a real-time implementation of LORETA including a three-dimensional representation enriched with interactive features to allow more precise investigations of brain parts. As a proof of concept, we performed a pilot study with an adaptive VR scene which changes according to the measured brain activities to achieve a neurofeedback session. Changes in brain activation are investigated during VR sessions in real-time as well as in post hoc analysis.

9.2 System

The implementation of our algorithm uses neuromore Studio[2] as base platform. Together it builds a system allowing us to retrieve data measured by a BCI device and processing it with our implementation in real-time. Furthermore, we can create adaptive VR-based neurofeedback sessions based on the measured EEG data[3].

9.2.1 Localizing Electrical Source

We use a multithreaded implementation of the standardized LORETA (sLORETA) algorithm for electrical source localization [295]. sLORETA allows precise source localization in a three-dimensional space with different settings regarding noise minimization, voxel density and the usage of a head model. The input of the algorithm is the current voltages of extracranial scalp measurements received by the attached electrodes. sLORETA then calculates a

---

[1] `www.neuroscape.ucsf.edu` - last access 2020-01-28

[2] `www.neuromore.com` - last access 2020-01-28

[3] The visualizations are available in this publication: Thomas Kosch, Mariam Hassib, and Albrecht Schmidt. *The Brain Matters: A 3D Real-Time Visualization to Examine Brain Source Activation Leveraging Neurofeedback*. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 1570–1576, New York, NY, USA, 2016. ACM

density distribution throughout a head and associates each voxel with a magnitude that describes the current density power at the voxel. Voxels are color-coded depending on the calculated magnitude ranging from blue (weak) over green (medium) to red (strong).

9.2.2 Visualizing Results

After retrieving electrode measurements, the algorithm begins processing the data based on the chosen head model and the number of voxels. We have decided to use a spherical head model because it is easier to reproduce and compare results from other research projects. The number of voxels can be set during program runtime, while a higher number of voxels increases the precision of the result at the cost of performance. Interactive features like zoom and rotation are supported to ease the analysis of brain source activation. The computation of the visualization algorithm is GPU accelerated. In contrast to current systems, we can visualize the activated areas in real-time for the observer (doctor, researcher) to directly see the reaction inside the brain.

9.2.3 Evaluation

The implementation is evaluated using the BESA simulator tool[4] to verify the correctness of the results. The BESA simulator provides the functionality to place an electrical source inside a head. These can be exported as a CSV file, which allows an import into the sLORETA implementation. The evaluation showed that the algorithm is working correctly and can be used in future user studies.

9.2.4 Input Stimulus

Our system uses VR as an input stimulus which enables us to create controllable simulations of real-world situations where the reaction of body and mind can be analyzed. The simulation can be changed based on the measured electrode values to match a VR scene to the cognitive load of a user. Depending on the result of the EEG analysis, the difficulty of tasks or environments can be modified for brain training exercises, for example, to decrease frustration or boredom. Frequent training using VR with suitable parameters can be used to increase overall brain activity. Treatment of neuropsychiatric diseases often aim to increase brain activity in the frontal lobe, since decreased brain activity in these areas often is seen in patients of depression, schizophrenia, substance abuse or obsessive-compulsive disorders [359].

---

[4] `www.besa.de/downloads/besa-simulator` - last access 2020-01-28

9.3   Impact of VR

Virtual reality enables controllable simulations of real-world situations where the reaction of the body and mind can be analyzed. The simulation can be changed to suit the VR scene to the cognitive load of a user. Depending on the result of an EEG the difficulty of tasks can be modified for brain training exercises to decrease frustration or boredom. Therefore it is important to understand what kind of neurophysiological effect different VR tasks have.

VR is often related to video games that provide an immersive experience. Video games contain goals that have to be accomplished by a user. The frontal lobe, which is responsible for motivation-related thinking, is activated during such tasks [5]. Decreased brain activity in this brain area often involves depression, schizophrenia, substance abuse or obsessive-compulsive disorders because of dopamine lack or depletion [359]. Frequent training using VR with suitable parameters can be used to increase brain activity in the frontal lobe, which may be used to treat neuropsychiatric diseases. Therefore VR is a popular method used in neurofeedback sessions for frontal lobe training.

9.4   Study

A pilot study with eight male participants was conducted to find visible changes of electrical source activation within the brain during a neurofeedback session that used VR to change the stimuli. The average age was 27.12 years (SD = 5.19). The participants were healthy and have not participated in a bio- or neurofeedback study before. The study took approximately 40 minutes per participant.

9.4.1   Apparatus

We used an OpenBCI[5] board with 16-channel support as EEG device for the study. OpenBCI and its electrodes are mounted on a 3D printed retainer which can be worn on the head. An Oculus Rift[6] was used to display VR content. The visualization uses 8000 voxels to render the results.

9.4.2   Method

The study is divided into a baseline task, a VR session and a repetition of the baseline task to see visible differences before and after the VR session. The OpenBCI is continuously

---

[5] `www.openbci.com` - last access 2020-01-28

[6] `www.oculus.com` - last access 2020-01-28

recording EEG data to enable post hoc analysis. Changes in brain activation are observed in real-time.

The baseline task is conducted to visualize initial source activation. The participant is asked to keep the eyes open for three minutes, followed by three minutes of keeping the eyes closed. Participants were seated in a dimmed room facing a white wall. This is a common task to estimate initial electrical brain source activation [21].

The VR session displays an infinite tunnel which is traversed by the participant for ten minutes. The traversal speed is influenced by the activation measured by the sum of three electrodes at the frontal lobe. High alpha band magnitudes at these electrodes cause an increasing traversal speed while low voltages slowdown the traversal speed. We told the participant to traverse the tunnel as fast as possible. This training aims to increase brain activation in the frontal lobe. To achieve this, a signal processing algorithm is constructed which focuses on frontal lobe training. The electrodes at the positions F3, F4 and Fz are used to estimate brain activation at the frontal lobe. The measured electrode values are summed up and processed in a Fast Fourier Transform (FFT) to estimate frequencies in the range between 8 Hz and 13 Hz. The estimated frequency is normalized to the interval $[0; 1]$. The normalized values are used to set the traversal speed of the VR tunnel, where 0 causes the slowest traversal speed and 1 the fastest traversal speed. In the last step, the baseline task is repeated. This is required to detect differences in brain activation before and after the VR session.

9.5   Results

Results have shown that significant electrical brain source changes could be found within all participants. Different brain activations were observed during the eyes-opened and eyes-closed task. Higher brain activation is noticed during the eyes-closed task. The most active part in the back of the brain which is responsible for visual processing.

Before and after conducting the VR session, a snapshot of brain activation is taken for comparison purposes. The sLORETA visualizations show clear differences regarding brain activation before and after conducting the VR session. The brain shows higher brain activation before the session than the baseline task, which could be the effect of excitement since the participants have not participated in a VR related neurofeedback session before. After conducting the VR session, higher brain activation was measured. Repetition of the baseline task shows higher brain activation than in the first baseline run.

9.6   Discussion

Investigating brain source localization has shown that changes are visible and measurable when using VR as a stimulus. Increased brain activity is observed when participants started closing their eyes. Previous research has shown that this is a natural effect [260]. The back

of the brain is permanently active since it is responsible for visual processing. Closing the eyes pushes activity in other brain regions responsible for imagination and multi-sensory actions which explain enhanced electrical source localization during the eyes-closed task.

Considering the effect of VR as the stimulus has shown increased brain activity in different brain regions. The cerebellum and the frontal lobe have shown improvement in brain activation. These regions are responsible for motion sensing and motivation-related thinking. The changes in motion and speed in the tunnel explain brain activation differences in the cerebellum, while the excitement and the desire to speed up the tunnel could explain enhanced activations in the frontal lobe. The repetition of the baseline task has shown sustainable brain activity after the VR session. This is usually not a prolonged effect since permanent brain activation changes require extensive neurofeedback training. Neurofeedback sessions have to be repeated several times for long term lasting effects. The study has shown that real-time brain source localization can be used to understand the effect of stimuli on a neuropsychological level. Improved therapy of neuropsychiatric diseases and tailored enhancements of neurofeedback sessions are possible to increase the treatment efficiency.

9.7   Study and Chapter Conclusion

We presented an interactive three-dimensional real-time implementation of the sLORETA algorithm which calculates a density distribution to estimate the electrical origin of brain source activation. The usage of VR in medical areas has been investigated to estimate neuronal impacts on the brain using EEG devices. Our implementation of sLORETA is used in a user study to find out if it is possible to see visible changes in brain activation during a VR session. The results show that changes in real-time can be observed when using VR scenes as an input stimulus. The outcomes can be used to optimize medical treatment methods or to identify neuropsychiatric diseases. Our visualization proves visible changes regarding brain activation for different stimuli. We see this as a first step to create visualizations that can be interpreted by non-expert users (**RQ6**). We believe that neurofeedback can be used for the treatment of brain-related diseases like Alzheimer's, epilepsy or attention deficit hyperactivity disorders (ADHD). The use of our brain visualization as an analysis tool by doctors can help to identify upcoming diseases or track the progression of existing diseases. The collected dataset is hosted on Github[7].

---

[7] `www.github.com/hcum/the-brain-matters` - last access 2020-01-28

Workload-Aware Reading Interfaces

*No entertainment is so cheap as
reading, nor any pleasure so lasting.*

Mary Wortley Montagu

Reading resembles an activity that interprets written text for pleasure or information consumption. It is estimated that approximately *3,500* years B.C. were the first attempts of written communication that took place. Reading and writing itself was reserved for a small population of people. However, with the increase and availability of information daily, novel reading techniques emerged to filter and consume relevant information. One such technique is RSVP, a paradigm that displays text in a word-by-word manner to present information. Instead of requiring readers to move their eyes to fixate on individual words whilst reading, readers would fixate on the display center while the words of the targeted text would be presented one after another at a chosen frequency [119]. Doing so enables full text to be accurately presented on smartphones [142, 289], smartwatches [96, 97], or smart glasses [324], in spite of the available display area (see Figure 10.1a) [154]. Furthermore, removing the need to move one's eyes and the ability to raise the rate of word presentation above one's normal reading speed allows for clear time-savings, which provide the opportunity to increase the presentation rate for faster text processing. The primary display parameters of most RSVP applications are *Text Alignment* and *Presentation Speed* [322]. This is either fixed or adjustable to user preferences. Past research has shown that these two parameters influence perceived workload and text comprehension [27, 64, 321].

The overall aim of RSVP is to gain more reading speed by aligning the readers' eye fixation around an Optimal Recognition Point (ORP). Commercial applications, such as *Spritz*, adopted the concept of ORPs where users can adjust various settings. This can reduce cognitive load and result in better comprehension [234, 309], which is highly disputed in the community [27]. *Presentation Speed* determines the rate of the presented words. Previous research found that *Presentation Speed* is an individual factor that has to be adjusted ac-
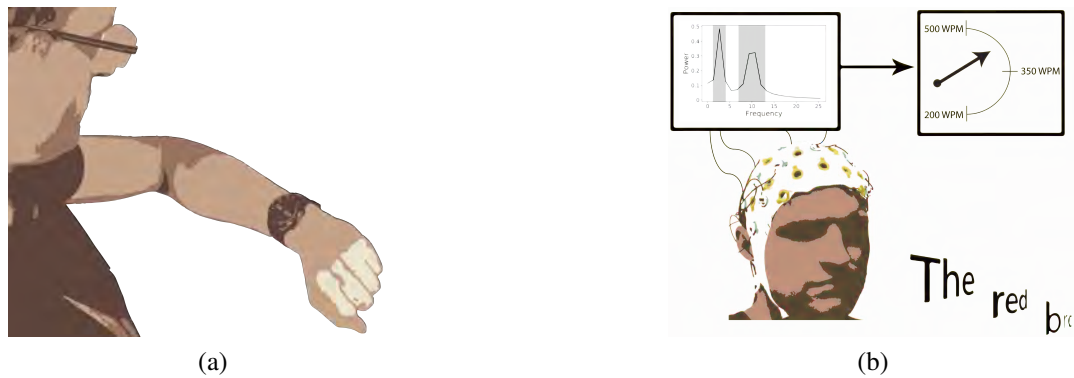
(a) (b)

Figure 10.1: (a): User reading on a device with small screen space. **(b):** Electroencephalography determines the current level of cognitive workload for different *Text Alignments* and *Presentation Speeds*.

cording to reading ability, text content, and individual preference [70]. Choosing unsuitable *Presentation Speeds* may have negative effects on the overall reading efficiency [68, 322]. For example, setting a high *Presentation Speed* may result in lower text comprehension and text retention since less time is available to process the presented text. In contrast, setting *Presentation Speeds* below the regular reading speed results in a loss of time. Since reading speed is an individual factor, it is difficult for novice users to estimate their optimal *Presentation Speed*. While users can modify the *Presentation Speed* during RSVP to estimate their preferred speed, the adjustment itself places additional workload on users which might compromise their assessment. Hence, the selected *Presentation Speed* remains static during the reading session. Given individual differences and variable user states for and throughout an RSVP reading session of even a single user, it is worthwhile to investigate the viability of developing a real-time measurement that could determine appropriate *Text Alignments* or *Presentation Speeds*.

In this chapter, we investigate the use of EEG as a direct metric for information processing whilst reading with RSVP with different *Text Alignments* and *Presentation Speeds* (see Figure 10.1b). EEG allows for real-time measurements of a user's brain activity with high temporal resolution across different scalp locations [282]. While text comprehension and subjectively perceived workload have previously been assessed with *post hoc* questionnaires for evaluating RSVP applications [98, 324], we propose that EEG could serve as a direct, implicit, and real-time measurement of cognitive workload during RSVP reading. This overcomes many of the limitations of questionnaires, which include the need for an explicit measurement phase as well as the mitigating influence of the user's subjective self-evaluation (*e.g.*, perceived workload) or their prior knowledge of the read content (*e.g.*, text comprehension). Here, we evaluate if EEG could feasibly be used to predict the individual current gain in reading speed, text comprehension, and subjectively perceived workload. We report on a user study that measures cognitive workload which is raised by the two RSVP parameters *Text Alignment* and *Presentation Speed* using EEG. Thereby, EEG as a direct physiological

measure to anticipate the current individual gain in reading speed, subjectively perceived workload, and text comprehension using predictive models from regression analyses. Finally, we discuss how our results contribute to the evaluation of future novel RSVP designs and how our findings can be used to select individual RSVP parameters based on cognitive workload:

$\rightarrow$ **RQ7**: How can RSVP reading parameters be selected based on cognitive workload?

This section is based on the following publications:

- Thomas Kosch, Albrecht Schmidt, Simon Thanheiser, and Lewis L. Chuang. One does not Simply RSVP: Mental Workload to Select Speed Reading Parameters using Electroencephalography. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, New York, NY, USA, 2020. ACM

- Thomas Kosch and Lewis L. Chuang. Investigating the Influence of RSVP Display Parameters on Working Memory Load using Electroencephalography. In *2nd International Conference on Neuroadaptive Technology*, 2019

10.1 Related Work

We surveyed related research that investigated the impact of RSVP on the reading experience, performance, working memory, and engagement.

10.1.1 Reading using RSVP

Past researchers investigated how RSVP can be used to increase, control, or manipulate the overall reading speed. While reading with RSVP, words are displayed word-by-word on a fixed space. This is achieved by updating the currently displayed word with a fixed frequency in a single-word-at-a-time notion [119, 393].

Using RSVP to read text is believed to increase the overall reading speed. Words are constantly presented in the same visual field of view and thus neglect eye movements, that require time, to jump from word to word [182, 315]. For example, Rubin and Turano [321] compared the reading speed and comprehension between the regular reading of a single page letter and previously calibrated RSVP speeds. Their findings show that RSVP speeds up the reading process while maintaining a text comprehension similar to regular reading. However, the presented approach relies on a calibration method that requires users to read the text

aloud. Based on the incorrect or correct reading, the RSVP speed was adjusted manually by the experimenter.

Since RSVP presents a word at a time, it reduces display-related space restrictions that enable the representation of longer texts on devices with small screen space. For example, RSVP has been successfully employed on mobile phones [289], smartwatches [96], and mobile augmented reality devices [324]. Since the requirements and context for RSVP designs differentiate between devices, the requirements for RSVP may be different depending on the used device as well as on the user requirements [188]. For example, users may prefer slower word representation speeds in a mobile scenario than in a static context. In the following, we present related research that investigated relevant RSVP designs and their parameters to provide suitable reading experiences.

10.1.2   Design of RSVP Parameters

RSVP reading operates by presenting single words or short phrases one after another. In contrast to regular text reading, it removes the need for eye movements and controls the rate at which text information is accessed by the reader. Past research was concerned with factors that optimize the overall reading performance.

The centering of text and the rate of its presentation henceforth referred to as *Text Alignment* and *Presentation Speed* respectively, can be argued to critical design parameters. A faster *Presentation Speed* reduces the time it takes to read a document. Meanwhile, effective *Text Alignment* ensures that the presented word or phrase is appropriately fixated by the reader upon presentation without the need for eye movements. Thus, a factor for effective *Text Alignment* is the fixation location of a word or sentence, where a large number of neighboring letters are visible to recognize the word in a short amount of time [49]. Some commercial RSVP implementations (*e.g.*, Spritz Inc.[1]) adopt an appropriate fixation location, also known as an Optimal Recognition Position (ORP), to eliminate unnecessary saccades and minimize eye movements. For instance, highlighting and centering words around their ORP can reduce the overall word processing time. However, reading with ORPs over longer periods suppresses parafoveal processing, increases subjectively perceived workload, and elicits a higher number of eye blinks which is an indicator for visual fatigue [27, 65]. Thus, Dingler et al. [98] explored how alternative ORP representation, such as underlining ORPs instead of coloring them, affect subjectively perceived workload and text comprehension. Although no significant results in subjective workload and text comprehension were found between underlined and colored ORPs, participants elaborated that they adjusted more quickly to colored ORPs instead to underlined ORPs. An alternative approach that utilizes colors in text representations is BeeLine Reader[2]. BeeLine Reader integrates color gradients within text lines to save time during return sweeps. Beyond that, Russel and Chaparro [322] investigated how different font sizes affected RSVP reading compared to regular reading. They

---

[1] `www.spritzinc.com` - last access 2020-01-28

[2] `www.beelinereader.com` - last access 2020-01-28

The
red
brown
fox
t

The
red
brown
fox
t

The
red
brown
fox
t

|·················| Eye movement |

(a) (b) (c)
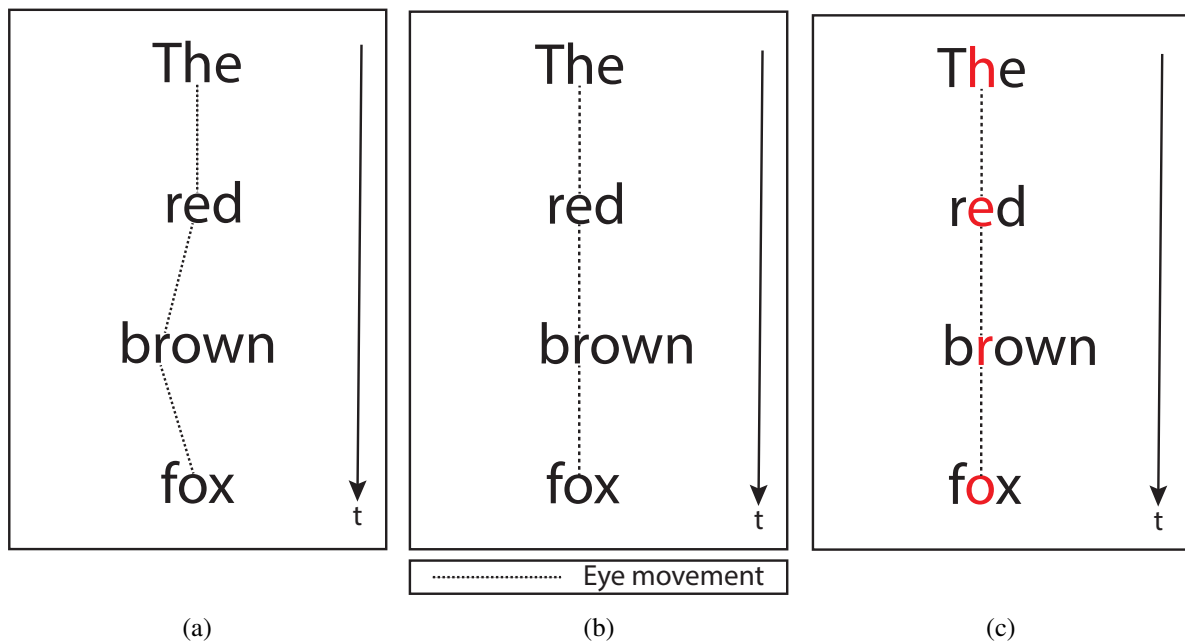
Figure 10.2: Word position alignment for each text representation. **(a):** Centered word positioning. **(b):** ORP-centered word without color. **(c):** ORP-centered word with color.

find that font sizes did not significantly influence text comprehension. However, participants stated a preference for font sizes that were readable and did not strain their eyes.

Presentation Speed yields another critical RSVP design parameter. An effective *Presentation Speed* has the potential to speed up reading times by accelerating the frequency of displayed words. However, as *Presentation Speed* increases, less time is available to cognitively process words. Previous studies found significant differences in text comprehension, whereas text comprehension and *Presentation Speed* are correlated negatively [322]. Previous studies have confirmed that the preferred reading speed is a factor that depends on individual preferences and text type [234].

10.1.3 Cognitive Processing during Reading

Reading elicits different strains of working memory regarding information retention. Daneman and Carpenter [88] revealed that reading performance and comprehension is individual among users. A reader with poor processing and storage functions may retain information for a shorter time compared to more experienced readers. However, at the cost of time, readers that take more time to process information can increase their overall text comprehension [242].

Various researchers presented approaches to quantify the exert of working memory for different RSVP design parameters that mainly include *Text Alignment* and *Presentation Speed*.

The awareness about the workload placed on readers enables to compare RSVP design parameters in terms of time-savings and information retention. For example, Dingler et al. [98] presented the use of the NASA-TLX questionnaire [160, 161] to measure several facets of experienced workload when reading with different RSVP design parameters. Further subjective measures, such as the Continuous Subjective Workload Assessment Technique (C-SWAT) [254], have been employed to measure the placed cognitive workload for various RSVP parameters [140]. However, the aforementioned measures are susceptible to subjective biases. When comparing RSVP design parameters in a study, participants might suggest a better reading performance due to previous knowledge about the presented text or lack of recall regarding their performance during the self-assessment itself.

To overcome the previous gaps from subjective assessments, physiological measurements have been proposed as an alternative to evaluating the suitability of RSVP design parameters regarding reading and information retention performance. Oliveira et al. present how EEG has been used to detect reading activity [285, 286, 287]. They incorporate several frequency bands for each electrode to train a k-nearest neighbor classifier. However, they apply their approach for regular text reading only and do not evaluate reading activity during different RSVP design parameters. A similar approach was taken by Yuan et al. [395], where reading comprehension was evaluated instead of reading activity during regular reading. Finally, Lees et al. [240] provide a literature review about RSVP and EEG. In their review, they present past work that utilized an event-related potential-based brain-computer interface paradigm during the RSVP of images and words for analysis or input. The aforementioned research investigates frequency bands as an indicator of reading activity and comprehension. Specific EEG bandwidths, such as the frontal theta bandwidth (4 Hertz to 8 Hertz) or alpha bandwidth (8 Hertz to 12 Hertz), are correlated with executive cognitive functions and brain resting states that whether synchronize (*i.e.*, theta power increases) or desynchronize (*i.e.*, alpha power decreases) when cognitive workload is raised [202].

Cognitive functions correlate with changes in the power of theta and alpha frequency bands of EEG measurements. In particular, high alpha power is associated with the default mode of a resting brain, which is exemplified by synchronized neural activity in the alpha bandwidth [259]. Hence, cortical processing of perceived stimuli disrupts this default mode, resulting in alpha-desynchronization that manifests as reduced alpha power. For instance, inter-individual differences in alpha power have shown to reflect memory performance at the occipital lobe [200]. In line with this, reading and text comprehension elicits lower alpha power as reading requires cognitive resources to memorize recently read propositions, text coherence, and information to put sentences into their context [80, 150]. The theta frequency bandwidth, on the other hand, is related to executive functions and correlates negatively with the default mode of a resting state [333]. An increase in theta power is indicative of focus or higher task engagement [30] across vigilance, learning, and memory tasks. In the context of the current work, high theta power is also prominent during the effective semantic processing of language [22]. Thus, theta power is expected to rise with increasing language processing activity.

Some researchers also adopt the theta-alpha ratio [202] as a hybrid metric for cortical activity that normalizes frontal theta power relative to alpha power. This entangles previous findings of frequency bands on brain resting states (*i.e.*, alpha) and semantic processing (*i.e.*, theta) into a metric that expresses cognitive processing or engagement during reading. Thus, a small theta-alpha quotient could correlate with low reading effort or no engagement in reading at all while a large theta-alpha quotient can indicate high reading engagement or increased cognitive demand.

10.1.4 Summary

Past research investigated how different settings of RSVP influence individual information processing, working memory, and engagement. However, previous work did not address how RSVP parameters can be implicitly evaluated without the need for individual user adjustments or reading interruptions, hence utilizing fixed RSVP reading parameters that do not take the mental demand into account. We close this gap by presenting a study that investigates the impact on EEG measures from three different *Text Alignments* and *Presentation Speeds*. We compare the alpha and theta bandwidth, subjective workload, and text comprehension of different RSVP parameter settings to regular reading and examine the placed cognitive workload. Finally, we evaluate alpha and theta bandwidths as a predictive metric for the current gain in speed relative to regular reading, subjectively experienced workload, and text comprehension.

10.2 Study

We designed and conducted a controlled user study to evaluate the influence of the two RSVP parameters *Text Alignment* and *Presentation Speed* on text comprehension, subjective workload, and measured cortical activity. The experiment consisted of a repeated-measures design comprising two factors with three levels each. They were *Text Alignment* (*centered*, *ORP*, *ORP* with colored letter) and *Presentation Speed* in words per minute (WPM) (200 WPM, 350 WPM, and 500 WPM). Text comprehension, subjective workload, and cortical activity (*i.e.*, power in the theta and alpha frequency band) served as the primary measures in our study. Comparisons are drawn to the baseline of normal non-RSVP reading from a PDF document and across the test conditions. We describe the independent variables in the following.

10.2.1 Text Alignment

RSVP can vary in terms of the alignment of presented words, relative to the display and each other. A *centered* condition presented words within a bounding box that is *centered* in the display (see Figure 10.2a). An *ORP* condition centered each word around the *ORP* (see
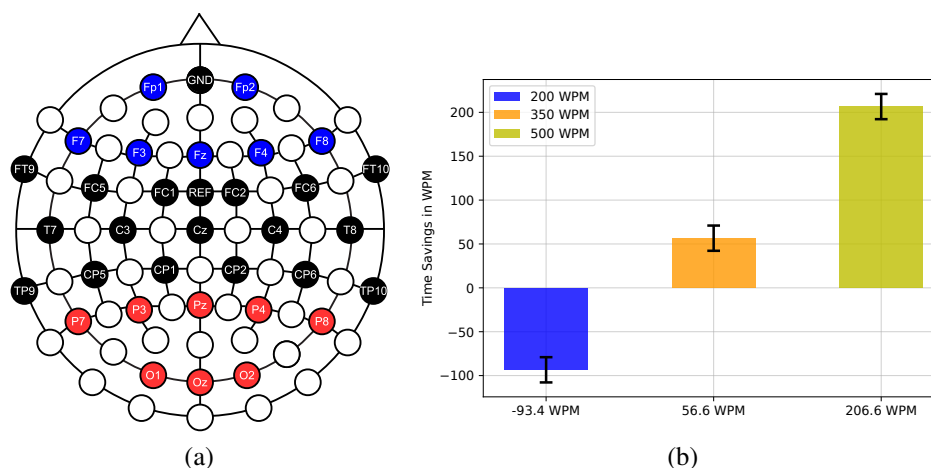
Figure 10.3: (a): 32 electrode layout used in the experiment. The ground electrode was placed on *Fpz* and the reference electrode on *FCz*. The red marked electrodes were used to analyze alpha frequencies while the blue electrodes were used to analyze theta frequencies. **(b):** Time difference of using RSVP compared to reading a PDF. Using 200 WPM slowed participants down, while 350 WPM and 500 WPM accelerated the reading times. Hence, 200 WPM did not provide effect time-savings compared to 350 WPM and 500 WPM. The error bars depict the standard error.

Figure 10.2b). An *ORP* with colored letter condition colored the *ORP* red in addition (see Figure 10.2c). This approach is similar to the algorithm employed by Spritz Inc. [27].

10.2.2 Presentation Speed

RSVP can vary in terms of how quickly they are presented measured in words per minute (WPM). Three *Presentation Speeds* were chosen with equal intervals, namely 200, 350, and 500 WPM. Previous work suggests that readers can understand RSVP content well at 200 WPM [68]. The conditions of 350 WPM and 500 WPM were chosen to investigate the influence of higher *Presentation Speeds* on cortical activity, subjective workload, and cortical activity [63, 352]. Hence, we hypothesized workload to increase with higher *Presentation Speeds*, albeit not necessarily at the cost of reading comprehension.

10.2.3 Participants

We recruited 18 participants (7 females, 11 males; age $M = 27$, $SD = 3.7$). All participants had a normal or corrected-to-normal vision and none of them reported a history of neurological disorders. Two participants have used RSVP once before. Participants were recruited through university mailing-lists. They received 15 Euro or lecture study points for their participation. We removed one participant from our analysis due to technical difficulties during the experiment.

10.2.4   Apparatus and Stimuli

The study was conducted in a self-contained room with constant luminance throughout experimentation. The room was divided into two areas, a test area for the participant and a control area for the experimenter. The experiment was controlled remotely with a laptop while participants were presented with stimuli via an LCD monitor (Dell U2715H; 27 inches diagonal; 2560 × 1440 pixel resolution; 60 Hertz refresh rate).

The reading distance approximated 50 cm from the participants' head to the display. As suggested by previous research, words were displayed with a font size that did not strain the participants' eyes to avoid confounding measures during the experiment [322]. These were consistently 20 points during the PDF trial and 55 points during the RSVP trials. The font size was adjusted to match a similar size relative to the screen resolution. EEG data was collected using a 32 active gel electrodes wireless system (BrainVision LiveAmp[3], Brain-Products GmbH; bandpass: 0.1-1000 Hz, no notch filter). Electrodes were placed in accordance to the International 10-20 layout (ground electrode: *Fpz*, reference electrode: *FCz*; see Figure 10.3a). Before testing, the impedance of all electrodes was kept to less than 10 $k\Omega$. EEG data was collected at the occipital and frontal lobe. Changes in alpha power at the occipital lobe are associated with respective changes in brain resting states, where high alpha power is associated with a distinct resting state. Respectively, low alpha power is associated with low brain resting states [143]. Theta power is an indicator of task engagement [201] when measured at the frontal lobe. Thereby, high theta power corresponds to high task engagement whereas low theta power suggests the opposite. EEG data were annotated with unique triggers for the start and end of the experiment as well as for each word presentation according to the relevant test condition. We adopted a set of ten short excerpts from the book *Speed Reading: A Course for Learners of English* [312, 324]. The presented excerpts had a mean average of $M = 547.9$ words ($SD = 3.7$) and were accompanied by a unique set of ten comprehension questions.

10.2.5   Procedure

Before testing, participants were required to perform some tasks to establish baseline measures. First, we derived the individual alpha frequency (IAF) by recording one minute of resting-state EEG activity for when the participants' eyes were opened and closed [201]. Next, we requested participants to read a one-page PDF text excerpt that was presented all at once via a PDF viewer. With this, we established the baseline for each participants' reading time and individual cortical activity. Ten comprehension questions and a raw NASA-TLX questionnaire were conducted after each reading trial. Testing consisted of nine RSVP presentation conditions which were counter-balanced according to the balanced Latin square [76]. After each condition, participants were presented with ten questions that tested for reading comprehension and a NASA-TLX questionnaire for evaluating subjective

---

[3] `www.brainproducts.com/productdetails.php?id=63` - last access 2020-01-28
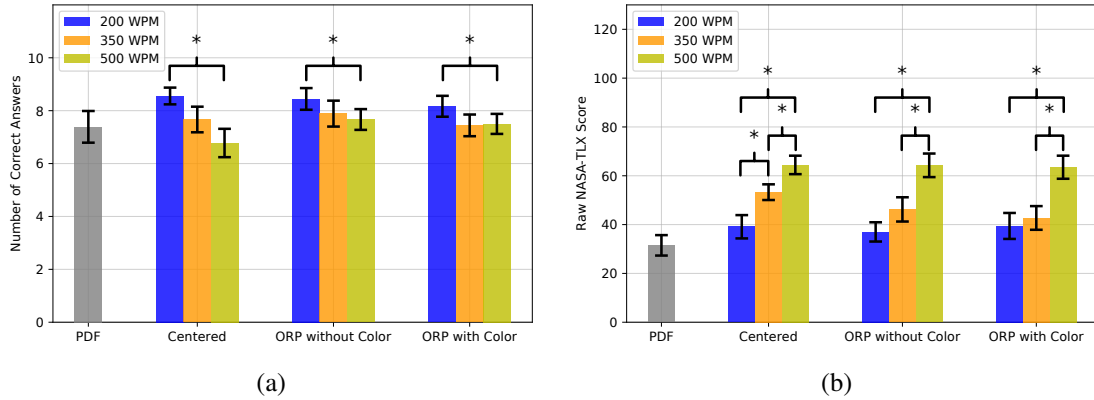
Figure 10.4: (a): Mean comprehension scores for the showed questionnaires. Increasing the *Presentation Speed* reduces the number of correct answers and may impact the overall comprehension. The error bars depict the standard error. Brackets indicate significant differences. **(b):** Mean raw NASA-TLX scores per condition. Less WPM induce less subjective compared to higher WPM. The error bars depict the standard error. Brackets indicate significant differences.

workload [161]. The full study took approximately 90 minutes to complete. Briefing and debriefing of the study purpose were performed before and after the study respectively. All participants provided signed informed consent.

10.2.6 Data Preprocessing

Four seconds of data were removed from the beginning and end of each recording to remove signals that are unrelated to cortical activity. The IAF bandwidth [201] was determined for each participant based on the separate spectral analyses of the EEG recording when the participants' opened their eyes and closed them; the mean power around the maximum difference peak (±2 Hz) was submitted to a spatio-spectral decomposition filter [279, 310]. The spatio-spectral decomposition derives the bandwidth power based on neighboring electrodes. A decomposition of recordings maximizes the signal power of the chosen frequency while simultaneously minimizing it at the neighboring electrodes, thus mitigating the appearance of noise for frequencies outside the alpha frequency bandwidth. We performed a short-time Fourier transform by dividing the signal into one-second slices with an overlap of half a second. After estimating the IAF from each of those electrodes, we averaged the individual frequency band per electrode. This procedure was applied to the electrodes *Pz*, *P3*, *P7*, *O1*, *Oz*, *O2*, *P4*, and *P8* located at the occipital lobe [143]. (see Figure 10.3a). A similar procedure was applied to extract theta power. The signal was filtered using a spatio-spectral decomposition (5 ± 2 Hz). Again, a short-time Fourier transform was carried out by dividing the signal into one-second slices with an overlap of half a second and were mean averaged to obtain the final theta power. This procedure was applied to the electrodes *Fp1*, *Fp2*, *F7*, *F3*, *Fz*, *F4*, and *F8* (see Figure 10.3a).

	Parameter	F	p	ω^2
Reading Comprehension	*Presentation Speed*	3.68	.04	0.08
	Text Alignment	0.85	.44	0.00
	Text Align. × Pr. Sp.	0.60	.67	0.00
Subjective Workload	*Presentation Speed*	18.5	<.001	0.27
	Text Alignment	1.39	.26	0.00
	Text Align. × Pr. Sp.	1.02	.40	0.00
IAF Power	*Presentation Speed*	491.19	<.001	0.61
	Text Alignment	2.37	.13	0.00
	Text Align. × Pr. Sp.	2.48	.11	0.00
Theta Power	*Presentation Speed*	338.09	<.001	0.58
	Text Alignment	1.00	.38	0.00
	Text Align. × Pr. Sp.	0.42	.79	0.00

Table 10.1: Summary of the confirmatory ANOVA results.

10.3 Results

The RSVP test conditions provided three measurements for each participant that we submitted for further analyses with a 3×3 repeated measures analysis of variance (ANOVA) for the factors of *Text Alignment* (*centered*; *ORP*; *ORP* with red color) and *Presentation Speed* (200, 350, 500). They were reading comprehension, raw NASA-TLX workload score, and IAF power. They respectively represented performance, subjective workload, and cortical activity. A Shapiro-Wilk test did not reveal a deviation of normality from the measures ($p \geq .05$).

We adopted an alpha-level of 0.05 for statistical significance testing and report the ω^2 for effect sizes of significant results. Greenhouse-Geisser corrections are reported when violations of sphericity are detected. Post hoc Bonferroni-corrected tests were performed between test conditions to investigate the significant main effect of *Text Alignment*. Planned linear contrasts were performed to understand significant main effects or interactions that included *Presentation Speed*. Post hoc pair-wise comparisons to baseline measures, derived from normal reading, was performed when applicable. A summary of the results can be found in Table 10.1.

10.3.1 Time-savings of RSVP

To evaluate the time-savings of RSVP, we subtracted the normal non-RSVP reading times of our participants from 200, 350, and 500 WPM. The mean reading time was 293.4 WPM ($SD = 59.3$). Hence, the mean time-savings were -93.4 WPM, 56.6 WPM, and 206.6 WPM,
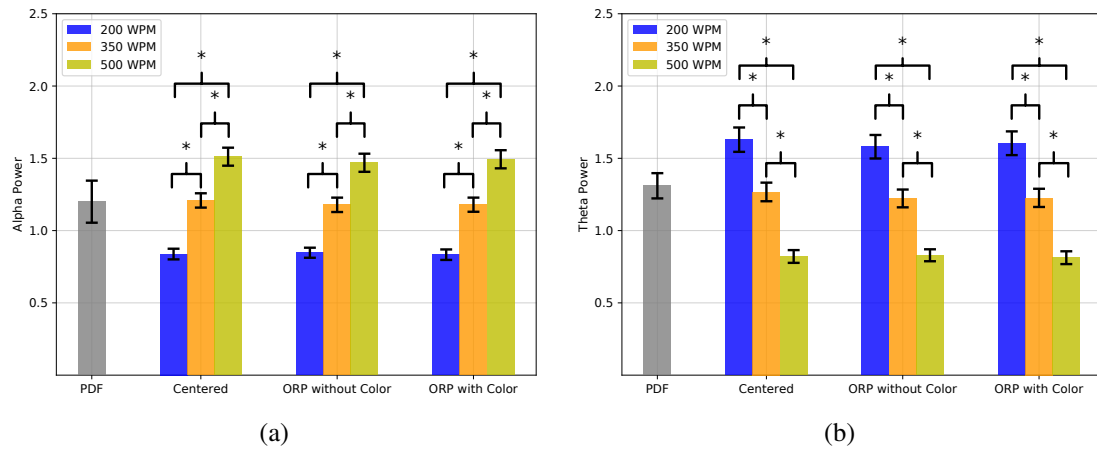
Figure 10.5: (a): Normalized mean IAF band power per condition. Increasing the WPM yields an increase in alpha power band while slower speeds elicits lower alpha powers. The error bars depict the standard error. Brackets indicate significant differences. **(b):** Normalized mean theta power per condition. Lower *Presentation Speeds* lead to higher theta power. The error bars depict the standard error. Brackets indicate significant differences.

for the *Presentation Speed* levels of 200 WPM, 350 WPM, and 500 WPM respectively. Figure 10.3b illustrates these time-savings. One-sample t-tests show that participants were significantly slower at the *Presentation speed* of 200 WPM ($t(16) = -6.69, p < .001, d = -1.58$) and significantly faster at 350 and 500 WPM ($t(16) = 4.05, p < .001, d = 0.95$; $t(16) = 14.8, p < .001, d = 3.49$), relative to their normal reading times.

10.3.2 Reading Comprehension

Reading comprehension was operationalized in terms of the number of correct answers out of the ten questions that were posed during each questionnaire. Figure 10.4a illustrates the mean number of correct answers per participant and condition. We found a significant main effect for *Presentation Speed* with a medium effect size, whereas no significant effect was found for *Presentation Speed* and *Text Alignment × Presentation Speed*.

10.3.3 Subjective Workload

The subjective workload was operationalized as raw NASA-TLX scores. Figure 10.4b shows the mean averaged raw NASA-TLX for each condition. There was a significant main effect of *Presentation Speed* with a large effect size. No significant effect was found for *Presentation Speed* and *Text Alignment × Presentation Speed*.

	IAF					Theta				
	F	p	R	R^2	RMSE	F	p	R	R^2	RMSE
Reading Speed Gain	3681.95	$< .001$	**.980**	**.961**	**0.11**	2266.56	$< .001$	**.984**	**.968**	**0.10**
Subjective Workload	25.24	$< .001$	.378	.143	20.88	28.95	$< .001$	.401	.161	20.66
Text Comprehension	7.47	.007	.217	.047	1.83	8.43	.004	.230	.053	1.82

Table 10.2: Results of the predictive models using a linear regression. All dependent variables derive a significant regression equation ($p < .05$). The normalized IAF and theta powers are reliable predictors for the current gain in reading speed according to the large R^2 value and low Root Mean Square Errors (RMSE). However, subjective workload and text comprehension provide less accurate predictive models. Bold values represent the most efficient results.

10.3.4 Cortical Activity for Cognitive Workload

Cortical activity was operationalized in terms of the mean power within the IAF bandwidth (ca. 10 ± 2 Hz) and theta bandwidth (ca. 5 ± 2 Hz). Deviations from resting-state EEG activity result in lower alpha power and higher theta power [202, 199]. This is referred to as alpha-desynchronization and theta-synchronization. This is a reliable classification feature for working memory [13], which describes the cognitive processing of information that is kept in short-term memory. The alpha bandwidth, characterized by a unimodal peak in spatio-spectral power, varies across individuals and age [201]. Thus, we determined the IAF bandwidth for each participant based on their peak frequency value across the eyes-opened and eyes-closed pre-test conditions. We defined ±2 Hz around the peak frequency as the individual alpha band. The mean alpha peak was $M = 9.5$ Hz ($SD = 1.95$), which provides a mean bandwidth of $7.5 - 11.5$ Hz. Since alpha and theta power are individual metrics which are different among participants, we normalized the bandwidths by their individual bandwidth power that is elicited during the PDF reading trial before aggregating the data.

IAF Power

There was a significant main effect for *Presentation Speed* with a large effect size. This was a significant linear trend whereby IAF power increased with increasing *Presentation Speed*. The main effect of *Text Alignment* and *Text Alignment × Presentation Speed* interaction were not statistically significant. Figure 10.5a shows the normalized mean IAF power for the employed *Text Alignments* and *Presentation Speeds*. Interestingly, the linear trend for cortical activity contradicts the percept of our participants for subjective workload. The current results suggest that there is less cognitive workload with increasing *Presentation Speed* based on cortical activity. This raises the possibility that high *Presentation Speed* might introduce subjective workload while restricting our participants' neuro-cognitive ability to process the presented words. To evaluate this possibility, we performed a post hoc tests to find a significant effect between 200 and 350 WPM ($t(16) = -21.77, p < .001, d = -5.28$), 200 and 500 WPM ($t(16) = -22.72, p < .001, d = -5.5$) as well as 350 and 500 WPM ($t(16) = -20.45, p < .001, d = -4.96$).

Theta Power

The analysis of theta power results in a significant main effect for *Presentation Speed* with a large effect size. No significant main effect was found for *Text Alignment* and *Text Alignment × Presentation Speed*. Similar to the analysis in the IAF bandwidth, a significant linear trend presented here indicates the possibility to restrict the participants' possibility to process words. Figure 10.5b depicts the normalized mean theta power per condition.

Number of Eye Blinks

We counted the number of eye blinks to investigate if the number of eye blinks significantly change with different *Presentation Speeds*. An increased number of eye blinks is an indicator of visual fatigue [65] that could arise from different presentation speeds. Since eye blinks introduce noise in EEG measures, we want to ensure that our EEG recordings result from cortical activity and not from noise. We use Python MNE to automatically count eye blinks using the electrodes *Fp1* and *Fp2* [361] to detect EOG artifacts using a threshold-based approach which was set to $200 \mu V$. We found that 200 WPM elicited the highest number of eye blinks ($M = 44.88$, $SD = 33.12$) followed by 350 WPM ($M = 23.55$, $SD = 8.99$) and 500 WPM ($M = 18.8$, $SD = 7.6$). Normalizing the results to blinks per minute shows 16.38 blinks per minute for 200 WPM, 15.04 blinks per minute for 350 WPM, and 17.16 for 500 WPM. Comparing each *Presentation Speed* using pairwise t-tests did not yield a significant effect. Due to minimal differences in blinks per minute and non-significant differences for each *Presentation Speed*, we do not assess distorted alpha or theta measures due to eye closure.

10.3.5   Presentation Speed Compared to Regular Reading

The main effect of *Presentation Speed* was consistently significant across all three measures. Therefore, we performed pair-wise comparisons between normal reading and the three levels of *Presentation Speed* on each measure. The median score of the three *Text Alignment* levels were treated as the representative score of the corresponding level. Subjective workload was the only measure with significant differences, whereby the *Presentation Speed* of 350 and 500 WPM resulted in significantly higher values of subjective workload compared to normal reading ($t(16) = 4.06, p < .001, d = 0.96; t(16) = 7.35, p < .001, d = 1.73$). There were no significant differences between the *Presentation Speed* levels and normal reading for reading comprehension and cortical activity. To summarize, the time-savings of RSVP at 350 and 500 WPM was associated with the cost of increasing subjective workload, without significant improvements in reading comprehension or changes in cortical activity.

10.4 Evaluating EEG as Predictive Metric

The results show significant differences in alpha and theta power for reading with different *Presentation Speeds*. We evaluate the efficiency of models that utilize EEG frequency bands to predict current gains in reading speed, subjective workload, and text comprehension using a linear regression analysis.

10.4.1 Independent and Dependent Variables

We use the IAF and theta power as independent variables for the regression analysis. Similar to the previous analysis and to mitigate person-dependent differences in the alpha and theta bandwidths, both bandwidths are normalized relative to the full-text baseline reading trial. For each RSVP condition concerning *Presentation Speeds* and *Text Alignments*, we calculated the mean bandwidth for the IAF and theta power for each RSVP reading trial. This resulted in nine data points for each bandwidth and each participant. Hence, a total of 153 data points were used to fit a function for the IAF and theta power. We use these data points to evaluate a predictive model using linear regression to evaluate the forecasting efficiency for the current gain in reading speed relative to the full-text reading trial, subjectively perceived workload, and text comprehension.

10.4.2 Predictive Performance

We describe the results of the regression analysis in the following. We report the significance of the regression equations and the accuracy of the fitted model. A summary of the results can be obtained from Table 10.2.

A significant regression equation was found for gains in reading speed ($F(1, 151) = 3681.95, p < .001, R^2 = .961$) with a linear trend for increasing *Presentation Speeds* and IAF power. In contrary, theta power resulted in a significant regression equation ($F(1, 151) = 2266.56, p < .001, R^2 = .968$), where theta power shows a decreasing linear trend with higher *Presentation Speeds*. This indicates a strong linear trend between the EEG frequency bands and gains in reading speed as suggested by the previous results. Similarly, a significant regression equation ($F(1, 151) = 25.24, p < .001, R^2 = .143$) between increasing raw NASA-TLX scores and IAF power was confirmed. Furthermore, we find a significant regression equation ($F(1, 151) = 28.95, p < .001, R^2 = .161$) between decreasing theta power and increasing raw NASA-TLX scores. However, the low R and R^2 scores show that the variance of the raw NASA-TLX scores is large, thus indicating a low predictive performance. Finally, we apply the same regression analysis for the text comprehension scores. Again, this results in a significant equation ($F(1, 151) = 7.47, p = .007, R^2 = .047$) with decreasing text comprehension for increasing *Presentation Speeds*. Theta power also shows a significant equation ($F(1, 151) = 8.43, p < .001, R^2 = .161$), where text comprehension decreases with decreasing
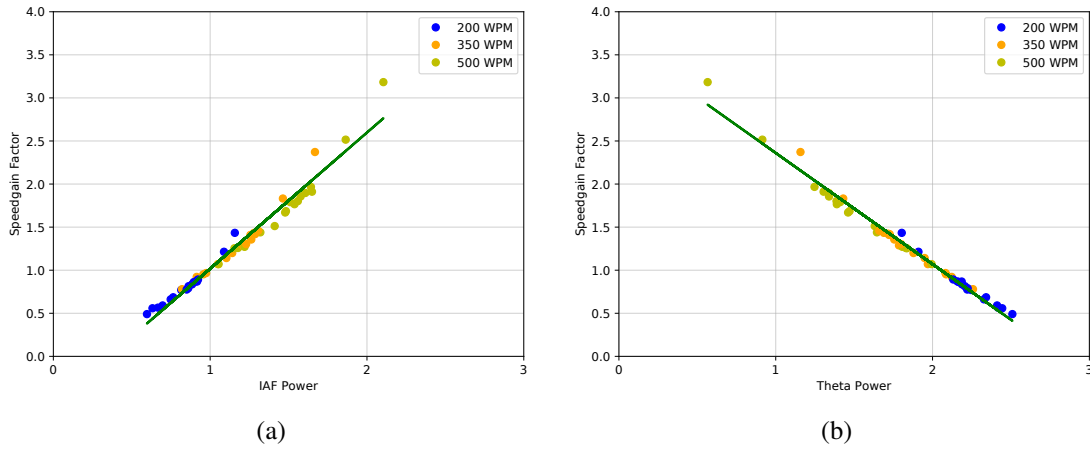
Figure 10.6: (a): A linear regression of the normalized IAF power results in an efficient prediction for gains in reading speed. The green line denotes the regression function. The colors resemble the *Presentation Speed*. **(b):** A linear regression of the normalized theta power results in an efficient prediction for gains in reading speed. The green line denotes the regression function. The colors resemble the *Presentation Speed*.

theta power. Similar to the raw NASA-TLX scores, the R and R^2 denote large variances among the data points and, therefore, a low predictive performance.

10.4.3 Towards a General Model for Predicting Reading Speed

The IAF and theta power show a strong linear trend regarding the current gain in reading speed. These results suggest a functional relationship between gains in reading speed and the employed EEG frequency bands that generalize to the individual user, hence obviating a dedicated calibration phase. First, we averaged the *Presentation Speeds* for each *Text Alignment*, resulting in 51 data points for all participants. We then performed a leave-one-participant-out validation between gains in reading speed and the measured frequency bandwidths. More specifically, we iteratively derived a linear regression model using all participants except one and used the remaining one for the validation. This resulted in a low averaged error rate for the IAF (RMSE = 0.12, R^2 = .94) and theta (RMSE = 0.11, R^2 = .97) power, indicating that the model fits interpersonal differences that may not require individual user calibration. Figure 10.6 shows the regression lines for the IAF and theta power relative to the gains in reading speed for each participant and each *Presentation Speed*.

10.5 Discussion

We conducted a study to investigate the influence of *Text Alignments* and *Presentation Speeds* on text comprehension, subjective workload, and cortical activity. Our results show that time-savings are achieved from the *Presentation Speeds* of 350 WPM onwards which corresponds with increased alpha power and decreased theta power in EEG. Faster *Presentation Speeds* significantly increases subjective workload and impairs text comprehension. Interestingly, *Text Alignment* did not influence any of our measurements. Thus, it is clear that *Presentation Speed* will continue to be a limiting factor. Until this is resolved, spatial factors such as the ideal presentation of text, which includes font readability could be relatively less important and should only be considered after optimizing *Presentation Speeds*. It is worth noting that *Text Alignment* could have minimal impact on wearable displays, which are small and unlikely to subtend to a larger visual angle than the fovea (*i.e.*, 2°). Thus, we focus on the implications for *Presentation Speeds* on future developments of RSVP readers.

10.5.1 Cognitive Workload of RSVP

Our results show that the physical manipulation of *Presentation Speeds* does not simply bring gains in reading speed. It has a notable impact on our limited capacity for information processing at the cortical level. While 350 and 500 WPM induced significant gains in reading speeds, relative to regular reading, this was accompanied by a significant cost in subjective workload as well as comprehension scores. These performance findings agree with previous findings [322]. We show that this is measurable in terms of EEG correlates with established cognitive processes, whereby higher *Presentation Speeds* commensurate with reduced working memory and engagement as respectively measured by IAF and theta power [317]. At 200 WPM, our participants were forced to read slower than their regular speed, which increased text comprehension scores that were commensurate with the EEG correlates of higher working memory load and engagement [317]. Thus, we show that physical manipulation of the speed of RSVP readers exerts an influence on how we process information for comprehension that can be measured with EEG. Thus, we show that ongoing measurements of cortical activity can reliably indicate the rate at which we can process and comprehend presented stimuli. In many cases, it could be desirable to manipulate RSVP speeds to facilitate information processing, in a way that could be verified by EEG measurements, as long as fast reading is not a pressing requirement.

10.5.2 Evaluating EEG as Measure for Predictive Models

We use the mean power in the IAF and theta bands to predict gains in reading speed, subjective workload, and text comprehension scores. We found that only gains in reading speed could be reliably predicted with both IAF and theta power. This implies a linear relationship between *Presentation Speeds* and the measured IAF and theta bandwidths. This forecast

can support the dynamic selection of RSVP parameters according to the current context (*i.e.*, reading in private or mobile spaces, reading a novel or a scientific article) to maximize time-savings, supporting the development of adaptive brain-computer interfaces. In contrast, subjective workload and text comprehension were not reliably associated with our EEG measurements. Non-linear models might be necessary to estimate subjective workload and text comprehension, given that NASA-TLX questionnaires are susceptible to differences in subjective perception.

10.5.3   Real-Time Assessment of RSVP Parameter Selections

EEG affords high sampling rates and can be employed as a real-time indicator for the evaluation of RSVP reader designs [116]. Since RSVP can be ubiquitously employed on devices with limited screen space, the reading performance can be evaluated in short time frames using EEG. Our predictive model shows that workload-aware computer interfaces can implicitly sense the current gain in reading speed depending on the bandwidth powers relative to a reading baseline. The *Presentation Speed* can be adjusted to a suitable reading level depending on the current level of cortical activity. Adaptive RSVP *Presentation Speeds* can then be deployed on-the-go, where textual information is available at refresh rates suitable for mental processing. With the ubiquitous availability of RSVP-enabled devices, baselines of regular reading speeds can be collected in any context to evaluate the current RSVP reading speed using EEG. RSVP systems can then dynamically select *Presentation Speeds* which implicitly suits the users' reading ability.

10.5.4   Limitations

The current study has several design limitations. We limited our manipulations to three *Presentation Speeds*. Therefore, it is not clear how cortical activity is affected for slower *Presentation Speeds* than outlined in the study. However, due to our findings, we expect a similar engagement of mental resources for slower *Presentation Speeds*. Furthermore, the comprehension tests may be confounded by knowledgeable participants who were aware of the answers without reading the text. Although we made sure that the participants were not knowledgeable about the used text excerpts, we can not exclude the use of common knowledge during the tests. Finally, our sample consisted of people who exclusively did not use RSVP regularly. Therefore, our results only confirm the observed effects of non-trained RSVP users.

10.5.5   Outlook

Although our findings show the feasibility of relying on EEG measurements to predict gains in reading speed. Any gains due to *Presentation Speed* depends on the baseline reading

speed of the individual user. Being able to predict gains means that RSVP readers can adaptively determine the tradeoff of reading speed gains to ensure that *Presentation Speed* is maintained at a level that allows for meaningful text comprehension. Aspects, such as mobile use, reading interruptions, outdoor scenarios, or different viewing postures have to be determined before RSVP-based brain-computer interfaces can be implemented in a way that allows for good signal acquisition. These are open research questions that were not addressed in our experimental design. However, our findings provide a first step towards the adaptive selection of *Presentation Speed* for RSVP readers, which would need to be manually determined otherwise.

10.6   Study Conclusion

We present a user study that investigates the cognitive workload raised by the RSVP design parameters *Text Alignment* and *Presentation Speeds* using Electroencephalography (EEG). We find that a *Presentation Speed* of 350 WPM increases the reading speed compared to regular reading while preserving a similar level of cortical activity and text comprehension. However, faster *Presentation Speeds* increase the subjectively perceived workload with a decrease in the overall text comprehension and cortical activity. No effect was observed for different *Text Alignments*, making *Presentation Speed* the critical design parameter that needs to be optimized first before modifying *Text Alignments*. Due to linear trends in the EEG measures and *Presentation Speeds*, we perform a linear regression analysis to evaluate the robustness of predictive models for gains in reading speed, subjective workload, and text comprehension. While we find that the current individual gains in reading speed can be reliably forecasted using EEG, subjective workload and text comprehension are less suitable variables for reliable predictions. Our results show that future RSVP interface designer benefit from the presented approach to design workload-aware user interfaces that dynamically select RSVP parameters to suit the individuals' cognitive workload using EEG measures. We publish the data set to foster and encourage research in this area[4].

10.7   Chapter Conclusion

In this chapter, we investigated if EEG provides suitable metrics to assess cognitive workload during information consumption. Thereby, we were able to derive a predictive model that can predict the current gain in reading speed using EEG metrics. This enables an approximation of the reading speed towards the workload that is elicited during regular reading.

The human mind possesses a limited capacity for attention and working memory. These two factors can change during a reading session. During tedious passages of a manuscript, the attention level can go down, hence being a marker for annotating the displayed text as *unread*

---

[4] `www.github.com/hcum/one-does-not-simply-rsvp` - last access 2020-01-28

and slowing down the reading speed. In contrast, exciting or easy parts of a manuscript can be sped up to increase the information processing capacities. The individual circadian cycle is a variable that contributes as well to this circumstance. Different reading performances can be expected when reading at different times. Thus, a method for setting adaptive reading speeds benefits the user. To answer **RQ7**, we conducted a study that utilized EEG as a real-time measure to evaluate different RSVP parameters. Going a step further towards workload-aware interfaces, we evaluate a predictive model that uses EEG bandwidth powers to efficiently forecast the current gain in reading speed.

In conclusion, there is a tradeoff between reading speed and comprehension. In this chapter, we found that faster reading speeds provide time-savings although text comprehension is negatively affected. This is not detrimental per se since the human mind is efficient in filtering and discarding irrelevant information. Workload-awareness can be used to proactively regulate the users' current cognitive state and match the optimal RSVP reading mode.

V

CONCLUSION

Chapter 11

Conclusion and Future Work

Here on this road of no release
Pale reign of misery
That even time cannot erase
On through my sorrows
Life will go on

André Olbrich, Hansi Kürsch

In this thesis, we investigated how workload-aware systems can sense and utilize mental demand to provide cognitive assistance. Cognitive workload serves as an extrinsic modality which represents an additional dimension in the paradigm of context-aware computing. Such workload-aware interfaces are envisioned in the domain of assistive computing where cognitive workload is sensed and "offloaded" to a system. Through a series of lab studies and field experiments, we explored opportune moments for diurnal cognitive assistance, evaluated EEG and eye tracking metrics as a measure for cognitive workload, and showcased several applications that benefit from real-time workload measurements. We provide a summary of the presented contributions and implications of this thesis as well as the limitations and future work in the following.

11.1 Summary

With the recent information growth and rising availability of ubiquitous computing systems, devices are competing for our attention and cognitive capabilities with regards to indirect as well as direct interaction. However, cognitive resources that map human cognition, such as working memory, are generally limited by the speed at which we can memorize information in the short-term memory, process or modify it, and provide the respective output (*e.g.*, phonological loop) [13]. If too many information chunks have to be remembered, the

cognitive system cannot keep up with additional incoming stimuli and information, and the likeliness of human errors and cognitive exhaustion rises. Hence, the treatment of the available cognitive resources is a crucial component to consider when designing workload-aware interfaces.

In Chapter 3 we started to investigate *how* cognitive assistance can be delivered regarding opportune moments that require cognitive assistance and pointed out potential strategies for cognitive assistance during a cooking session. Posing the question "*What are the design requirements for systems that provide opportune cognitive augmentation?*" (**RQ1**) we conducted a qualitative inquiry in a living facility sheltering persons with cognitive impairments. This user group is used to report the need for cognitive assistance at every opportunity to avoid dangerous situations and injuries. These explicit statements allowed us to analyze a series of observations of a cooking process and conduct interviews with caretakers. This resulted in five design implications for future assistive computing systems that provide cognitive support. Thereby, persons with cognitive impairments revealed their need for cognitive assistance explicitly. After becoming aware of how users encounter situations that require cognitive assistance, we investigated the research question "*What are suitable in-situ feedback modalities for cognitive support?*" (**RQ2**). We evaluated three in-situ feedback modalities regarding their acceptance as well as perception assuming that visual, auditory, and tactile sensation represent primary sources of information consumption. Using an assembly task designed to provoke errors, we studied the acceptance of visual, auditory, and tactile in-situ feedback. Visual in-situ feedback provided the highest user preference and performance in terms of task completion time and the overall number of errors. This was followed by auditory and tactile feedback, indicating a clear preference for visual feedback. This is potentially linked to the visual attention capabilities of humans which are responsible for recognizing up to 80 % of the presented information. Hence, notifying users and augmenting interfaces visually anticipates the highest benefit for cognitive support.

Chapter 3 and Chapter 4 provide us with situations that elicit the need for cognitive augmentation and their potential feedback modalities. These can be used to augment or "disaugment" the user depending on the current level of perceived cognitive workload level. Our prior approach required users to actively state their need for cognitive assistance. However, it is still unclear *when* cognitive assistance should be provided without stating an explicit need for it. Post hoc measures, such as questionnaire-based methods, provide measures for cognitive demand but lack real-time insights and are susceptible to subjective biases. To address this issue, we investigated objective EEG and eye tracking metrics as a real-time indicator for cognitive demand. With the research question, "*Does electroencephalography provide measures of cognitive workload for user interface evaluation?*" (**RQ3**) we investigated EEG signals emitted by the human brain as a direct physiological measure for cognitive workload. During an assembly task with two different instruction modalities, where both varied in their difficulty, we found that EEG is a suitable real-time measurement modality for working memory assessments. Hence, the cognitive workload generated by an interface can be synchronized with the user interaction to reveal user interface designers which aspects of their interface design induces workload. Finally, a system can use this measure to provide real-

time interface adjustments. However, EEG headsets are required to be body-worn. Thus, we further investigated eye gaze properties for cognitive workload assessment. Contrary to EEG, eye gaze can be recorded contactless without the need for body-worn sensors. In this context, we leverage eye gaze to investigate the research question *"Do eye gaze metrics enable the classification of cognitive workload states?"* (**RQ4**). First, we exploited changes in smooth pursuit eye movements during different levels of cognitive workload. We found that changes in eye movements can be efficiently measured between different task difficulties. Hence, a SVM-based classifier revealed a classification accuracy of 99.5% for person-independent classification between two workload states and a 84.5% person-dependent classification accuracy between four workload states using smooth pursuit eye movements. This shows that eye movements are a suitable indicator for different workload levels. We showcase use cases that may exploit natural eye movements to estimate workload states. Since situations exist that do not allow the integration of moving objects to elicit smooth pursuit eye movements, we extended our research to pupil diameter measures. Since changes in pupil dilation inform about cognitive workload states, we leveraged a math multiplication task to induce workload. Using a SVM-based classifier, we presented an adaptive system that alternates between simple and difficult task complexities based on the pupil diameter alone. We assess that an adaptive system enhances the overall task engagement while reducing the experienced frustration. The presented approach indicates the viability of using physiological sensing for implicit workload-aware measures.

The last part of the thesis looks at potential applications *which* make use of the prior presented research. We learned that visual in-situ feedback during a cooking session may help persons with cognitive impairments for autonomous meal preparation. We designed and prepared an assistive system that utilizes visual and auditory feedback to support people with cognitive impairments through a cooking process, thus investigating the research question *"Does in-situ feedback provide cognitive support during a cooking task?"* (**RQ5**). A two-week case study in a sheltered living facility was carried out where tenants with cognitive impairments cooked with caretakers and the assistive system. Although being slower using visual in-situ feedback, the tenants appreciated being able to independently prepare their meal through cognitive assistance. This underpins the physiological user interface evaluation pipeline presented in Chapter 5. However, the interpretation of complex physiological data can become overwhelming for inexperienced users, posing the research question *"How can complex physiological signals be visualized to be utilized by non-expert users?"* (**RQ6**). We developed and evaluated a visualization that displays the generation of brain activity inside the brain. Such visualizations are suitable to understand how stimuli affect the generation of cortical activity and, to this end, the impact of neurofeedback loops on neuronal processes. Finally, we investigated how cognitive workload is affected by different RSVP reading parameters. Since the RSVP parameter space is large in terms of *Presentation Speed* and *Text Alignments*, we pose the question *"How can RSVP reading parameters be selected based on cognitive workload?"* (**RQ7**) using EEG. The results of the study show that larger *Presentation Speed* increases the subjectively perceived workload while decreasing cortical activity. However, different *Text Alignments* did not show significant changes in cognitive workload, making *Presentation Speed* the sole critical factor. A predictive model using a

regression analysis shows that the results are applicable when predicting the current reading speed.

11.2 Contribution and Results

During this thesis, we followed an experimental approach to answer seven research questions that investigated how ubiquitous computing technologies can be used to sense cognitive workload to tailor task engagement while avoiding frustration and boredom. In the following we will briefly summarize the three key contributions:

11.2.1 Identifying Opportune Moments for Cognitive Assistance

Our studies in Chapter 3 and 4 showed *how* and *when* cognitive assistance is required during the accomplishment of daily routine tasks. Phases that require constant attention might be prone to user-related errors, increase the likeliness for injuries, and deteriorate the concentration level. Such situations render opportunities for visual, auditory, or tactile feedback to alert and notify users. After encountering a favorable moment for cognitive assistance, our results revealed that visual feedback was most preferred followed by auditory and tactile feedback (see Chapter 4). We evaluated visual in-situ feedback during a cooking scenario and found that mental workload is decreased on a cognitive and behavioral level (see Chapter 8).

11.2.2 Methods for Quantifying Cognitive Workload

We investigated eye tracking and EEG as sensing modalities for cognitive workload detection. We investigated EEG signals as an implicit marker for cognitive workload (see Chapter 5). Thereby, EEG showed a clear distinction of workload states between two assembly instruction designs. We describe a method that analyzes deviations in smooth pursuit eye movements to reliably classify different levels of cognitive workload on a person-dependent and person-independent basis between 88.1% and 99.5% (see Chapter 6). Further investigations showed that the pupil diameter is a pivotal indicator for cognitive workload. Finally, we employed a math task with two task difficulties to show that classifications between 64% and 99% can be achieved (see Chapter 7). However, person-dependent differences and changing lighting conditions are biasing factors that are subject to future research when classifying based on pupil dilation. This resulted in physiological design pipelines to implicitly assess cognitive workload in real-time and rich data sets that pave the way for future research in this area.

11.2.3   Tools for Workload-Aware User Interfaces

After showcasing workload sensing methods, we presented use cases and their reference implementations for workload-aware systems. Specifically, we evaluated the validity of a physiological evaluation pipeline in Chapter 5, self-awareness through visualizations in Chapter 9, and potential interfaces that adapt to cognitive workload levels in Chapter 10. Hence, ubiquitous systems are evaluated by physiological measures with complementary physiological feedback. This tailors cognitive engagement to the individual user. With these results in mind, we presented a visualization to foster self-awareness regarding the perceived cognitive workload and present predictive models that can be used by future RSVP reading interfaces for automatic parameter selection. Complemented by the data sets presented in the Chapters 5, 6, 9, and 10 as well as a reference implementation of an assistive system in Chapter 8 we share tools for the development of further workload-aware systems.

11.3   Limitations

Throughout our studies, we have employed empirical approaches to answer the stated research questions. We initially conducted lab studies (*e.g.*, assessing mental load through physiological sensing) before applying our methods within in-the-wild deployments. This was necessary to isolate confounding variables that might have influenced the validity of our results. After confirming the results, we transitioned our prototypes and methods into in-the-wild deployments. However, this transition resulted in the loss of control for variables that might lead to different results depending on the environment and individual user.

During our studies, we involved persons with cognitive impairments. These were included in the design process of workload-aware user interfaces since they explicitly require the need for cognitive assistance and supervision. The perception of task difficulty and support modality depends on the level of cognitive impairment. To investigate this, we included a large variety of cognitive impairments (*i.e.*, ranging from mild cognitive impairments to autism and Trisomy 21). Here, we did not investigate how individual cognitive impairments perceive the design of the presented interfaces.

We employed physiological sensing to sense cognitive workload states. Such measures enable automatic activation of assistive functions. While these can provide cognitive augmentation to support efficient task accomplishment, unsuitable assistance can be detrimental for the user experience. We assess EEG and eye tracking as an efficient marker for cognitive workload estimation in real-time. However, the analyzed features and space of physiological modalities presented in this thesis are far from exhaustive. The combination of multiple sensors for additional physiological responses may provide different results based on the used modalities. Another limitation is the susceptibility of noise during physiological sensing. Our studies were conducted in controlled lab environments to minimize noise. While this supports the validity of our results, we are not yet able to confirm similar results within in-the-wild settings. Thus, this is considered as part of future work.

Finally, individual differences in *cognitive constitution* and *physiological responses* vary per individual. While some approaches showed robust classification for cognitive workload (see Chapter 6), they are limited to the classification of two workload states (*i.e.*, binary classification). The fine-grained distinction of workload states often requires calibration to the individual user. While this might be acceptable in private spaces, interaction in public spaces (*e.g.*, interaction in the context of smart cities) benefits from immediate and calibration-free interaction. However, we believe that with the further integration of sensing devices (*e.g.*, smartphones) and ubiquitous activity tracking (*e.g.*, fitness trackers) implicit calibration without the users' awareness will be possible.

11.4 Future Work

This thesis presented design implications, sensors, and applications for systems that utilize workload as input for user interaction. While we focus on the design and the measurement modalities to support workload-related tasks, our methods can be transferred to numerous applications. In the following, we present future research that can be continued by picking up the topics presented in this thesis.

The use of machine learning provides a huge potential for classifying cognitive states in real-time. The presented methods are far from exhaustive and advanced methods, such as deep learning, become common to automate the extraction of features from physiological data. Functional Brain Networks (FBNs) [293] and Convolutional Neural Networks (CNNs) [348] are becoming popular methods for robust human state classification regardless of individual differences in cognition. However, the scarce availability of efficient methods for classification through deep learning and the need for huge amounts of data are a problem and provide possibilities for future research which includes the combination of sensors to enhance the overall classification performance. With the computing power available nowadays, these research topics can be picked up **immediately**.

The integration of physiological sensing into devices such as smartphones or fitness trackers has become a trend in the last few years since physiological sensors became smaller in size and are seamlessly integrated into wearables. This rise in devices that collect physiological data is credited to the increased acceptance of users to *share* their physiological data for certain *benefits*. For example, in the context of fitness trackers, activity data is shared for insights into health and activity states. After researching the area of *learning* the physiological behavior of humans, wearable devices require a seamless integration into the user's internal or external environment. Research on the unobtrusive integration of physiological sensing into the users' environments represents a major research question that can be picked up in a **mid-term** time horizon, *i.e.*, between three to five years.

Finally, ethics and user acceptance pose another important aspect of future workload-aware interfaces. With systems that can "gaze" into our cognitive states and mind, responsible handling of the underlying data and their interpretation must be established. Context-aware

devices communicate sensitive data with vendors to enhance the user experience. Cognition represents such an additional sensitive variable where its availability to third parties should be carefully considered. This research area requires a reliable physiological workload model of the user with hardware that is already able to reliably sense physiological signals. Therefore, we expect this research to be picked up in the **long-term** horizon, *i.e.*, ten years from now.

11.5 Final Remarks

With more and more tools becoming available to support our cognition, the question of what happens when removing the tool remains: Do humans depend on their tools for cognitive support? Can they still cope with daily chores when these tools are removed?

Technologies strive to make difficult tasks easier for humans, whether it be on a physical or mental level. Douglas Engelbart [111, 112] elaborated in the *60s* that technologies will play a pivotal role in augmenting the human intellect. Removing these tools will likely result in the use of support mechanisms that have been used before. This includes the integration of objects in the environment which is known as *active externalism* [75], a concept that has accompanied humankind throughout history. Support for working memory was once achieved by, for example, using fingers to count or memorization of road forks for navigation. While most of these tasks were substituted by technologies, a symbiosis between the environment and technology still exists. We believe that technologies make humans more efficient by combining computing efficiency with human flexibility. While these advantages are likely to fade away when removing such tools, humans will compensate for this lack of skill by other means.

A final remark remains on the ethical aspects of workload-aware systems that rely on physiological data collection. What are the caveats to consider assistive technologies that *gaze into the mind*?. Ubiquitous technologies, in general, collect huge amounts of data to provide benefits to the user. Including workload-awareness into a system extends the already large data collection corpus. This data analysis intends to provide a *benefit* at a certain *cost*. For example, the benefit might be an enhanced user experience at the cost of data sharing. In particular, physiological data that is transferred to third parties enables the derivation of patterns regarding, health, intellect, cognition, or stress levels. Transparent data collection and processing pipelines need to be disclosed by the respective stakeholders. The responsibility lies in the collecting institution to follow the guidelines of modern ethics [403] that inform and protect the user.

VI

LISTINGS AND BIBLIOGRAPHY

LIST OF FIGURES

LIST OF TABLES

Bibliography

[1] United nations: United nations convention on the rights of persons with disabilities. `http://www.un.org/disabilities/convention/conventionfull.shtml`, 2008. [last accessed 02-May-2016].

[2] Sylvia Ahern and Jackson Beatty. Pupillary responses during information processing vary with scholastic aptitude test scores. *Science*, 205(4412):1289–1292, 1979.

[3] Ulf Ahlstrom and Ferne J. Friedman-Berg. Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, 36(7):623–636, 2006.

[4] Motoyuki Akamatsu, I. Scott MacKenzie, and Thierry Hasbroucq. A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics*, 38(4):816–827, 1995.

[5] L. Anatole, Fabien Lotte, Richard B. Reilly, Robert Leeb, Michitaka Hirose, and Mel Slater. Brain-computer interfaces, virtual reality, and videogames. *Computer*, (10):66–72, 2008.

[6] Charles W. Anderson and Zlatko Sijercic. Classification of eeg signals from four subjects during five mental tasks. In *Solving engineering problems with neural networks: proceedings of the conference on engineering applications in neural networks (EANN'96)*, pages 407–414. Turkey, 1996.

[7] E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva. A user study of visualization effectiveness using eeg and cognitive load. *Computer Graphics Forum*, 30(3):791–800, 2011.

[8] Amaya Arcelus, Megan Howell Jones, Rafik Goubran, and Frank Knoefel. Integration of smart home technologies in a health monitoring system for the elderly. In *Advanced Information Networking and Applications Workshops, 2007, AINAW'07. 21st International Conference on*, volume 2, pages 820–825. IEEE, 2007.

[9] Nomy Arpaly. *Unprincipled virtue: An inquiry into moral agency*. Oxford University Press, 2002.

[10] Mark H. Ashcraft and Elizabeth P. Kirk. The relationships among working memory, math anxiety, and performance. *Journal of experimental psychology: General*, 130(2):224, 2001.

[11] Mark H. Ashcraft and Jeremy A. Krause. Working memory, math performance, and math anxiety. *Psychonomic bulletin & review*, 14(2):243–248, 2007.

[12] World Medical Association. World medical association declaration of helsinki. ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization*, 79(4):373, 2001.

[13] Alan Baddeley. Working memory. *Science*, 255(5044):556–559, 1992.

[14] Alan D. Baddeley and Graham Hitch. Working memory. volume 8 of *Psychology of Learning and Motivation*, pages 47 – 89. Academic Press, 1974.

[15] Alan D. Baddeley and Graham Hitch. Working memory. *Psychology of learning and motivation*, 8:47–89, 1974.

[16] A. Bannat, F. Wallhoff, G. Rigoll, F. Friesdorf, H. Bubb, S. Stork, H. J. Müller, A. Schubö, M. Wiesbeck, and Michael F. Zäh. Towards optimal worker assistance: a framework for adaptive selection and presentation of assembly instructions. In *Proceedings of the 1st international workshop on cognition for technical systems, Cotesys*, 2008.

[17] Fred Barbash. Hawaii missile mess: that was no 'wrong button.'take a look. *Washington Post January*, 16, 2018.

[18] Graham R. Barnes. Cognitive processes involved in smooth pursuit eye movements. *Brain and cognition*, 68(3):309–326, 2008.

[19] Solon Barocas and Helen Nissenbaum. Big data's end run around procedural privacy protections. *Commun. ACM*, 57(11):31–33, October 2014.

[20] Robert J. Barry, Adam R. Clarke, Stuart J. Johnstone, Christopher A. Magee, and Jacqueline A. Rushby. Eeg differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology*, 118(12):2765 – 2773, 2007.

[21] Robert J. Barry, Adam R. Clarke, Stuart J. Johnstone, Christopher A. Magee, and Jacqueline A. Rushby. Eeg differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology*, 118(12):2765–2773, 2007.

[22] Marcel C. M. Bastiaansen, Marieke van der Linden, Mariken ter Keurs, Ton Dijkstra, and Peter Hagoort. Theta responses are involved in lexical—semantic retrieval during language processing. *Journal of Cognitive Neuroscience*, 17(3):530–541, 2005.

[23] Thomas Baumgartner, Lilian Valko, Michaela Esslen, and Lutz Jäncke. Neural correlate of spatial presence in an arousing and noninteractive virtual reality: an eeg and psychophysiology study. *CyberPsychology & Behavior*, 9(1):30–45, 2006.

[24] Tom L. Beauchamp and James F. Childress. *Principles of biomedical ethics*. Oxford University Press, USA, 2001.

[25] Nikolaus Bee and Elisabeth André. Writing with your eye: A dwell time free writing system adapted to the nature of human eye gaze. In *International Tutorial and Research Workshop on Perception and Interactive Technologies for Speech-Based Systems*, pages 111–122. Springer, 2008.

[26] Hauke Behrendt, Markus Funk, and Oliver Korn. Ethical implications regarding assistive technology at workplaces. In *Collective Agency and Cooperation in Natural and Artificial Systems*, pages 109–130. Springer, 2015.

[27] Simone Benedetto, Andrea Carbone, Marco Pedrotti, Kevin Le Fevre, Linda Amel Yahia Bey, and Thierry Baccino. Rapid serial visual presentation in reading: The case of spritz. *Computers in Human Behavior*, 45:352 – 358, 2015.

[28] Simone Benedetto, Marco Pedrotti, Luca Minin, Thierry Baccino, Alessandra Re, and Roberto Montanari. Driver workload and eye blink duration. *Transportation research part F: traffic psychology and behaviour*, 14(3):199–208, 2011.

[29] Chris Berka, Daniel J. Levendowski, Milenko M. Cvetinovic, Miroslav M. Petrovic, Gene Davis, Michelle N. Lumicao, Vladimir T. Zivkovic, Miodrag V. Popovic, and Richard Olmstead. Real-time analysis of eeg indexes of alertness, cognition, and memory acquired with a wireless eeg headset. *International Journal of Human–Computer Interaction*, 17(2):151–170, 2004.

[30] Chris Berka, Daniel J. Levendowski, Michelle N. Lumicao, Alan Yau, Gene Davis, Vladimir T. Zivkovic, Richard E. Olmstead, Patrice D. Tremoulet, and Patrick L. Craven. Eeg correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine*, 78(Supplement 1):B231–B244, 2007.

[31] Ashok J. Bharucha, Vivek Anand, Jodi Forlizzi, Mary Amanda Dew, Charles F. Reynolds, Scott Stevens, and Howard Wactlar. Intelligent assistive technology applications to dementia care: Current capabilities, limitations, and future challenges. *The American Journal of Geriatric Psychiatry*, 17(2):88 – 104, 2009.

[32] Dominik Bial, Dagmar Kern, Florian Alt, and Albrecht Schmidt. Enhancing outdoor navigation systems through vibrotactile feedback. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 1273–1278. ACM, 2011.

[33] Ann Blandford, Dominic Furniss, and Stephann Makri. Qualitative hci research: Going behind the scenes. *Synthesis Lectures on Human-Centered Informatics*, 9(1):1–115, 2016.

[34] Rubén Blasco, Álvaro Marco, Roberto Casas, Diego Cirujano, and Richard Picking. A smart kitchen for ambient assisted living. *Sensors (Basel, Switzerland)*, 14(1):1629–1653, 2013.

[35] Jonas Blattgerste, Benjamin Strenge, Patrick Renner, Thies Pfeiffer, and Kai Essig. Comparing conventional and augmented reality instructions for manual assembly tasks. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '17, pages 75–82, New York, NY, USA, 2017. ACM.

[36] Sarah Blum, Stefan Debener, Reiner Emkes, Nils Volkening, Sebastian Fudickar, and Martin G. Bleichner. EEG Recording and Online Signal Processing on Android: A Multiapp Framework for Brain-Computer Interfaces on Smartphone, 2017.

[37] Patrick Boissy, Hélène Corriveau, François Michaud, Daniel Labonté, and Marie-Pier Royer. A qualitative study of in-home robotic telepresence for home care of community-living elderly subjects. *Journal of Telemedicine and Telecare*, 13(2):79–84, 2007.

[38] Leonardo Bonanni, Chia-Hsun Lee, and Ted Selker. Attention-Based Design of Augmented Reality Interfaces. 2005.

[39] Christopher Boorse. *Disability and Medical Theory*, pages 55–88. Springer Netherlands, Dordrecht, 2010.

[40] Bruno Bouchard, Sylvain Giroux, and Abdenour Bouzouane. A smart home agent for plan recognition of cognitively-impaired patients. *Journal of Computers (Finland)*, 1(5):53–62, 2006.

[41] Wolfram Boucsein. *Electrodermal activity*. Springer Science & Business Media, 2012.

[42] Stacy M. Branham and Shaun K. Kane. Collaborative accessibility: How blind and sighted companions co-create accessible home spaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2373–2382. ACM, 2015.

[43] Stacy M. Branham and Shaun K. Kane. The invisible work of accessibility: how blind employees manage accessibility in mixed-ability workplaces. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pages 163–171. ACM, 2015.

[44] Philip Brey. Freedom and privacy in ambient intelligence. *Ethics and Information Technology*, 7(3):157–166, 2005.

[45] Karel A. Brookhuis and Dick de Waard. On the assessment of (mental) workload and other subjective qualifications. *Ergonomics*, 45(14):1026–1030, 2002.

[46] Anne-Marie Brouwer, Maarten A. Hogervorst, Jan B. F. van Erp, Tobias Heffelaar, Patrick H. Zimmerman, and Robert Oostenveld. Estimating workload using eeg spectral power and erps in the n-back task. *Journal of Neural Engineering*, 9(4):045008, 2012.

[47] Anne-Marie Brouwer, Maarten A. Hogervorst, Jan B. F. Van Erp, Tobias Heffelaar, Patrick H. Zimmerman, and Robert Oostenveld. Estimating workload using eeg spectral power and erps in the n-back task. *Journal of neural engineering*, 9(4):045008, 2012.

[48] S. J. Brownsell, David A. Bradley, R. Bragg, P. Catlin, and J. Carlier. Do community alarm users want telecare? *Journal of Telemedicine and Telecare*, 6(4):199–204, 2000.

[49] Marc Brysbaert and Tatjana Nazir. Visual constraints in written word recognition: evidence from the optimal viewing-position effect. *Journal of Research in Reading*, 28(3):216–228, 2005.

[50] Martha D. Buffum, Evelyn Hutt, Victor T. Chang, H. Michael, and A. Lynn Snow. Cognitive impairment and pain management: review of issues and challenges. *Journal of rehabilitation research & development*, 44(2):315–330, 2007.

[51] A. Bulling and H. Gellersen. Toward mobile eye-based human-computer interaction. *IEEE Pervasive Computing*, 9(4):8–12, October 2010.

[52] Andreas Bulling, Daniel Roggen, and Gerhard Tröster. It's in your eyes: Towards context-awareness and mobile hci using wearable eog goggles. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, UbiComp '08, pages 84–93, New York, NY, USA, 2008. ACM.

[53] Andreas Bulling and Thorsten O. Zander. Cognition-aware computing. *IEEE Pervasive Computing*, 13(3):80–83, 2014.

[54] Christopher G. Burns and Stephen H. Fairclough. Use of auditory event-related potentials to measure immersion during a computer game. *International Journal of Human-Computer Studies*, 73:107–114, 2015.

[55] George Bush, Phan Luu, and Michael I. Posner. Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, 4(6):215 – 222, 2000.

[56] Sarah Buss. Valuing autonomy and respecting persons: Manipulation, seduction, and the basis of moral constraints. *Ethics*, 115(2):195–235, 2005.

[57] Sarah Buss and Andrea Westlund. Personal autonomy. 2002.

[58] Sebastian Büttner, Markus Funk, Oliver Sand, and Carsten Röcker. Using head-mounted displays and in-situ projection for assistive systems – a comparison. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2016. ACM.

[59] Sebastian Büttner, Henrik Mucha, Markus Funk, Thomas Kosch, Mario Aehnelt, Sebastian Robert, and Carsten Röcker. The design space of augmented and virtual reality applications for assistive environments in manufacturing: A visual approach. In *Proceedings of the 10th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2017. ACM.

[60] Sebastian Büttner, Oliver Sand, and Carsten Röcker. Extending the design space in industrial manufacturing through mobile projection. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 1130–1133. ACM, 2015.

[61] Sebastian Büttner, Oliver Sand, and Carsten Röcker. Extending the design space in industrial manufacturing through mobile projection. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '15, pages 1130–1133, New York, NY, USA, 2015. ACM.

[62] Andreas Butz, Michael Schneider, and Mira Spassova. Searchlight–a lightweight search function for pervasive environments. In *Pervasive Computing*, pages 351–356. Springer, 2004.

[63] Kate Cain, Jane Oakhill, and Peter Bryant. Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of educational psychology*, 96(1):31, 2004.

[64] Monica S. Castelhano and Paul Muter. Optimizing the reading of electronic text using rapid serial visual presentation. *Behaviour & Information Technology*, 20(4):237–247, 2001.

[65] Monica S. Castelhano and Paul Muter. Optimizing the reading of electronic text using rapid serial visual presentation. *Behaviour & Information Technology*, 20(4):237–247, 2001.

[66] Marie Chan, Eric Campo, Daniel Estève, and Jean-Yves Fourniols. Smart homes-current features and future perspectives. *Maturitas*, 64(2):90–97, 2009.

[67] Yao-Jen Chang, Ya-Shu Kang, Yao-Sheng Chang, and Hung-Huan Liu. Arcoach 2.0: Optimizing a vocational prompting system based on augmented reality for people with cognitive impairments. In *Proceedings of the 17th International ACM SIGACCESS*

Conference on Computers and Accessibility, ASSETS '15, pages 313–314, New York, NY, USA, 2015. ACM.

[68] Chien-Hsiung Chen and Yu-Hung Chien. Effects of rsvp display design on visual performance in accomplishing dual tasks with small screens. *International Journal of Design*, 1(1), 2007.

[69] Daniel Chen, Jamie Hart, and Roel Vertegaal. Towards a Physiological Model of User Interruptability. In Cécilia Baranauskas, Philippe Palanque, Julio Abascal, and Simone Diniz Junqueira Barbosa, editors, *Human-Computer Interaction – INTERACT 2007*, volume 4663, pages 439–451. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[70] Hsuan-Chih Chen. Reading normal versus rapid, sequential text formats: Effects of text structure and reading ability, 1983.

[71] Rui Chen, Tiantian Xie, Yingtao Xie, Tao Lin, and Ningjiu Tang. Do speech features for detecting cognitive load depend on specific languages? In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI '16, pages 76–83, New York, NY, USA, 2016. ACM.

[72] Pei Yu Chi, Jen Hao Chen, Hao Hua Chu, and Jin Ling Lo. Enabling calorie-aware cooking in a smart kitchen. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5033 LNCS:116–127, 2008.

[73] Pei-Yu Peggy Chi, Hao-Hua Chu, Jen-hao Chen, and Bing-Yu Chen. Enabling nutrition-aware cooking in a smart kitchen. *Persuasive 2008, Lncs 5033*, pages 116–127, 2007.

[74] Burcu Cinaz, Bert Arnrich, Roberto Marca, and Gerhard Tröster. Monitoring of mental workload levels during an everyday life office-work scenario. *Personal Ubiquitous Comput.*, 17(2):229–239, February 2013.

[75] Andy Clark and David Chalmers. The extended mind. *Analysis*, 58(1):7–19, 1998.

[76] William G. Cochran and Gertrude M. Cox. Experimental designs. 1950.

[77] Mark Coeckelbergh. Health care, capabilities, and ai assistive technologies. *Ethical Theory and Moral Practice*, 13(2):181–190, Apr 2010.

[78] Han Collewijn and Ernst P. Tamminga. Human smooth and saccadic eye movements during voluntary pursuit of different target motions on different backgrounds. *The Journal of physiology*, 351:217, 1984.

[79] R. Contreras, J. Ghajar, S. Bahar, and M. Suh. Effect of cognitive load on eye-target synchronization during smooth pursuit eye movement. *Brain research*, 1398:55–63, 2011.

[80] Anne E. Cook, Jennifer G. Halleran, and Edward J. O'Brien. What is readily available during reading? a memory-based view of text processing. *Discourse Processes*, 26(2-3):109–129, 1998.

[81] Karen L. Courtney. Privacy and senior willingness to adopt smart home information technology in residential care facilities. 2008.

[82] Karen L. Courtney, George Demeris, Marilyn Rantz, and Marjorie Skubic. Needing smart home technologies: the perspectives of older adults in continuing care retirement communities. 2008.

[83] Nelson Cowan. The magical mystery four: How is working memory capacity limited, and why? *Current directions in psychological science*, 19(1):51–57, 2010.

[84] Mihaly Csikszentmihalyi. Flow and the psychology of discovery and invention. *HarperPerennial, New York*, 39, 1997.

[85] Mihaly Csikszentmihalyi. *Flow: The psychology of happiness*. Random House, 2013.

[86] Tim R. H. Cutmore and Daniel A. James. Identifying and reducing noise in psychophysiological recordings. *International Journal of Psychophysiology*, 32(2):129 – 150, 1999.

[87] Anthony J. Cuvo, Paula K. Davis, Mark F. O'Reilly, Brenda M. Mooney, and Ruth Crowley. Promoting stimulus control with textual prompts and performance feedback for persons with mild disabilities. *Journal of applied behavior analysis*, 25(2):477– 489, 1992.

[88] Meredyth Daneman and Patricia A. Carpenter. Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4):450 – 466, 1980.

[89] Laura Davy. Philosophical inclusive design: Intellectual disability and the limits of individual autonomy in moral and political theory. *Hypatia*, 30(1):132–148, 2014.

[90] Melissa Dawe. Desperately seeking simplicity: How young adults with cognitive disabilities and their families adopt assistive technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 1143–1152, New York, NY, USA, 2006. ACM.

[91] Alexander De Luca, Roman Weiss, and Heiko Drewes. Evaluation of eye-gaze interaction methods for security enhanced pin-entry. In *Proceedings of the 19th australasian conference on computer-human interaction: Entertaining user interfaces*, pages 199–202. ACM, 2007.

[92] George Demiris, Brian K. Hensel, Marjorie Skubic, and Marilyn Rantz. Senior residents' perceived need of and preferences for "smart home" sensor technologies. *International journal of technology assessment in health care*, 24(01):120–124, 2008.

[93] George Demiris, Marilyn J. Rantz, Myra A. Aud, Karen D. Marek, Harry W. Tyrer, Marjorie Skubic, and Ali A. Hussam. Older adults' attitudes towards and perceptions of 'smart home' technologies: a pilot study. *Medical Informatics and the Internet in Medicine*, 29(2):87–94, 2004.

[94] Antonio Di Carlo, Marzia Baldereschi, Luigi Amaducci, Stefania Maggi, Francesco Grigoletto, Guglielmo Scarlato, and Domenico Inzitari. Cognitive impairment without dementia in older people: Prevalence, vascular risk factors, impact on disability. the italian longitudinal study on aging. *Journal of the American Geriatrics Society*, 48(7):775–782, 2000.

[95] Tilman Dingler. Cognition-aware systems as mobile personal assistants. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, pages 1035–1040, New York, NY, USA, 2016. ACM.

[96] Tilman Dingler, Rufat Rzayev, Valentin Schwind, and Niels Henze. Rsvp on the go: Implicit reading support on smart watches through eye tracking. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers*, ISWC '16, pages 116–119, New York, NY, USA, 2016. ACM.

[97] Tilman Dingler, Rufat Rzayev, Alireza Sahami Shirazi, and Niels Henze. Designing consistent gestures across device types: Eliciting rsvp controls for phone, watch, and glasses. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 419:1–419:12, New York, NY, USA, 2018. ACM.

[98] Tilman Dingler, Alireza Sahami Shirazi, Kai Kunze, and Albrecht Schmidt. Assessment of stimuli for supporting speed reading on electronic devices. In *Proceedings of the 6th Augmented Human International Conference*, AH '15, pages 117–124, New York, NY, USA, 2015. ACM.

[99] Debra Dobbs, J. Kevin Eckert, Bob Rubinstein, Lynn Keimig, Leanne Clark, Ann Christine Frankowski, and Sheryl Zimmerman. An ethnographic study of stigma and ageism in residential care or assisted living. *The Gerontologist*, 48(4):517–526, 2008.

[100] Michael Doppelmayr, Wolfgang Klimesch, T. Pachinger, and B. Ripper. Individual differences in brain dynamics: important implications for the calculation of event-related band power. *Biological cybernetics*, 79(1):49–57, 1998.

[101] Andrew T. Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4):455–470, Nov 2002.

[102] Andrew T. Duchowski. A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, & Computers*, 34(4):455–470, 2002.

[103] Andrew T. Duchowski. *Eye tracking methodology: Theory and practice*, volume 373. Springer Science & Business Media, 2017.

[104] Andrew T. Duchowski. Gaze-based interaction: A 30 year retrospective. *Computers & Graphics*, 73:59–69, 2018.

[105] Andrew T. Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 282:1–282:13, New York, NY, USA, 2018. ACM.

[106] Andrew T. Duchowski, Vinay Shivashankaraiah, Tim Rawls, Anand K. Gramopadhye, Brian J. Melloy, and Barbara Kanki. Binocular eye tracking in virtual reality for inspection training. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ETRA '00, pages 89–96, New York, NY, USA, 2000. ACM.

[107] William Easley, Michele A. Williams, Ali Abdolrahmani, Caroline Galbraith, Stacy M. Branham, Amy Hurst, and Shaun K. Kane. Let's get lost: Exploring social norms in predominately blind environments. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2034–2040. ACM, 2016.

[108] Florian Echtler, Fabian Sturm, Kay Kindermann, Gudrun Klinker, Joachim Stilla, Jörn Trilk, and Hesam Najafi. *The Intelligent Welding Gun: Augmented Reality for Experimental Vehicle Construction*, pages 333–360. Springer London, London, 2004.

[109] Mai ElKomy, Yomna Abdelrahman, Markus Funk, Tilman Dingler, Albrecht Schmidt, and Slim Abdennadher. Abbas: An adaptive bio-sensors based assistive system. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, New York, NY, USA, 2017. ACM.

[110] Stephen A. Engel, David E. Rumelhart, Brian A. Wandell, Adrian T. Lee, Gary H. Glover, Eduardo-Jose Chichilnisky, and Michael N. Shadlen. fmri of human visual cortex. *Nature*, 1994.

[111] Douglas Engelbart. Augmenting human intellect. *Summary report AFOSR-3223 under Contract AF*, 49(638):1024, 1962.

[112] Douglas C Engelbart. Toward augmenting the human intellect and boosting our collective iq. *Communications of the ACM*, 38(8):30–33, 1995.

[113] Randall W. Engle, Stephen W. Tuholski, James E. Laughlin, and Andrew R. A. Conway. Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General*, 128(3):309, 1999.

[114] Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. Orbits: gaze interaction for smart watches using smooth pursuit eye movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 457–466. ACM, 2015.

[115] Nir Eyal. Using informed consent to save trust. *Journal of Medical Ethics*, 2012.

[116] Stephen H. Fairclough. Fundamentals of physiological computing. *Interacting with computers*, 21(1-2):133–145, 2009.

[117] A. Fink, R. H. Grabner, C. Neuper, and A. C. Neubauer. Eeg alpha band dissociation with increasing task demands. *Cognitive Brain Research*, 24(2):252 – 259, 2005.

[118] Bruce J. Fisch and Rainer Spehlmann. *Fisch and Spehlmann's EEG primer: basic principles of digital and analog EEG*. Elsevier Health Sciences, 1999.

[119] Kenneth I. Forster. Visual perception of rapidly presented word sequences of varying complexity. *Perception & Psychophysics*, 8(4):215–221, Jul 1970.

[120] Harry G. Frankfurt. *Freedom of the Will and the Concept of a Person*, pages 127–144. Humana Press, Totowa, NJ, 1988.

[121] C. Ailie Fraser, Tovi Grossman, and George Fitzmaurice. Webuild: Automatically distributing assembly tasks among collocated workers to improve coordination. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 1817–1830, New York, NY, USA, 2017. ACM.

[122] Christopher Frauenberger, Julia Makhaeva, and Katharina Spiel. Designing smart objects with autistic children: Four design exposès. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 130–139. ACM, 2016.

[123] Jérémy Frey, Maxime Daniel, Julien Castet, Martin Hachet, and Fabien Lotte. Framework for electroencephalography-based evaluation of user experience. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 2283–2294, New York, NY, USA, 2016. ACM.

[124] Jérémy Frey, Christian Mühl, Fabien Lotte, and Martin Hachet. Review of the Use of Electroencephalography as an Evaluation Method for Human-Computer Interaction. *arXiv:1311.2222 [cs]*, November 2013. arXiv: 1311.2222.

[125] Bruce H. Friedman and Julian F. Thayer. Facial muscle activity and EEG recordings: redundancy analysis. *Electroencephalography and Clinical Neurophysiology*, 79(5):358 – 360, 1991.

[126] R. Buckminster Fuller and Kiyoshi Kuromiya. *Critical path*. Macmillan, 1981.

[127] M. Funk and A. Schmidt. Cognitive assistance in the workplace. *IEEE Pervasive Computing*, 14(3):53–55, July 2015.

[128] Markus Funk, Andreas Bächler, Liane Bächler, Oliver Korn, Christoph Krieger, Thomas Heidenreich, and Albrecht Schmidt. Comparing projected in-situ feedback at the manual assembly workplace with impaired workers. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, page 1. ACM, 2015.

[129] Markus Funk, Andreas Bächler, Liane Bächler, Thomas Kosch, Thomas Heidenreich, and Albrecht Schmidt. Working with augmented reality? a long-term analysis of in-situ instructions at the assembly workplace. In *Proceedings of the 10th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2017. ACM.

[130] Markus Funk, Tilman Dingler, Jennifer Cooper, and Albrecht Schmidt. Stop helping me - i'm bored! why assembly assistance needs to be adaptive. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015.

[131] Markus Funk, Juana Heusler, Elif Akcay, Klaus Weiland, and Albrecht Schmidt. Haptic, auditory, or visual? towards optimal error feedback at manual assembly workplaces. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2016. ACM.

[132] Markus Funk, Thomas Kosch, Scott W. Greenwald, and Albrecht Schmidt. A benchmark for interactive augmented reality instructions for assembly tasks. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, pages 253–257. ACM, 2015.

[133] Markus Funk, Thomas Kosch, Romina Kettner, Oliver Korn, and Albrecht Schmidt. motioneap: An overview of 4 years of combining industrial assembly with augmented reality for industry 4.0. In *Proceedings of the 16th International Conference on Knowledge Technologies and Data-driven Business*, i-KNOW '16, New York, NY, USA, 2016. ACM.

[134] Markus Funk, Thomas Kosch, and Albrecht Schmidt. Interactive worker assistance: Comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions. pages 934–939, 2016.

[135] Markus Funk, Thomas Kosch, Katrin Wolf, Pascal Knierim, Sven Mayer, and Albrecht Schmidt. Automatic projection positioning based on surface suitability. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays*, PerDis'16, New York, NY, USA, 2016. ACM.

[136] Markus Funk, Sven Mayer, and Albrecht Schmidt. Using in-situ projection to support cognitively impaired workers at the workplace. In *Proceedings of the 17th international ACM SIGACCESS conference on Computers & accessibility*, 2015.

[137] Markus Funk, Sven Mayer, and Albrecht Schmidt. Using in-situ projection to support cognitively impaired workers at the workplace. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '15, pages 185–192, New York, NY, USA, 2015. ACM.

[138] Markus Funk, Alireza Sahami Shirazi, Sven Mayer, Lars Lischke, and Albrecht Schmidt. Pick from here! - an interactive mobile cart using in-situ projection for order picking. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015.

[139] Edith Galy, Magali Cariou, and Claudine Mélan. What is the relationship between mental workload factors and cognitive load types? *International Journal of Psychophysiology*, 83(3):269 – 275, 2012.

[140] Erin Gannon, Jibo He, Xuefei Gao, and Barbara Chaparro. Rsvp reading on a smart watch. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 60(1):1130–1134, 2016.

[141] Steffen Gauglitz, Cha Lee, Matthew Turk, and Tobias Höllerer. Integrating the physical environment into mobile remote collaboration. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '12, pages 241–250, New York, NY, USA, 2012. ACM.

[142] Tsvetozar Georgiev. Investigation of the user's text reading speed on mobile devices. In *Proceedings of the 13th International Conference on Computer Systems and Technologies*, CompSysTech '12, pages 329–336, New York, NY, USA, 2012. ACM.

[143] A. Gevins, M. E. Smith, L. McEvoy, and D. Yu. High-resolution eeg mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral Cortex*, 7(4):374–385, 1997.

[144] Alan Gevins and Michael E. Smith. Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics Science*, 4(1-2):113–131, 2003.

[145] Alan Gevins, Michael E. Smith, Harrison Leong, Linda McEvoy, Susan Whitfield, Robert Du, and Georgia Rush. Monitoring working memory load during computer-based tasks with eeg pattern recognition methods. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(1):79–91, 1998.

[146] Alan Gevins, Michael E. Smith, Linda McEvoy, and Daphne Yu. High-resolution eeg mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral cortex*, 7(4):374–385, 1997.

[147] Christiane Glatz, Stas S. Krupenia, Heinrich H. Bülthoff, and Lewis L. Chuang. Use the right sound for the right job: Verbal commands and auditory icons for a task-management system favor different information processes in the brain. In *Proceedings*

of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18, New York, NY, USA, 2018. Association for Computing Machinery.

[148] Pierre Gloor. Hans berger on electroencephalography. *American Journal of EEG Technology*, 9(1):1–8, 1969.

[149] Benedikt Gollan, Michael Haslgrübler, and Alois Ferscha. Demonstrator for extracting cognitive load from pupil dilation for attention management services. In *Ubi-Comp'16*, pages 1566–1571, New York, NY, USA, 2016. ACM.

[150] Arthur C. Graesser, Murray Singer, and Tom Trabasso. Constructing inferences during narrative text comprehension. *Psychological review*, 101(3):371, 1994.

[151] Madeleine A. Grealy, David A. Johnson, and Simon K. Rushton. Improving cognitive function after brain injury: the use of exercise and virtual reality. *Archives of physical medicine and rehabilitation*, 80(6):661–667, 1999.

[152] David Grimes, Desney S. Tan, Scott E. Hudson, Pradeep Shenoy, and Rajesh P.N. Rao. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 835–844, New York, NY, USA, 2008. ACM.

[153] DMA Gronwall. Paced auditory serial-addition task: a measure of recovery from concussion. *Perceptual and motor skills*, 44(2):367–373, 1977.

[154] Wei Guo and Jingtao Wang. Smartrsvp: Facilitating attentive speed reading on small screen wearable devices. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, pages 1640–1647, New York, NY, USA, 2017. ACM.

[155] Pavel Gurevich, Joel Lanir, Benjamin Cohen, and Ran Stone. Teleadvisor: A versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 619–622, New York, NY, USA, 2012. ACM.

[156] Eija Haapalainen, SeungJun Kim, Jodi F. Forlizzi, and Anind K. Dey. Psychophysiological measures for assessing cognitive load. In *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, UbiComp '10, pages 301–310, New York, NY, USA, 2010. ACM.

[157] Ian Hacking. *The taming of chance*, volume 17. Cambridge University Press, 1990.

[158] Peter A. Hancock and Paula A. Desmond. *Stress, workload, and fatigue*. CRC Press, 2000.

[159] Sandra G. Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 50, pages 904–908. Sage Publications, 2006.

[160] Sandra G. Hart. Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, 2006.

[161] Sandra G. Hart and Lowell E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In Peter A. Hancock and Najmedin Meshkati, editors, *Human Mental Workload*, volume 52 of *Advances in Psychology*, pages 139 – 183. North-Holland, 1988.

[162] George W. Hartmann. Gestalt psychology: A survey of facts and principles. 1935.

[163] Atsushi Hashimoto, Naoyuki Mori, Takuya Funatomi, Yoko Yamakata, Koh Kakusho, and Michihiko Minoh. Smart Kitchen : A User Centric Cooking Support System. *Proceedings of IPMU*, pages 848–854, 2008.

[164] Richard P. Heitz, Josef C. Schrock, Tabitha W. Payne, and Randall W. Engle. Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology*, 45(1):119–129, 2008.

[165] Suzana Herculano-Houzel. The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, 3, 2009.

[166] Eckhard H. Hess and James M. Polt. Pupil size in relation to mental activity during simple problem-solving. *Science*, 143(3611):1190–1192, 1964.

[167] James Hodge, Sarah Foley, Rens Brankaert, Gail Kenning, Amanda Lazar, Jennifer Boger, and Kellie Morrissey. Relational, flexible, everyday: Learning from ethics in dementia research. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–16, New York, NY, USA, 2020. Association for Computing Machinery.

[168] Hunter G. Hoffman. Virtual-reality therapy. *SCIENTIFIC AMERICAN-AMERICAN EDITION-*, 291:58–65, 2004.

[169] Nina Hollender, Cristian Hofmann, Michael Deneke, and Bernhard Schmitz. Integrating cognitive load theory and concepts of human-computer interaction. *Computers in Human Behavior*, 26(6):1278 – 1288, 2010. Online Interactivity: Role of Technology in Behavior Change.

[170] Keum-Shik Hong, Noman Naseer, and Yun-Hee Kim. Classification of prefrontal and motor cortex signals for three-class fnirs–bci. *Neuroscience Letters*, 587:87 – 92, 2015.

[171] Clare J. Hooper, Anne Preston, Madeline Balaam, Paul Seedhouse, Daniel Jackson, Cuong Pham, Cassim Ladha, Karim Ladha, Thomas Plötz, and Patrick Olivier. The french kitchen: Task-based learning in an instrumented kitchen. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, UbiComp '12, pages 193–202, New York, NY, USA, 2012. ACM.

[172] Matthias Hoppe, Marinus Burger, Albrecht Schmidt, and Thomas Kosch. Dronos: A flexible open-source prototyping framework for interactive drone routines. In *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 2019.

[173] Matthias Hoppe, Pascal Knierim, Thomas Kosch, Markus Funk, Lauren Futami, Stefan Schneegass, Niels Henze, Albrecht Schmidt, and Tonja Machulla. Vrhaptic-drones: Providing haptics in virtual reality through quadcopters. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, pages 7–18. ACM, 2018.

[174] Matthias Hoppe, Thomas Kosch, Pascal Knierim, Markus Funk, and Albrecht Schmidt. Are drones ready for takeoff? reflecting on challenges and opportunities in human-drone interfaces. In *International Workshop on Human-Drone Interaction at the CHI Conference on Human Factors in Computing Systems*, 2019.

[175] Matthias Hoppe, Yannick Weiß, Marinus Burger, and Thomas Kosch. Don't drone yourself in work: Discussing dronos as a framework for human-drone interaction. In *2nd International Workshop on Human-Drone Interaction at the CHI Conference on Human Factors in Computing Systems*, 2020.

[176] Thomas E. Hutchinson, K. Preston White, Worthy N. Martin, Kelly C. Reichert, and Lisa A. Frey. Human-computer interaction using eye-gaze input. *IEEE Transactions on systems, man, and cybernetics*, 19(6):1527–1534, 1989.

[177] Shamsi T. Iqbal and Eric Horvitz. Disruption and recovery of computing tasks: Field study, analysis, and directions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 677–686, New York, NY, USA, 2007. ACM.

[178] Poika Isokoski. Text input methods for eye trackers using off-screen targets. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 15–21. ACM, 2000.

[179] R. J. Jacob and Keith S. Karn. Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind*, 2(3):4, 2003.

[180] Robert J. K. Jacob. What you look at is what you get: Eye movement-based interaction techniques. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '90, pages 11–18, New York, NY, USA, 1990. ACM.

[181] Ole Jensen, Jack Gelfand, John Kounios, and John E. Lisman. Oscillations in the alpha band (9-12 hz) increase with memory load during retention in a short-term memory task. *Cerebral Cortex*, 12(8):877, 2002.

[182] James F. Juola, David Haugh, Scott Trast, F. Richard Ferraro, and Michael Lieb-
haber. Reading with and without eye movements. In J.K. O'REGAN and A. LEVY-
SCHOEN, editors, *Eye Movements from Physiology to Cognition*, pages 499 – 508.
Elsevier, Amsterdam, 1987.

[183] Marcel A Just and Patricia A Carpenter. A theory of reading: From eye fixations to
comprehension. *Psychological review*, 87(4):329, 1980.

[184] Marcel Adam Just and Patricia A Carpenter. Eye fixations and cognitive processes.
Cognitive Psychology, 8(4):441 – 480, 1976.

[185] Daniel Kahneman and Jackson Beatty. Pupil diameter and load on memory. *Science*,
154(3756), 1966.

[186] Michael J. Kane, Andrew R. A. Conway, Timothy K. Miura, and Gregory J. H.
Colflesh. Working memory, attention control, and the n-back task: a question of
construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cog-
nition*, 33(3):615, 2007.

[187] Ted J. Kaptchuk. Effect of interpretive bias on research evidence. *Bmj*,
326(7404):1453–1455, 2003.

[188] Lari Kärkkäinen and Jari Laarni. Designing for small display screens. In *Proceedings
of the Second Nordic Conference on Human-computer Interaction*, NordiCHI '02,
pages 227–230, New York, NY, USA, 2002. ACM.

[189] Jakob Karolus, Annika Kilian, Thomas Kosch, Albrecht Schmidt, and Paweł W. Woz-
niak. Hit the thumb jack! using electromyography to augment the piano keyboard.
In *Proceedings of the 2020 Designing Interactive Systems Conference*, DIS '20, New
York, NY, USA, 2020. ACM.

[190] Jakob Karolus, Hendrik Schuff, Thomas Kosch, Paweł W. Wozniak, and Albrecht
Schmidt. Emguitar: Assisting guitar playing with electromyography. In *Proceedings
of the 2018 Designing Interactive Systems Conference*, DIS '18, pages 651–655, New
York, NY, USA, 2018. ACM.

[191] Ivo Käthner, Selina C. Wriessnegger, Gernot R. Müller-Putz, Andrea Kübler, and
Sebastian Halder. Effects of mental workload and fatigue on the p300, alpha and theta
band power during operation of an erp (p300) brain–computer interface. *Biological
psychology*, 102:118–129, 2014.

[192] Z. A. Keirn and J. I. Aunon. A new mode of communication between man and his
surroundings. *IEEE Transactions on Biomedical Engineering*, 37(12):1209–1214,
Dec 1990.

[193] Mohamed Khamis, Florian Alt, and Andreas Bulling. A field study on spontaneous gaze-based interaction with a public display using pursuits. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 863–872. ACM, 2015.

[194] Mohamed Khamis, Ozan Saltuk, Alina Hang, Katharina Stolz, Andreas Bulling, and Florian Alt. Textpursuits: Using text for pursuits-based interaction and calibration with public displays. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016.

[195] M. Asif Khawaja, Fang Chen, and Nadine Marcus. Using language complexity to measure cognitive load for adaptive interaction design. In *Proceedings of the 15th International Conference on Intelligent User Interfaces*, IUI '10, pages 333–336, New York, NY, USA, 2010. ACM.

[196] M. Asif Khawaja, Natalie Ruiz, and Fang Chen. Think before you talk: An empirical study of relationship between speech pauses and cognitive load. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat*, OZCHI '08, pages 335–338, New York, NY, USA, 2008. ACM.

[197] Peter Kiefer, Ioannis Giannopoulos, Andrew Duchowski, and Martin Raubal. Measuring cognitive load for map tasks through pupil diameter. In *International Conference on Geographic Information Science*, pages 323–337. Springer, 2016.

[198] Yoshifumi Kitamura, Yoshihisa Yamaguchi, Imamizu Hiroshi, Fumio Kishino, and Mitsuo Kawato. Things happening in the brain while humans learn to use new tools. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 417–424, New York, NY, USA, 2003. ACM.

[199] W. Klimesch, H. Schimke, and G. Pfurtscheller. Alpha frequency, cognitive load and memory performance. *Brain Topography*, 5(3):241–251, 1993.

[200] W. Klimesch, F. Vogt, and M. Doppelmayr. Interindividual differences in alpha and theta power reflect memory performance. *Intelligence*, 27(4):347 – 362, 1999.

[201] Wolfgang Klimesch. Eeg alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. *Brain Research Reviews*, 29(2):169 – 195, 1999.

[202] Wolfgang Klimesch, Bärbel Schack, and Paul Sauseng. The functional significance of theta and upper alpha oscillations. *Experimental psychology*, 52(2):99–108, 2005.

[203] Jeff Klingner, Rakshit Kumar, and Pat Hanrahan. Measuring the task-evoked pupillary response with a remote eye tracker. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 69–72. ACM, 2008.

[204] Pascal Knierim, Markus Funk, Thomas Kosch, Anton Fedosov, Tamara Müller, Benjamin Schopf, Marc Weise, and Albrecht Schmidt. Ubibeam++: Augmenting interactive projection with head-mounted displays. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*, NordiCHI '16, pages 112:1–112:6, New York, NY, USA, 2016. ACM.

[205] Pascal Knierim, Thomas Kosch, Alexander Achberger, and Markus Funk. Flyables: Exploring 3d interaction spaces for levitating tangibles. In *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, TEI '18, New York, NY, USA, 2018. ACM.

[206] Pascal Knierim, Thomas Kosch, Johannes Groschopp, and Albrecht Schmidt. Opportunities and challenges of text input in portable virtual reality. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, New York, NY, USA, 2020. ACM.

[207] Pascal Knierim, Thomas Kosch, Matthias Hoppe, and Albrecht Schmidt. Challenges and opportunities of mixed reality systems in education. *Mensch und Computer 2018– Proceedings*, 2018.

[208] Pascal Knierim, Thomas Kosch, Gabrielle LaBorwit, and Albrecht Schmidt. Altering the speed of reality? exploring visual slow-motion to amplify human perception using augmented reality. In *Proceedings of the 1st Augmented Humans International Conference*, AHs '20, New York, NY, USA, 2020. ACM.

[209] Pascal Knierim, Thomas Kosch, Valentin Schwind, Markus Funk, Francisco Kiss, Stefan Schneegass, and Niels Henze. Tactile drones - providing immersive tactile feedback in virtual reality through quadcopters. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, New York, NY, USA, 2017. ACM.

[210] Oliver Korn, Markus Funk, Stephan Abele, Thomas Hörz, and Albrecht Schmidt. Context-aware assistive systems at the workplace: analyzing the effects of projection and gamification. In *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*, page 38. ACM, 2014.

[211] Oliver Korn, Markus Funk, and Albrecht Schmidt. Assistive systems for the workplace: Towards context-aware assistance. *Assistive Technologies for Physical and Cognitive Disabilities*, pages 121–133, 2015.

[212] Oliver Korn, Markus Funk, and Albrecht Schmidt. Design approaches for the gamification of production environments. a study focusing on acceptance. In *Proceedings of the 8th International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2015. ACM.

[213] Oliver Korn, Markus Funk, and Albrecht Schmidt. Towards a gamification of industrial production. a comparative study in sheltered work environments. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, New York, NY, USA, 2015. ACM.

[214] Oliver Korn, Markus Funk, and Albrecht Schmidt. Towards a gamification of industrial production: A comparative study in sheltered work environments. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '15, pages 84–93, New York, NY, USA, 2015. ACM.

[215] Oliver Korn, Albrecht Schmidt, and Thomas Hörz. The potentials of in-situ-projection for augmented workplaces in production: a study with impaired persons. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 979–984. ACM, 2013.

[216] A. V. Korshakov, A. A. Frolov, and P. D. Bobrov. On-line automatic suppression of artifacts in multi-dimensional signals using ica. In *Abstr. 10th Europ. Conf. on Non-Destructive Testing*, page 370, 2010.

[217] Thomas Kosch, Yomna Abdelrahman, Markus Funk, and Albrecht Schmidt. One size does not fit all: Challenges of providing interactive worker assistance in industrial settings. In *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*, UbiComp '17, pages 1006–1011, New York, NY, USA, 2017. ACM.

[218] Thomas Kosch, Robin Boldt, Matthias Hoppe, Pascal Knierim, and Markus Funk. Exploring the optimal point of view in third person out-of-body experiences. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2016. ACM.

[219] Thomas Kosch and Lewis L. Chuang. Investigating the Impact of Assistive Technologies on Working Memory Load in Manual Assembly through Electroencephalography. *Human Neuroscience Archive*, (114).

[220] Thomas Kosch and Lewis L. Chuang. Investigating the Influence of RSVP Display Parameters on Working Memory Load using Electroencephalography. In *2nd International Conference on Neuroadaptive Technology*, 2019.

[221] Thomas Kosch, Markus Funk, Albrecht Schmidt, and Lewis L. Chuang. Identifying Cognitive Assistance with Mobile Electroencephalography: A Case Study with In-Situ Projections for Manual Assembly. *Proc. ACM Hum.-Comput. Interact.*, 2(EICS):11:1–11:20, June 2018.

[222] Thomas Kosch, Markus Funk, Daniel Vietz, Marc Weise, Tamara Müller, and Albrecht Schmidt. Dronectrl: A tangible remote input control for quadcopters. In *The*

31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings, pages 120–122. ACM, 2018.

[223] Thomas Kosch, Mariam Hassib, Daniel Buschek, and Albrecht Schmidt. Look into My Eyes: Using Pupil Dilation to Estimate Mental Workload for Task Complexity Adaptation. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, pages LBW617:1–LBW617:6, New York, NY, USA, 2018. ACM.

[224] Thomas Kosch, Mariam Hassib, Robin Reutter, and Florian Alt. Emotions on the go: Mobile emotion assessment in real-time using facial expressions. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '20, New York, NY, USA, 2020. ACM.

[225] Thomas Kosch, Mariam Hassib, and Albrecht Schmidt. The Brain Matters: A 3D Real-Time Visualization to Examine Brain Source Activation Leveraging Neurofeedback. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '16, pages 1570–1576, New York, NY, USA, 2016. ACM.

[226] Thomas Kosch, Mariam Hassib, Paweł W. Woźniak, Daniel Buschek, and Florian Alt. Your Eyes Tell: Leveraging Smooth Pursuit for Assessing Cognitive Workload. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 436:1–436:13, New York, NY, USA, 2018. ACM.

[227] Thomas Kosch, Jakob Karolus, Havy Ha, and Albrecht Schmidt. Your skin resists: Exploring electrodermal activity as workload indicator during manual assembly. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, EICS '19, pages 8:1–8:5, New York, NY, USA, 2019. ACM.

[228] Thomas Kosch, Romina Kettner, Markus Funk, and Albrecht Schmidt. Comparing Tactile, Auditory, and Visual Assembly Error-Feedback for Workers with Cognitive Impairments. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '16, pages 53–60, New York, NY, USA, 2016. ACM.

[229] Thomas Kosch, Pascal Knierim, Pawel Wozniak, and Albrecht Schmidt. Chances and challenges of using assistive systems in education. *Mensch und Computer 2017–Proceedings*, 2017.

[230] Thomas Kosch, Albrecht Schmidt, Simon Thanheiser, and Lewis L. Chuang. One does not Simply RSVP: Mental Workload to Select Speed Reading Parameters using Electroencephalography. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, New York, NY, USA, 2020. ACM.

[231] Thomas Kosch, Kevin Wennrich, Daniel Topp, Marcel Muntzinger, and Albrecht Schmidt. The Digital Cooking Coach: Using Visual and Auditory In-situ Instructions to Assist Cognitively Impaired During Cooking. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '19, pages 156–163, New York, NY, USA, 2019. ACM.

[232] Thomas Kosch and Paweł W. Woźniak. Keep assembling and carry on: A satirical view on solving the workforce problem through cognitively impaired labor. 2019.

[233] Thomas Kosch, Paweł W. Woźniak, Erin Brady, and Albrecht Schmidt. Smart Kitchens for People with Cognitive Impairments: A Qualitative Study of Design Requirements. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI'18, pages 271:1–271:12, New York, NY, USA, 2018. ACM.

[234] Judith F. Kroll. Comprehension and memory in rapid sequential reading. *Attention and performance VIII*, 8:395, 1980.

[235] Jan-Louis Kruger, Esté Hefer, and Gordon Matthew. Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts. In *Proceedings of the 2013 Conference on Eye Tracking South Africa*, pages 62–66. ACM, 2013.

[236] Marc Langheinrich. Privacy by design — principles of privacy-aware ubiquitous systems. In Gregory D. Abowd, Barry Brumitt, and Steven Shafer, editors, *Ubicomp 2001: Ubiquitous Computing*, pages 273–291, Berlin, Heidelberg, 2001. Springer Berlin Heidelberg.

[237] III Lawrence J. Prinzel, Pope Alan T., Freeman Frederick G., Scerbo Mark W., and Mikulka Peter J. Empirical analysis of eeg and erps for pyschophysiological adaptive task allocation. Technical report, 2001.

[238] Johnny Chung Lee and Desney S. Tan. Using a low-cost electroencephalograph for task classification in hci research. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology*, UIST '06, pages 81–90, New York, NY, USA, 2006. ACM.

[239] Sang Hoon Lee, Doug Hyun Han, Seman Oh, In Kyoon Lyoo, Young Sik Lee, Perry F. Renshaw, and Scott E Lukas. Quantitative electroencephalographic (qeeg) correlates of craving during virtual reality therapy in alcohol-dependent patients. *Pharmacology Biochemistry and Behavior*, 91(3):393–397, 2009.

[240] Stephanie Lees, Natalie Dayan, Hubert Cecotti, Paul McCullagh, Liam Maguire, Fabien Lotte, and Damien Coyle. A review of rapid serial visual presentation-based brain–computer interfaces. *Journal of Neural Engineering*, 15(2):021001, jan 2018.

[241] R. John Leigh and David S. Zee. *The neurology of eye movements*, volume 90. Oxford University Press, USA, 2015.

[242] Alan M. Lesgold and Charles A. Perfetti. Interactive processes in reading comprehension. *Discourse Processes*, 1(4):323–336, 1978.

[243] E. C. Leuthardt, K. J. Miller, G. Schalk, R. P. N. Rao, and J. G. Ojemann. Electrocorticography-based brain computer interface-the seattle experience. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14(2):194–198, June 2006.

[244] Simon P Levine, Heidi M Horstmann, and Ned L Kirsch. people with cognitive impairment in accessing assistive technologies. *J Head Trauma Rehabil*, 7(3):46–58, 1992.

[245] Li Jiang, Da-you Liu, and Bo Yang. Smart home research. *Proceedings of 2004 International Conference on Machine Learning and Cybernetics (IEEE Cat. No.04EX826)*, 2(August):659–663, 2004.

[246] Calvin Liang, Jakob Karolus, Thomas Kosch, and Albrecht Schmidt. On the suitability of real-time assessment of programming proficiency using gaze properties. In *Proceedings of the 7th ACM International Symposium on Pervasive Displays*, PerDis '18, pages 31:1–31:2, New York, NY, USA, 2018. ACM.

[247] Yulan Liang and John D. Lee. Driver cognitive distraction detection using eye movements. In *Passive Eye Monitoring*, pages 285–300. Springer, 2008.

[248] Dachuan Liu, Bo Dong, Xing Gao, and Haining Wang. *Exploiting Eye Tracking for Smartphone Authentication*, pages 457–477. Springer International Publishing, Cham, 2015.

[249] Markus Löchtefeld, Sven Gehring, Johannes Schöning, and Antonio Krüger. Shelftorchlight: Augmenting a shelf using a camera projector unit. In *Conference on Pervasive Computing*, volume 10. Citeseer, 2010.

[250] Dillon James Lohr and Oleg V. Komogortsev. A comparison of smooth pursuit- and dwell-based selection at multiple levels of spatial accuracy. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, pages 2760–2766, New York, NY, USA, 2017. ACM.

[251] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger. A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, 15(3):031005, 2018.

[252] F. Lotte, M. Congedo, A. Lecuyer, F. Lamarche, and B. Arnaldi. A review of classification algorithms for eeg-based brain-computer interfaces. *Journal of Neural Engineering*, 4(2):R1, 2007.

[253] Fabien Lotte. *A Tutorial on EEG Signal-processing Techniques for Mental-state Recognition in Brain–Computer Interfaces*, pages 133–161. Springer London, London, 2014.

[254] Ameersing Luximon and Ravindra S Goonetilleke. Continuous subjective workload assessment technique. *Ergonomics for Global Quality and Productivity*, 1:1–4, 1998.

[255] Manon Maheux, Étienne Bisaillon-Sicotte, Shirin Tabrizi, Jorge L. Armony, Jean-Marc Lina, and Pierre Jolicoeur. Optimal measurements of hemodynamic response latency in fnirs using the jackknife approach. *Psychophysiology*, 54(1):62–73, 2017.

[256] Päivi Majaranta and Kari-Jouko Räihä. Twenty years of eye typing: systems and design issues. In *Proceedings of the 2002 symposium on Eye tracking research & applications*, pages 15–22. ACM, 2002.

[257] Plonsey Malmivuo, Jaakko Malmivuo, and Robert Plonsey. *Bioelectromagnetism: Principles and Applications of Bioelectric and Biomagnetic Fields*. Oxford University Press, 1995. Google-Books-ID: H9CFM0TqWwsC.

[258] Valeria Manera, Pierre-David Petit, Alexandre Derreumaux, Ivan Orvieto, Matteo Romagnoli, Graham Lyttle, Renaud David, and Philippe H. Robert. 'kitchen and cooking,' a serious game for mild cognitive impairment and alzheimer's disease: a pilot study. *Frontiers in Aging Neuroscience*, 7:24, 2015.

[259] Dante Mantini, Mauro G. Perrucci, Cosimo Del Gratta, Gian L. Romani, and Maurizio Corbetta. Electrophysiological signatures of resting state networks in the human brain. *Proceedings of the National Academy of Sciences*, 104(32):13170–13175, 2007.

[260] Esther Marx, Angela Deutschländer, Thomas Stephan, Marianne Dieterich, Martin Wiesmann, and Thomas Brandt. Eyes open and eyes closed as rest conditions: impact on brain activation patterns. *Neuroimage*, 21(4):1818–1824, 2004.

[261] S. G. Mason and G. E. Birch. A general framework for brain-computer interface design. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11(1):70–85, March 2003.

[262] Stefan Mattes and Anders Hallén. Surrogate distraction measurement techniques: The lane change test. *Driver distraction: Theory, effects, and mitigation*, pages 107–121, 2009.

[263] R. Matthews, P. J. Turner, N. J. McDonald, K. Ermolaev, T. M. Manus, R. A. Shelby, and M. Steindorf. Real time workload classification from an ambulatory wireless eeg system using hybrid eeg electrodes. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5871–5875, Aug 2008.

[264] G. I. McCalla, J. E. Greer, VS Kumar, P. Meagher, J. A. Collins, R. Tkatch, and B. Parkinson. A peer help system for workplace training. *B. d. Boulay, & R. Mizoguchi (Eds.), AI-ED*, 97(8):183–190, 1997.

[265] Colin McGinn. Can we solve the mind–body problem? *Mind*, 98(391):349–366, 1989.

[266] Bruce Mehler, Bryan Reimer, and J. A. Dusek. Mit agelab delayed digit recall task (n-back). *Cambridge, MA: Massachusetts Institute of Technology*, 2011.

[267] Alex Mihailidis, Jennifer N. Boger, Tammy Craig, and Jesse Hoey. The coach prompting system to assist older adults with dementia through handwashing: An efficacy study. *BMC Geriatrics*, 8(1):28, Nov 2008.

[268] Linda Miksch and Charlotte Schulz. Disconnect to reconnect: The phenomenon of digital detox as a reaction to technology overload, 2018. Student Paper.

[269] Franklin Miller and Alan Wertheimer. *The ethics of consent: theory and practice*. Oxford University Press, 2010.

[270] George A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.

[271] Matthew W. Miller, Jeremy C. Rietschel, Craig G. McDonald, and Bradley D. Hatfield. A novel approach to the physiological measurement of mental workload. *International Journal of Psychophysiology*, 80(1):75–78, 2011.

[272] Kenzaburo Miyawaki, Mutsuo Sano, Syunichi Yonemura, and Mihoko Matsuoka. A cooking support system for people with higher brain dysfunction. In *Proceedings of the ACM Multimedia 2009 Workshop on Multimedia for Cooking and Eating Activities*, CEA '09, pages 47–52, New York, NY, USA, 2009. ACM.

[273] Christian Mühl, Camille Jeunet, and Fabien Lotte. Eeg-based workload estimation across affective contexts. *Frontiers in neuroscience*, 8, 2014.

[274] Tim R. Mullen, Christian A. E. Kothe, Yu Mike Chi, Alejandro Ojeda, Trevor Kerth, Scott Makeig, Tzyy-Ping Jung, and Gert Cauwenberghs. Real-time neuroimaging and cognitive monitoring using wearable dry eeg. *IEEE Transactions on Biomedical Engineering*, 62(11):2553–2567, 2015.

[275] Klaus-Robert Müller, Michael Tangermann, Guido Dornhege, Matthias Krauledat, Gabriel Curio, and Benjamin Blankertz. Machine learning for real-time single-trial eeg-analysis: From brain-computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, 167(1):82 – 90, 2008. Brain-Computer Interfaces (BCIs).

[276] Bjørn Kåre Myskja, Rune Nydal, and Anne Ingeborg Myhr. We have never been elsi researchers–there is no need for a post-elsi shift. *Life Sciences, Society and Policy*, 10(1):9, 2014.

[277] Noman Naseer and Keum-Shik Hong. fnirs-based brain-computer interfaces: a review. *Frontiers in Human Neuroscience*, 9:3, 2015.

[278] Luis Fernando Nicolas-Alonso and Jaime Gomez-Gil. Brain Computer Interfaces, a Review. *Sensors*, 12(2):1211–1279, February 2012.

[279] Vadim V. Nikulin, Guido Nolte, and Gabriel Curio. A novel method for reliable and fast extraction of neuronal eeg/meg oscillations on the basis of spatio-spectral decomposition. *NeuroImage*, 55(4):1528 – 1535, 2011.

[280] Helen Nissenbaum. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009.

[281] Adam Nowak, Pascal Knierim, Andrzej Romanowski, Albrecht Schmidt, and Thomas Kosch. What does the oscilloscope say?: Comparing the efficiency of in-situ visualisations during circuit analysis. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI EA '20, New York, NY, USA, 2020. ACM.

[282] Paul L. Nunez and Ramesh Srinivasan. *Electric fields of the brain: the neurophysics of EEG*. Oxford University Press, USA, 2006.

[283] N. Y. L. Oei, W. T. A. M. Everaerd, B. M. Elzinga, S. van Well, and B. Bermond. Psychosocial stress impairs working memory at high loads: An association with cortisol levels and memory retrieval. *Stress*, 9(3):133–141, 2006. PMID: 17035163.

[284] Takehiko Ohno. Features of eye gaze interface for selection tasks. In *Computer Human Interaction, 1998. Proceedings. 3rd Asia Pacific*, pages 176–181. IEEE, 1998.

[285] Inês Oliveira, Ovidiu Grigore, Nuno M. Guimarães, and Carlos Duarte. Experiences in reading detection with eeg signals. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, SAC '10, pages 1236–1237, New York, NY, USA, 2010. ACM.

[286] Inês Oliveira, Ovidiu Grigore, and Nuno Guimarães. Reading detection based on electroencephalogram processing. In *WSEAS International Conference. Proceedings. Recent Advances in Computer Engineering*, number 13. WSEAS, 2009.

[287] Inês Oliveira and Nuno Guimarães. A tool for mental workload evaluation and adaptation. In *Proceedings of the 4th Augmented Human International Conference*, AH '13, pages 138–141, New York, NY, USA, 2013. ACM.

[288] Patrick Olivier, Guangyou Xu, Andrew Monk, and Jesse Hoey. Ambient kitchen: Designing situated services using a high fidelity prototyping environment. In *Proceedings of the 2Nd International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '09, pages 47:1–47:7, New York, NY, USA, 2009. ACM.

[289] Gustav Öquist and Kristin Lundin. Eye movement study of reading text on a mobile phone using paging, scrolling, leading, and rsvp. In *Proceedings of the 6th International Conference on Mobile and Ubiquitous Multimedia*, MUM '07, pages 176–183, New York, NY, USA, 2007. ACM.

[290] World Health Organization et al. *World report on disability.* World Health Organization, 2011.

[291] Adrian M. Owen, Kathryn M. McMillan, Angela R. Laird, and Ed Bullmore. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, 25(1):46–59, 2005.

[292] R. Parasuraman and J. Beatty. Brain events underlying detection and recognition of weak sensory signals. *Science*, 210(4465):80–83, 1980.

[293] Hae-Jeong Park and Karl Friston. Structural and functional brain networks: From connections to cognition. *Science*, 342(6158), 2013.

[294] Roberto D. Pascual-Marqui, Christoph M. Michel, and Dietrich Lehmann. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of psychophysiology*, 18(1):49–65, 1994.

[295] Roberto Domingo Pascual-Marqui et al. Standardized low-resolution brain electromagnetic tomography (sloreta): technical details. *Methods Find Exp Clin Pharmacol*, 24(Suppl D):5–12, 2002.

[296] A. Pauzié. A method to assess the driver mental workload: The driving activity load index (dali). *IET Intelligent Transport Systems*, 2:315–322(7), December 2008.

[297] Annie Pauzié. A method to assess the driver mental workload: The driving activity load index (dali). *IET Intelligent Transport Systems*, 2(4):315–322, 2008.

[298] Peter Peltonen, Esko Kurvinen, Antti Salovaara, Giulio Jacucci, Tommi Ilmonen, John Evans, Antti Oulasvirta, and Petri Saarikko. It's mine, don't touch!: Interactions at a large multi-touch display in a city centre. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1285–1294, New York, NY, USA, 2008. ACM.

[299] Robert Perneczky, Corina Pohl, Christian Sorg, Julia Hartmann, Katja Komossa, Panagiotis Alexopoulos, Stefan Wagenpfeil, and Alexander Kurz. Complex activities of daily living in mild cognitive impairment: conceptual and diagnostic issues. *Age and ageing*, 35(3):240–245, 2006.

[300] Anat Perry, Nikolaus F. Troje, and Shlomo Bentin. Exploring motor system contributions to the perception of social information: Evidence from eeg activity in the mu/alpha frequency range. *Social Neuroscience*, 5(3):272–284, 2010. PMID: 20169504.

[301] Bernd Petzold, Michael F. Zaeh, Berthold Faerber, Barbara Deml, Hans Egermeier, Johannes Schilp, and Stella Clarke. A study on visual, auditory, and haptic feedback for assembly tasks. *Presence: teleoperators and virtual environments*, 13(1):16–21, 2004.

[302] Vsevolod Peysakhovich. *Study of pupil diameter and eye movements to enhance flight safety. Etude de diamètre pupillaire et de mouvements oculaires pour la sécurité aérienne.* PhD thesis, Université de Toulouse, Université Toulouse III-Paul Sabatier, 2016.

[303] Ken Pfeuffer, Melodie Vidal, Jayson Turner, Andreas Bulling, and Hans Gellersen. Pursuit calibration: Making gaze calibration less tedious and more flexible. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 261–270. ACM, 2013.

[304] Bastian Pfleging, Drea K. Fekety, Albrecht Schmidt, and Andrew L. Kun. A model relating pupil diameter to mental workload and lighting conditions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5776–5788. ACM, 2016.

[305] Gert Pfurtscheller and F. H. Lopes Da Silva. Event-related eeg/meg synchronization and desynchronization: basic principles. *Clinical neurophysiology*, 110(11):1842–1857, 1999.

[306] Claudio Pinhanez. The everywhere displays projector: A device to create ubiquitous graphical interfaces. In *Ubicomp 2001: Ubiquitous Computing*, pages 315–331. Springer, 2001.

[307] Martha E. Pollack. Intelligent Technology for an Aging Population: The Use of AI to Assist Elders with Cognitive Impairment. *AI Magazine*, 26(2):9, 2005.

[308] Alex Poole and Linden J. Ball. Eye tracking in hci and usability research. In *Encyclopedia of human computer interaction*, pages 211–219. IGI Global, 2006.

[309] Molly Potter. Rapid serial visual presentation (rsvp): A method for studying language processing. *New methods in reading comprehension research*, 1984.

[310] John G. Proakis. *Digital signal processing: principles algorithms and applications.* Pearson Education India, 2001.

[311] Dale Purves, George J. Augustine, David Fitzpatrick, Lawrence C. Katz, Anthony-Samuel LaMantia, James O. McNamara, and S. Mark Williams. Types of eye movements and their functions. 2001.

[312] Elizabeth Quinn and Ian Stephen Paul Nation. *Speed reading: A course for learners of English.* Oxford University Press, 1974.

[313] U. Rajendra Acharya, K. Paul Joseph, N. Kannathal, Choo Min Lim, and Jasjit S. Suri. Heart rate variability: a review. *Medical and Biological Engineering and Computing*, 44(12):1031–1051, Dec 2006.

[314] Matthias Rauterberg and Erich Styger. Positive effects of sound feedback during the operation of a plant simulator. In *Human-Computer Interaction*, pages 35–44. Springer, 1994.

[315] Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372, 1998.

[316] Paul Richard, Grigore Burdea, Daniel Gomez, and Philippe Coiffet. A comparison of haptic, visual and auditive force feedback for deformable virtual objects. In *Proceedings of the Internation Conference on Automation Technology (ICAT)*, pages 49–62, 1994.

[317] Vincenzo Romei, Tonia Rihs, Verena Brodbeck, and Gregor Thut. Resting electroencephalogram alpha-power over posterior sites indexes baseline visual cortex excitability. *Neuroreport*, 19(2):203–208, 2008.

[318] F. David Rose, Barbara M. Brooks, and Albert A. Rizzo. Virtual reality in brain damage rehabilitation: review. *CyberPsychology & Behavior*, 8(3):241–262, 2005.

[319] Beate Rossler. *The value of privacy*. John Wiley & Sons, 2018.

[320] R. N. Roy, S. Charbonnier, A. Campagne, and S. Bonnet. Efficient mental workload estimation using task-independent eeg features. *Journal of Neural Engineering*, 13(2):026019, 2016.

[321] Gary S. Rubin and Kathleen Turano. Reading without saccadic eye movements. *Vision research*, 32(5):895–902, 1992.

[322] Mark C. Russell and Barbara S. Chaparro. Exploring effects of speed and font size with rsvp. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 45(6):640–644, 2001.

[323] Stefan Rüther, Thomas Hermann, Maik Mracek, Stefan Kopp, and Jochen Steil. An assistance system for guiding workers in central sterilization supply departments. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*, page 3. ACM, 2013.

[324] Rufat Rzayev, Paweł W. Woźniak, Tilman Dingler, and Niels Henze. Reading on smart glasses: The effect of text position, presentation type and walking. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 45:1–45:9, New York, NY, USA, 2018. ACM.

[325] T. L. Saaty and M. S. Ozdemir. Why the magic number seven plus or minus two. *Mathematical and Computer Modelling*, 38(3):233 – 244, 2003.

[326] Dario D. Salvucci and Joseph H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ETRA '00, pages 71–78, New York, NY, USA, 2000. ACM.

[327] Dario D. Salvucci and Niels A. Taatgen. *The multitasking mind*. Oxford University Press, 2010.

[328] Oliver Sand, Sebastian Büttner, Volker Paelke, and Carsten Röcker. *smARt.Assembly – Projection-Based Augmented Reality for Supporting Assembly Workers*, pages 643–652. Springer International Publishing, Cham, 2016.

[329] Angela L. Sauer, Andra Parks, and Patricia C. Heyn. Assistive technology effects on the employment outcomes for people with cognitive disabilities: a systematic review. *Disability and Rehabilitation: Assistive Technology*, 5(6):377–391, 2010.

[330] Christian Scharinger, Alexander Soutschek, Torsten Schubert, and Peter Gerjets. Comparison of the working memory load in n-back and working memory span tasks by means of eeg frequency band power and p300 amplitude. *Frontiers in human neuroscience*, 11, 2017.

[331] Florian Schaule, Jan Ole Johanssen, Bernd Bruegge, and Vivian Loftness. Employing consumer wearables to detect office workers' cognitive load for interruption management. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(1):32:1–32:20, March 2018.

[332] Menja Scheer, Heinrich H Bülthoff, and Lewis L Chuang. Steering demands diminish the early-p3, late-p3 and ron components of the event-related potential of task-irrelevant environmental sounds. *Frontiers in human neuroscience*, 10, 2016.

[333] René Scheeringa, Marcel CM Bastiaansen, Karl Magnus Petersson, Robert Oostenveld, David G. Norris, and Peter Hagoort. Frontal theta eeg activity correlates negatively with the default mode network in resting state. *International journal of psychophysiology*, 67(3):242–251, 2008.

[334] Jürgen Scheible, Arnd Engeln, Michael Burmester, Gottfried Zimmermann, Tobias Keber, Uwe Schulz, Sabine Palm, Markus Funk, and Uwe Schaumann. Smartkitchen media enhanced cooking environment. In *Proceedings of the 6th International Conference on the Internet of Things*, IoT'16, pages 169–170, New York, NY, USA, 2016. ACM.

[335] Carlos H. Schenck, Scott R. Bundlie, Andrea L. Patterson, and Mark W. Mahowald. Rapid eye movement sleep behavior disorder: a treatable parasomnia affecting older adults. *Jama*, 257(13):1786–1789, 1987.

[336] Simon Schenk, Marc Dreiser, Gerhard Rigoll, and Michael Dorr. Gazeeverywhere: Enabling gaze-only user interaction on an unmodified desktop pc in everyday scenarios. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 3034–3044, New York, NY, USA, 2017. ACM.

[337] Marcia J. Scherer. Outcomes of assistive technology use on quality of life. *Disability and Rehabilitation*, 18(9):439–448, 1996.

[338] Marcia J Scherer, Tessa Hart, Ned Kirsch, and Maria Schulthesis. Assistive technologies for cognitive disabilities. *Critical Reviews™ in Physical and Rehabilitation Medicine*, 17(3), 2005.

[339] Christina Schneegass, Thomas Kosch, Andrea Baumann, Marius Rusu, Mariam Hassib, and Heinrich Hussmann. Braincode: Electroencephalography-basedcomprehension detection during reading and listening. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, New York, NY, USA, 2020. ACM.

[340] Christina Schneegass, Thomas Kosch, Albrecht Schmidt, and Heinrich Hussmann. Investigating the potential of eeg for implicit detection of unknown words for foreign language learning. In David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris, editors, *Human-Computer Interaction – INTERACT 2019*, pages 293–313, Cham, 2019. Springer International Publishing.

[341] M. Schneider. The semantic cookbook: sharing cooking experiences in the smart kitchen. *3rd IET International Conference on Intelligent Environments IE 07*, pages 416–423, 2007.

[342] Daniela Schoofs, Diana Preuß, and Oliver T. Wolf. Psychosocial stress induces working memory impairments in an n-back paradigm. *Psychoneuroendocrinology*, 33(5):643 – 653, 2008.

[343] Andrew Sears and Vicki L. Hanson. Representing users in accessibility research. *ACM Transactions on Accessible Computing (TACCESS)*, 4(2):7, 2012.

[344] Charlott Sellberg and Tarja Susi. Technostress in the office: a distributed cognition perspective on human–technology interaction. *Cognition, Technology & Work*, 16(2):187–201, May 2014.

[345] Linda E. Sibert and Robert J. K. Jacob. Evaluation of eye gaze interaction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 281–288. ACM, 2000.

[346] Eva Siegenthaler, Francisco M. Costela, Michael B. McCamy, Leandro L. Di Stasi, Jorge Otero-Millan, Andreas Sonderegger, Rudolf Groner, Stephen Macknik, and Susana Martinez-Conde. Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *European Journal of Neuroscience*, 39(2):287–294, 2014.

[347] David L. Sparks. The brainstem control of saccadic eye movements. *Nature Reviews Neuroscience*, 3(12):952–964, 2002.

[348] Julian Steil, Marion Koelle, Wilko Heuten, Susanne Boll, and Andreas Bulling. Privaceye: Privacy-preserving head-mounted eye tracking using egocentric scene image and eye movement features. In *Proceedings of the 11th ACM Symposium on Eye*

Tracking Research & Applications, ETRA '19, pages 26:1–26:10, New York, NY, USA, 2019. ACM.

[349] Emmanuel Stip and Vincent Rialle. Environmental cognitive remediation in schizophrenia: Ethical implications of "smart home" technology. *The Canadian Journal of Psychiatry*, 50(5):281–291, 2005.

[350] A. Stipacek, R.H. Grabner, C. Neuper, A. Fink, and A.C. Neubauer. Sensitivity of human eeg alpha band desynchronization to different working memory components and increasing levels of memory load. *Neuroscience Letters*, 353(3):193 – 196, 2003.

[351] Els Stuyven, Koen Van der Goten, André Vandierendonck, Kristl Claeys, and Luc Crevits. The effect of cognitive load on saccadic eye movements. *Acta psychologica*, 104(1):69–85, 2000.

[352] H. Lee Swanson and Rollanda O'Connor. The role of working memory and fluency practice on the reading comprehension of students who are dysfluent readers. *Journal of Learning Disabilities*, 42(6):548–575, 2009.

[353] John Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2):257–285, 1988.

[354] John Sweller. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4):295–312, 1994.

[355] John Sweller, Jeroen J. G. Van Merrienboer, and Fred G. W. C. Paas. Cognitive architecture and instructional design. *Educational psychology review*, 10(3):251–296, 1998.

[356] Kazuo Tanaka, Kazuyuki Matsunaga, and Hua O. Wang. Electroencephalogram-based control of an electric wheelchair. *Robotics, IEEE Transactions on*, 21(4):762–766, 2005.

[357] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 73–80. ACM, 2003.

[358] Z. Tang, J. Guo, S. Miao, S. Acharya, and J. H. Feng. Ambient intelligence based context-aware assistive system to improve independence for people with autism spectrum disorder. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 3339–3348, 2016.

[359] Sibel Tekin and Jeffrey L. Cummings. Frontal–subcortical neuronal circuits and clinical neuropsychiatry: an update. *Journal of psychosomatic research*, 53(2):647–654, 2002.

[360] Lucia Terrenghi, Otmar Hilliges, and Andreas Butz. Kitchen stories: Sharing recipes with the Living Cookbook. *Personal and Ubiquitous Computing*, 11(5):409–414, 2007.

[361] Zoran Tiganj, Mamadou Mboup, Christophe Pouzat, and Lotfi Belkoura. An algebraic method for eye blink artifacts detection in single channel eeg recordings. In Selma Supek and Ana Sušac, editors, *17th International Conference on Biomagnetism Advances in Biomagnetism – Biomag2010*, pages 175–178, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg.

[362] Angela K. Troyer and Kelly J. Murphy. Memory for intentions in amnestic mild cognitive impairment: Time-and event-based prospective memory. *Journal of the International Neuropsychological Society*, 13(2):365–369, 2007.

[363] Yi-Fang Tsai, Erik Viirre, Christopher Strychacz, Bradley Chase, and Tzyy-Ping Jung. Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, space, and environmental medicine*, 78(Supplement 1):B176–B185, 2007.

[364] Ioannis Tsochantaridis, Thomas Hofmann, Thorsten Joachims, and Yasemin Altun. Support vector machine learning for interdependent and structured output spaces. In *Proceedings of the twenty-first international conference on Machine learning*, page 104. ACM, 2004.

[365] Outi Tuisku, Päivi Majaranta, Poika Isokoski, and Kari-Jouko Räihä. Now dasher! dash away!: longitudinal study of fast text entry by eye gaze. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, pages 19–26. ACM, 2008.

[366] Marilyn L. Turner and Randall W. Engle. Is working memory capacity task dependent? *Journal of memory and language*, 28(2):127–154, 1989.

[367] Anirudh Vallabhaneni, Tao Wang, and Bin He. *Brain—Computer Interface*, pages 85–121. Springer US, Boston, MA, 2005.

[368] Jeroen Van Den Hoven. Information technology, privacy, and the protection of personal data. *Information technology and moral philosophy*, 301, 2008.

[369] Hidde van der Meulen, Andrew L. Kun, and Orit Shaer. What are we missing?: Adding eye-tracking to the hololens to improve gaze estimation accuracy. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, ISS '17, pages 396–400, New York, NY, USA, 2017. ACM.

[370] Michael Veale and Reuben Binns. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2):2053951717743530, 2017.

[371] Boris M. Velichkovsky and John Paulin Hansen. New technological windows into mind: There is more in eyes and brains for human-computer interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '96, pages 496–503, New York, NY, USA, 1996. ACM.

[372] Eduardo Velloso, Markus Wirth, Christian Weichel, Augusto Esteves, and Hans Gellersen. Ambigaze: Direct control of ambient devices by gaze. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, pages 812–817. ACM, 2016.

[373] Trent W. Victor, Joanne L. Harbluk, and Johan A. Engström. Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2):167 – 190, 2005. The relationship between distraction and driving performance: towards a test regime for in-vehicle information systemsIn-vehicle information systems.

[374] Mélodie Vidal, Andreas Bulling, and Hans Gellersen. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 439–448. ACM, 2013.

[375] Arno Villringer and Britton Chance. Non-invasive optical spectroscopy and imaging of human brain function. *Trends in Neurosciences*, 20(10):435 – 442, 1997.

[376] Daniel Vogel and Ravin Balakrishnan. Interactive public ambient displays: Transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, pages 137–146, New York, NY, USA, 2004. ACM.

[377] Alexandra Voit, Tonja Machulla, Dominik Weber, Valentin Schwind, Stefan Schneegass, and Niels Henze. Exploring notifications in smart home environments. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '16*, pages 942–947, 2016.

[378] Maria Vukovic, Vidhyasaharan Sethu, Jessica Parker, Lawrence Cavedon, Margaret Lech, and John Thangarajah. Estimating cognitive load from speech gathered in a complex real-life training exercise. *International Journal of Human-Computer Studies*, 124:116 – 133, 2019.

[379] Jenny Waycott, Hilary Davis, Anja Thieme, Stacy Branham, John Vines, and Cosmin Munteanu. Ethical encounters in hci: Research in sensitive settings. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, pages 2369–2372, New York, NY, USA, 2015. ACM.

[380] Jenny Waycott, Cosmin Munteanu, Hilary Davis, Anja Thieme, Wendy Moncur, Roisin McNaney, John Vines, and Stacy Branham. Ethical encounters in human-computer interaction. In *Proceedings of the 2016 CHI Conference Extended Abstracts*

on Human Factors in Computing Systems, CHI EA '16, pages 3387–3394, New York, NY, USA, 2016. ACM.

[381] Mark Weiser. Ubiquitous computing. *Computer*, (10):71–72, 1993.

[382] Mark Weiser and John Seely Brown. *The Coming Age of Calm Technology*, pages 75–85. Springer New York, New York, NY, 1997.

[383] Pierre Wellner. The digitaldesk calculator: tangible manipulation on a desk top display. In *Proceedings of the 4th annual ACM symposium on User interface software and technology*, pages 27–33. ACM, 1991.

[384] John D. Wells, Damon E. Campbell, Joseph S. Valacich, and Mauricio Featherman. The effect of perceived novelty on the adoption of information technology innovations: a risk/reward perspective. *Decision Sciences*, 41(4):813–843, 2010.

[385] Ari Widyanti, Addie Johnson, and Dick de Waard. Adaptation of the rating scale mental effort (rsme) for use in indonesia. *International Journal of Industrial Ergonomics*, 43(1):70 – 76, 2013.

[386] Brenda Wiederhold and Giuseppe Riva. *Annual Review of CyberTherapy and Telemedicine: Positive Technology and Health Engagement for Healthy Living and Active Ageing*. 06 2013.

[387] Frederik Wiehr, Alexandra Voit, Dominik Weber, Sven Gehring, Christoph Witte, Daniel Kärcher, Niels Henze, and Antonio Krüger. Challenges in designing and implementing adaptive ambient notification environments. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp '16*, pages 1578–1583, 2016.

[388] Michele A. Williams, Caroline Galbraith, Shaun K. Kane, and Amy Hurst. Just let the cane hit it: how the blind and sighted see navigation differently. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*, pages 217–224. ACM, 2014.

[389] Ian H. Witten, Eibe Frank, Mark A. Hall, and Christopher J. Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[390] Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, and Theresa M. Vaughan. Brain–computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, June 2002.

[391] Jonathan R. Wolpaw, Dennis J. McFarland, Gregory W. Neat, and Catherine A. Forneris. An eeg-based brain-computer interface for cursor control. *Electroencephalography and clinical neurophysiology*, 78(3):252–259, 1991.

[392] Mikolaj Wozniak, Sebastian Feger, Matthias Hoppe, Thomas Kosch, Stephanie Van de Sandt, Jakob Karolus, Luke Haliburton, Pawel Wozniak, and Jasmin Niess. Madrone: Using drones to facilitate connectedness across geographic boundaries. In *2nd International Workshop on Human-Drone Interaction at the CHI Conference on Human Factors in Computing Systems*, 2020.

[393] Sheryl R. Young. Rsvp: A task, reading aid, and research tool. *Behavior Research Methods, Instruments, & Computers*, 16(2):121–124, Mar 1984.

[394] Kun Yu, Julien Epps, and Fang Chen. Cognitive load evaluation of handwriting using stroke-level features. In *Proceedings of the 16th International Conference on Intelligent User Interfaces*, IUI '11, pages 423–426, New York, NY, USA, 2011. ACM.

[395] Yueran Yuan, Kai-min Chang, Jessica Nelson Taylor, and Jack Mostow. Toward unobtrusive measurement of reading comprehension using low-cost eeg. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, LAK '14, pages 54–58, New York, NY, USA, 2014. ACM.

[396] Johannes Zagermann, Ulrike Pfeil, and Harald Reiterer. Measuring cognitive load using eye tracking technology in visual computing. In *Proceedings of the Sixth Workshop on Beyond Time and Errors on Novel Evaluation Methods for Visualization*, BELIV '16, page 78–85, New York, NY, USA, 2016. Association for Computing Machinery.

[397] Thorsten O. Zander, Lena M. Andreessen, Angela Berg, Maurice Bleuel, Juliane Pawlitzki, Lars Zawallich, Laurens R. Krol, and Klaus Gramann. Evaluation of a dry eeg system for application of passive brain-computer interfaces in autonomous driving. *Frontiers in Human Neuroscience*, 11:78, 2017.

[398] N Jane Zbrodoff and Gordon D Logan. What everyone finds: The problem-size effect. 2005.

[399] Yanxia Zhang, Andreas Bulling, and Hans Gellersen. Sideways: a gaze interface for spontaneous interaction with situated displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 851–860. ACM, 2013.

[400] Josef Zihl, D. Von Cramon, and Norbert Mai. Selective disturbance of movement vision after bilateral brain damage. *Brain*, 106(2):313–340, 1983.

[401] F. R. H. Zijlstra and L. Van Doorn. *The construction of a scale to measure perceived effort*. University of Technology, 1985.

[402] Ferdinand Rudolf Hendrikus Zijlstra. Efficiency in work behaviour: A design approach for modern tools. 1993.

[403] Andrej Zwitter. Big data ethics. *Big Data & Society*, 1(2):2053951714559253, 2014.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig und ohne unerlaubte Beihilfe angefertigt wurde.

München, den 28.01.2020

Thomas-Andreas Kosch