

Chair for Computer Science 10 (Media Computing and Human-Computer Interaction)

RWTH AACHEN UNIVERSITY

Investigating the Robustness of EEG-Frequency Band Measurements to Movement-Induced Noise

Bachelor's Thesis at the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Wiebke Merle Neumeuer

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Ulrik Schroeder

Registration date: 10.06.2026
Submission date: 30.09.2026

Zentrales Prüfungsamt/Central Examination Office

RWTH AACHEN UNIVERSITY

Eidesstattliche Versicherung
Declaration of Academic Integrity

Neumeyer, Wiebke Mette

Name, Vorname/Last Name, First Name

Matrikelnummer (freiwillige Angabe)
Student ID Number (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/Masterarbeit* mit dem Titel
I hereby declare under penalty of perjury that I have completed the present paper/bachelor's thesis/master's thesis* entitled

Investigating the Robustness of EEG-Frequency Band Measurements to Movement-Induced Noise

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt; dies umfasst insbesondere auch Software und Dienste zur Sprach-, Text- und Medienproduktion. Ich erkläre, dass für den Fall, dass die Arbeit in unterschiedlichen Formen eingereicht wird (z.B. elektronisch, gedruckt, geplottet, auf einem Datenträger) alle eingereichten Versionen vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.
independently and without unauthorized assistance from third parties (in particular academic ghostwriting**.** I have not used any other sources or aids than those indicated; this includes in particular software and services for language, text, and media production. In the event that the work is submitted in different formats (e.g. electronically, printed, plotted, on a data carrier), I declare that all the submitted versions are fully identical. I have not previously submitted this work, either in the same or a similar form to an examination body.

Roisym, 29.09.26

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen/Please delete as appropriate

Belehrung:
Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt
Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.
§ 156 StGB (German Criminal Code): False Unsworn Declarations
Whosoever before a public authority competent to administer unsworn declarations (including Declarations of Academic Integrity) falsely submits such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment for a term not exceeding three years or to a fine.
§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt
(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.
(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.
§ 161 StGB (German Criminal Code): False Unsworn Declarations Due to Negligence
(1) If an individual commits one of the offenses listed in §§ 154 to 156 due to negligence, they are liable to imprisonment for a term not exceeding one year or to a fine.
(2) The offender shall be exempt from liability if they correct their false testimony in time. The provisions of § 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:
I have read and understood the above official notification:

Roisym, 29.09.26

Ort, Datum/City, Date

Unterschrift/Signature

Contents

Abstract ix

Überblick xi

Acknowledgments xiii

Conventions xv

1 Introduction 1

1.1 Outline 3

2 Related Work 5

2.1 Electroencephalography 5

2.1.1 Cognitive Load Index 8

2.1.2 Engagement Index 8

2.2 EEGs in HCI 9

2.3 Artifacts in EEG Data 10

2.4 EEG Preprocessing 11

2.4.1 Signal-to-Noise Ratio 11

3 Survey on Typical HCI User Study Settings 13

3.1 Research Question 13

3.2 Method 14

3.2.1 Survey Design 14

3.2.2 Pilot Survey 15

3.2.3 Participants 16

3.3 Survey Results 16

4 User Study 21

4.1 Research Question 21

4.2 Method 22

4.2.1 Recording Method 22

4.2.2 Final Study Process 23

4.2.3 Pilot Study 24

4.2.4 Task Design 25

Reading 27

Memory 28

ABC 29

Mental Math 30

4.2.5 Counterbalancing 30

4.2.6 Participants 32

4.3 Data Analysis 32
4.3.1 Evaluation Method 34
4.3.2 Results 35
Reading 35
Memory 37
ABC 38
Mental Math 38
SNR - Raw 39
4.3.3 Questionnaire 42

5 Discussion **47**
5.1 Findings 47
5.2 Implications 49
5.3 Limitations 49
5.3.1 Excursus - Alternative Data Analysis 50

6 Summary and Future Work **53**
6.1 Summary and Contributions 53
6.2 Future Work 54

A Survey Design **57**

B Study Questionnaire **67**

C Mental Math Tasks **77**

D Final Study Layout **79**

Bibliography **81**

Index **87**

List of Figures and Tables

2.1 Muse 10-20 Electrode Placement 6

2.2 First Muse from Interaxon 7

3.1 Typical Settings Survey Result 17

3.2 Typical Tasks Survey Result 17

3.3 Typical Motions Survey Result 18

4.1 Electrode Check Example Screen 23

4.2 Restriction Setup 26

4.3 Latin Squares 31

4.4 R Code Basis 33

4.5 Statistical Decision Tree 33

4.6 Boxplot Reading 36

4.7 Boxplot Memory 37

4.8 Boxplot ABC 38

4.9 Boxplot Mental Math 39

4.10 Collected Boxplots of Raw SNR 42

5.1 Manual SNR 51

5.2 Provided SNR 52

C.1 Math Tasks 77

D.1 Counterbalanced Conditions 79

Abstract

In recent years, Electroencephalography (EEG) has become an increasingly important field of research outside of medical applications. This also holds for the field of Human-Computer Interaction (HCI). EEGs can be used to actively control interfaces without conventional input methods, but also as an evaluation tool. They can give more factual insight into a person's state of mind. However, they are highly susceptible to movement which can impact the data quality.
While there are methods to preprocess data, these do not guarantee good data quality. With fewer numbers of electrodes, preprocessing is limited in its efficiency. This is why this thesis aimed to investigate the limits and possibilities with EEGs as an evaluation tool in HCI studies. A study using a commercial-grade EEG headset was conducted. The study included four different tasks and different movements that were consciously introduced and restricted to evaluate the robustness to these movements. To evaluate the data quality, the Signal-to-Noise Ratio was used.
It was found that the program used for evaluation computed an unreliable Signal-to-Noise Ratio leading to distorted results. Based on these computed values, however, it can be said that high intensity movements had a large impact on the data quality. For less intense movements, the impact and its significance varied.
For future work, it is recommended to reevaluate the calculation of the Signal-to-Noise Ratio and, in the course of this, adapt it to be more customizable in the definition of signal and noise.

Überblick

In den letzten Jahren hat sich die Nutzung von Elektroenzephalografie (EEG) als ein wichtiges Forschungsfeld auch außerhalb von medizinischen Nutzungen etabliert. Dies ist auch der Fall für den Bereich der Mensch-Computer-Interaktion (HCI). Hier werden EEGs unter anderem dafür verwendet, Interfaces zu kontrollieren, und ermöglichen damit den Verzicht auf konventionelle Eingabemethoden. Aber auch als Evaluierungsmethoden spielen sie eine wachsende Rolle. EEGs ermöglichen es, einen faktenbasierteren Einblick in kognitive Prozesse des Menschen zu erlangen. Jedoch sind EEG-Daten leicht verfälschbar durch Körperbewegungen. Zwar gibt es Methoden zur Reinigung der Daten, doch sind diese insbesondere bei kleineren EEGs mit weniger Elektroden limitiert. Deshalb ist das Ziel dieser Arbeit, die Grenzen und Möglichkeiten von einem kommerziellen EEG zu evaluieren. In der Studie wurden vier verschiedene Aufgaben unter vier verschiedenen Arten von Bewegungen untersucht. Zur Evaluierung wurde das Signal-Rausch-Verhältnis verwendet, welches eine übliche Evaluationsmethode der Datenqualität ist.
Im Zuge der Datenanalyse wurde herausgefunden, dass das benutzte Programm zur Evaluierung, eine fehlerhafte Berechnung nutzt und somit die Ergebnisse der Analyse verzerrt. Basierend auf der Analyse der erhobenen Daten, wurde festgestellt, dass intensivere Bewegungen eine starke Auswirkung auf die Daten hatten. Bei milderen Bewegungen variierten die Ergebnisse.
Für zukünftige Arbeiten wird empfohlen, die Berechnung des Signal-Rausch-Verhältnisses anzupassen und individualisierbarer zu gestalten.

Acknowledgments

Firstly, I would like to thank Prof. Dr. Jan Borchers for examining my thesis. I also thank Prof. Dr. Ulrik Schroeder for being my secondary advisor.

I thank my thesis advisor, Kristina Hörrmann, who always had helpful feedback and always helped me with all my questions. I really enjoyed working on this thesis with your help and am very grateful for the opportunity to write my thesis in this field.

I would also like to thank all people who participated in my study and made this thesis possible.

I also want to thank my parents, brother and friends who motivated me and cheered me on throughout the whole duration of my thesis.

Finally, I want to thank my cat for his continuous silent support.

Conventions

Throughout this thesis we use the following conventions:

- The thesis is written in American English.
- The first person is written in plural form.
- Unidentified third persons are described in third-person plural.

Where appropriate, paragraphs are summarized by one or two sentences that are positioned at the margin of the page.

This is a summary of a paragraph.

Chapter 1

Introduction

EEGs are applied in various research fields.

The origin of Electroencephalography (EEG) lies in neuroscience and medicine. EEGs enable the measuring of brain signals via electrodes placed on the scalp. Through this, they enable the recording of five distinct frequencies of the brain. These frequencies can be used for evaluation for numerous applications [Rao, 2013]. They helped deepen the understanding of our brain and advanced the research of brain diseases such as Alzheimer's disease [Arjoonsingh et al., 2024]. In recent years, there has been an increased interest in EEGs as tools outside purely medical applications as well. Especially the increased availability and affordability of EEGs with consumer-grade off-the-shelf products played a huge role in these advancements. Through them, opportunities to adopt EEGs in non-medical applications have increased [Jia, 2026]. Examples of consumer-grade headsets include the Muse[1], Emotiv[2], OpenBCI[3] and NeuroSky[4].

EEGs have a wider range of application in HCI.

These also play a role for the field of Human-Computer Interaction (HCI). In HCI, EEGs have found various forms for application. The most common domains for the usage of EEGs include interactive control and gaming, affective computing, workload and user experience evaluation, and there is an emerging usage of EEG in an extended reality

[1] https://choosemuse.com/, *last accessed September 30, 2026*
[2] https://www.emotiv.com/de/, *last accessed September 30, 2026*
[3] https://openbci.com/, *last accessed September 30, 2026*
[4] https://neurosky.com/, *last accessed September 30, 2026*

(XR) and virtual reality (VR) related context [Jia, 2026]. A popular and common application of EEGs as an input and control mechanism are Brain-Computer Interfaces (BCIs). BCI applications can range from entertainment, for example as input methods in games, to medical applications for people with motor or communication disabilities as such systems can replace conventional input hardware like keyboards [Ortega-Gijon and Mezura-Godoy, 2019].

EEGs have potential as evaluation tool in HCI for a more factual analysis of users.

EEGs also find application as evaluation tools in HCI. As of now, to evaluate a user's reaction to an interface, methods such as interviews or scales like the NASA Task Load Index (NASA-TLX) are used [Antonenko et al., 2010; Lee and Tan, 2006]. Passive observations of task performance can also be used as a form of evaluation [Chikhi et al., 2022; Ortega-Gijon and Mezura-Godoy, 2019]. These methods, however, are flawed in certain ways. For one, interviews and scale rating systems are primarily subjective measures and do not necessarily reflect the whole reaction of a participant. Observation methods used for evaluating task performance also do not consider the amount of brain resources someone had to use to achieve the observed performance and are not considered efficient [Chikhi et al., 2022; Kumar and Kumar, 2016]. EEGs can help balance these flaws of conventional methods. With the frequencies they are able to record, more factual calculations of cognitive load and engagement become possible. Concepts such as the cognitive load are fundamental in the study of human performance [Chikhi et al., 2022].

EEGs are highly susceptible to artifacts.

However, an issue that arises with the usage of EEGs is their susceptibility to artifacts. These can be of both internal and external origins. Internal artifacts describe unwanted physiological movements. This can include ocular or muscle movement for example. External artifacts refer to environmental interferences which include hardware disturbances and full body movements [Kyriaki et al., 2024].

Preprocessing methods help better data quality but are limited.

Although EEG data is easily impacted by such artifacts, there are various methods available to preprocess the data. These include manual cleaning, filtering and independent component analysis [Lai et al., 2018]. Usually a combination of multiple methods is recommended to ensure the best possible data quality [Downey and Ferris, 2023].

However, these methods are limited and, if applied too extensively, pose the danger of over-cleaning the data, resulting in data loss [Kyriaki et al., 2024; Lopez et al., 2022].

This study investigates how much noise we can introduce.

This is why our study aimed to investigate the robustness and reliability of EEG data that are recorded with a commercial-grade headset. The headset we used for this purpose was the Muse (2016) by Interaxon. As EEGs are sensitive to movements, their usage is usually restricted to participants being stationary. Recordings that are contaminated by movements tend to be excluded in the final analysis [Tamburro et al., 2021]. Our study focused on controllable movement artifacts. We introduced them purposefully and of varying degrees of intensity. Moreover, we investigated how and if better results are achievable if certain restrictions alongside the movement were introduced.

The data quality based on the program's calculation yielded varying results.

It was found that depending on the intensity of the movement, the impact varies. Movement-intense recordings showed a significant impact on data. Less intense movement varied in the resulting magnitude. However, the computed results indicated an issue with the computation of the SNR in this thesis.

1.1 Outline

In the course of this thesis, the second chapter covers related work. This includes the basics of EEGs such as the recordable frequencies and the Cognitive Load Index (CLI) and Engagement Index (ENG) that are computable with the different power bands. Then, EEGs in HCI, the various forms of artifacts and some methods to preprocess and evaluate the quality of data are covered.
The third chapter covers a pre-study we conducted in the form of an online survey. Here, we establish the basis for typical HCI user studies on which we build the final user study in Chapter 4 on. Chapter 4 covers the user study with the purpose of investigating EEG noise-robustness to movement-induced noise. This includes the design of the study and the results calculated.
Chapter 5 discusses the implications and limitations of the

results we found. Lastly, Chapter 6 concludes this thesis with a summary and outlook for future work.

Chapter 2

Related Work

In this chapter, basics regarding EEGs are covered. This includes the hardware available as well as the interpretation of the frequencies that are recordable with EEGs. The usage of the individual frequency bands in calculations is covered as well. Then, this chapter looks at the usage of EEGs with a Human-Computer Interaction focus. Additionally, it is explained what artifacts in EEG data are and what methods there are to lessen the artifacts' impact on the data. Lastly, this chapter quickly introduces the Signal-to-Noise Ratio, a measure used to evaluate data quality.

2.1 Electroencephalography

EEGs use a standardized electrode placement.

Electroencephalography (EEG) is a popular, non-invasive method in research for recording electrical signals of the brain [Rao, 2013]. The recording process involves electrodes being placed on assigned spots on the scalp. These spots are commonly referred to as the 10-20 system, which has been established as the international standard for EEGs [Oostenveld and Praamstra, 2001]. Figure 2.1 (b) shows example locations of the electrodes in this system. Depending on the EEG used, the number and position in the 10-20 system of the used electrodes can vary.

EEG quality ranges from research- to consumer-grade.

Generally, commercial EEG headsets can range

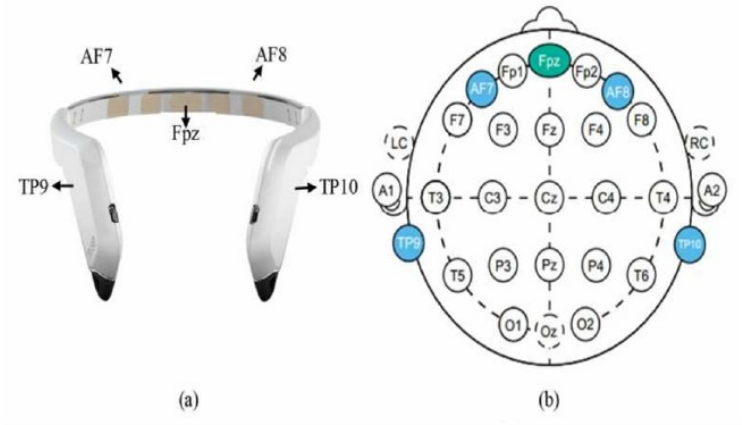


Figure 2.1: (a) The Muse headset and (b) its electrode placement in the standard 10-20 system marked in blue (right). There are four electrodes for information gain placed on the forehead (AF7, AF8) and behind the ears (TP9, TP10). Additionally, there is a reference electrode (Fpz) the middle of the forehead [Sawan and Awad, 2023].

from research-grade to consumer-grade quality [Vincenzo et al., 2026]. Research-grade products offer a high performance but are more expensive, whereas consumer-grade EEGs prioritize affordability and the ease of use for the average consumer. However, this typically leads to fewer electrodes, lower sampling rates and less adaptability of the EEG to a person's head [Vincenzo et al., 2026]. An example of a commercial-grade EEG is the Muse by Interaxon.[1] It provides four dry electrodes for information gain and an additional reference electrode. It is adjustable to the forehead by pulling and pushing the headband, also marked in yellow in Figure 2.2. For our purpose, we will use the first Muse made by Interaxon in 2016 which is no longer manufactured as of summer 2025.[1]

EEGs can record a total of five distinct frequencies. These include :

[1] https://choosemuse.com/, *last accessed September 30, 2026*

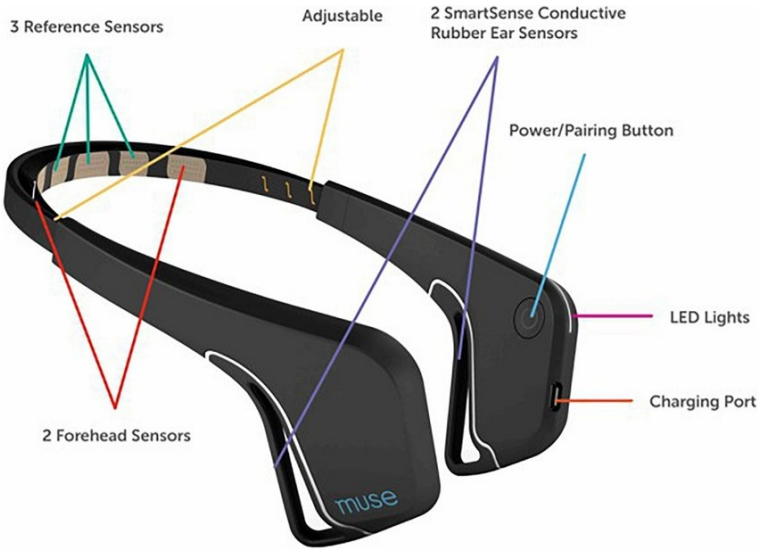


Figure 2.2: The first Muse from 2016 by Interaxon used for our study [Krigolson et al., 2021].

Delta (δ): Very low frequencies from 0.5 to 4 Hz are part of the delta power band. These frequencies are primarily dominant during deep sleep or unconsciousness [Gkintoni and Halkiopoulos, 2025].

Theta (θ): Theta frequencies describe lower frequencies ranging from 4 to 8 Hz. They are commonly associated with memory encoding, drowsiness or mediation [Müller-Putz et al., 2015].

Alpha (α): The alpha power includes frequencies between 8 and 13 Hz and is commonly linked to relaxed wakefulness. This can be seen especially when one closes their eyes [Müller-Putz et al., 2015].

Beta (β): Higher frequencies between 13 and 30 Hz are part of the beta power band. Beta power is typically observed during concentration, cognitive load or motor planning [Gkintoni and Halkiopoulos, 2025].

Gamma (γ): Gamma power describes very high frequencies over 30 Hz. It is usually associated with arousal and perceptual binding mechanisms. However, it has a lower amplitude and is highly susceptible to noise [Müller-Putz et al., 2015], making consumer-grade EEGs less

reliable to measure gamma frequencies accurately [Cannard et al., 2021].

CLI and ENG can give more insight into a persons state of mind.

While the separate power bands are already informative, using the ratios between certain frequencies can help gain a more in-depth insight into a person's state of mind. Two popular examples are the Cognitive Load Index (CLI) and Engagement Index (ENG).

2.1.1 Cognitive Load Index

There is no gold standard for the term "cognitive load" in HCI.

There is a lack of definition for the term of "cognitive load" in the field of Human-Computer Interaction (HCI). Other terms such as "mental workload" or "cognitive workload" are used interchangeably to describe the same effects [Kosch et al., 2023]. Kosch et al. [2023] point out that these terms are significantly different and neither rooted nor perfectly suited for the usage in HCI. Generally, cognitive load is used to describe the cognitive demand caused by and the mental effort required for a given task. The cognitive load can be calculated by using ratios between the alpha and theta power bands. It is defined as:

The cognitive load is the ratio between θ and α.

$$CLI = \frac{\theta}{\alpha}$$

[Gkintoni and Halkiopoulos, 2025]. This corresponds to an increase of theta band being linked to an increase of cognitive demand and task difficulty and an increase of alpha band activity being linked to decreased attention [Raufi and Longo, 2022].

A persistently high CLI has negative impacts on a person's mind.

A continuously high CLI affects a person's performance in tasks, can be harmful for learning effects and can result in more human errors, which are especially expensive in critical tasks [Khan et al., 2025; Kosch et al., 2023; Kumar and Kumar, 2016].

2.1.2 Engagement Index

The Engagement Index equals to β divided by the sum of θ and α.

The Engagement Index (ENG) is used to indicate how at-

tentive and engaged a person is during a certain task [Pope et al., 1995]. Several ways ways of calculating have been proposed [Coelli et al., 2015]. The most frequently used formula is:

$$ENG = \frac{\beta}{\theta + \alpha}$$

This corresponds to increased beta power being linked to an increase in brain activity, and assumes that an increase of theta and alpha power is related to lower alertness. A higher ENG value indicates a higher level of mental attentiveness and capability to react quickly to external events [Coelli et al., 2015].

2.2 EEGs in HCI

EEGs have a strong medical background but are established as a subfield of HCI for numerous applications.

EEGs have a wide range of use cases they are applied in [Soufineyestani et al., 2020]. Originally, EEGs were commonly used for medical purposes in clinics, for instance in neurology for the diagnosis of brain diseases [Lee and Tan, 2006; Soufineyestani et al., 2020]. However, their importance for HCI research has been growing over the years as well. Since 1996, there has been a steadily increasing number of articles related to the integration of brain signaling into HCI research [Putze et al., 2022]. The usage of brain signals in HCI can vary greatly, especially from explicitly controlling interfaces to HCI evaluation. Brain-Computer Interfaces (BCIs), a subfield of HCI [Riedl et al., 2010], are an example for applications of brain signaling where an interface is used as an active input, for instance. BCIs can also give people with motor or communication disabilities the opportunity to communicate with their surroundings by recording and translating their brain signals and replacing conventional input methods needed otherwise [Chang, 2019]. These applications have been extended to able-bodied users in non-medical contexts as well. Non-medical applications can include purposes such as neurogaming [Soufineyestani et al., 2020].

Using CLI and ENG can be used for EEGs as evaluation tool.

However, EEGs also have potential as evaluation tool. As of now, HCI evaluation relies heavily on interviews, questionnaires or tools such as the NASA-Task Load Index (NASA-TLX), which include subjective answers and

rating scales [Antonenko et al., 2010; Lee and Tan, 2006]. While assessing emotions with EEGs is not yet possible [Suhaimi et al., 2020], EEGs can be applied to study the cognitive load, engagement and stress levels [Soufineyestani et al., 2020]. Especially cognitive load, engagement and possibly emotion assessment can be advanced greatly by the usage of EEGs [Frey et al., 2013]. These measures can address problems such as a possible bias or external manipulation and can provide real-time information as well. Conventional evaluation methods like the NASA-TLX cannot provide this necessarily [Ortega-Gijon and Mezura-Godoy, 2019]. Moreover, EEGs can help in understanding a user's intention more thoroughly [Chang, 2019].

2.3 Artifacts in EEG Data

Artifacts in EEG recording can be divided into internal and external.

EEG recordings are susceptible to various forms of artifacts. These can be of both internal and external origin. Internal artifacts describe physiological activity such as ocular, cardiac or smaller muscle movement. In this thesis the most dominant types of internal artifacts are ocular movements, such as eye blinking or eye movement, and muscle movements, like swallowing and talking. External, or extra-physiological artifacts, refer to instrumental disturbances, such as electrode displacement, interferences or bigger movements, such as head or full-body movement [Islam et al., 2016].

Low- and high-noise describe the intensity of movement associated with an artifact.

Furthermore, Kyriaki et al. [2024] divide tasks into low- and high-noise. Low-noise tasks include singular tasks that do not engage a large amount of muscles. These tasks can include mouse movement or clicking, computer games and learning tasks like mental math. High-noise tasks, in general, include movements which require more muscles. Additionally, Kyriaki et al. [2024] divide these tasks further into reduced and elevated. Reduced-high-noise refers to movements engaging multiple muscles at once while a person is sitting or standing. Talking, typing on a keyboard and head movement are examples of such tasks. Lastly, elevated-high-noise includes tasks where a participant is in motion.

2.4 EEG Preprocessing

As artifacts are not completely preventable, there are various preprocessing methods aimed to reduce noise and non-neural activity. The most prevalent methods are filtering and Blind Source Separation (BSS) methods, mainly using ICA models [Kyriaki et al., 2024]. Filtering is often used to reduce the effects of very high or low frequencies and improve their interpretation [Peng, 2019]. For that, high-pass, low-pass, bandpass and notch filters are available. High-pass filters typically range from 0.1 to 1 Hz, low-pass filters are commonly in a range of 40 to 100 Hz. When using a bandpass filter, the frequencies between an upper and lower bound are kept. Lastly, notch filters are mostly used to filter out line noise at 50 to 60 Hz [Newson and Thiagarajan, 2019].

Some of the most important preprocessing methods include filtering and ICA.

ICA is one of the most popular and most investigated methods for movement-induced artifacts [Gorjan et al., 2022]. When using ICA, the data is decomposed into a set of independent components (ICs). ICs can be visually inspected and rejected if they are considered too noisy [Peng, 2019]. Ideally, before ICA is used, the data is prepared by filtering it, for instance. This generally improves the quality of ICA. However, ICA is more reliable when recording with more electrodes [Gorjan et al., 2022]. Additionally, manual artifact rejection still plays a major part during the preprocessing. During this step, EEG data is inspected and certain parts are rejected and deleted from the recordings. It is usually recommended to use a combination of multiple methods, since as of now there is no single method that can clear artifacts of EEG data completely and reliably [Downey and Ferris, 2023].

ICA decomposes the signals into independent components that can be rejected manually.

However, these methods cannot eliminate all noise and pose the danger of information loss especially if they are over-cleaned. This risk is higher when an EEG has fewer electrodes [Gorjan et al., 2022; Lai et al., 2018].

There is a danger of over-cleaning the data.

2.4.1 Signal-to-Noise Ratio

The Signal-to-Noise Ratio (SNR) describes the ratio be-

SNR describes the quality of data.

tween the signals and noise, both usually measured in root-mean-square (RMS) amplitude. The signal describes the part of the data that is desired, anything else is considered noise. Often, the SNR is expressed in decibels (dB) and defined as:

$$SNR_{dB} = 20log_{10}(\frac{signal}{noise})$$

It is used to describe the quality of the recorded data. A higher SNR usually implies that the data is cleaner and is more desirable [Semmlow, 2018]. When computing the SNR with power instead of amplitude, the calculation changes to

$$SNR_{dB} = 10log_{10}(\frac{signal}{noise})$$

Chapter 3

Survey on Typical HCI User Study Settings

In this chapter, the pre-study is covered which was conducted in the form of an online survey. The following covers the research question we formulated. Then, the structure and contents of the final survey are explained. Lastly, the answers received from the survey are presented. The following explains the basis on which the final design of the study is later built.

3.1 Research Question

In order to make our study design more representative of HCI user studies, a survey was conducted with SoSciSurvey[1], a free German survey platform. The goal was to identify common tasks, movements and settings researchers tend to use in their studies. Thus, we ended up with our research question:

RQ1 What noise does typically occur in HCI user studies?

[1] www.soscisurvey.de, *last accessed September 30, 2026*

Noise in the context of RQ1 refers to movements and motions participants use in studies. This type of noise can be consciously introduced and regulated in our final study design.

3.2 Method

In the following, the survey design, a pilot survey and the demographics are covered.

3.2.1 Survey Design

The survey consisted of the demographics and three questions.

The survey consisted of three pages. The first page was the informed consent participants had to agree to before proceeding. The next two pages included the demographics and three different multiple-choice questions. Pre-made answers were provided to make completing the survey less time-intensive. However, the last option of every question allowed participants to add their own remarks.

Demographics covered expertise, the home institution and research interests.

The demographics included four questions. The first two questions were single-choice. The first question asked for a participant's current occupation (students were also included as an occupation). To get a better feeling, for how much experience participants had, the second question covered the amount of studies a researcher had conducted in total. The given answer options were divided into four intervals. As we reached out to multiple universities, the home institution was also asked for. For the universities we directly reached out to, pre-made answers appeared once they typed in their institution's name. Lastly, participants were asked to indicate their research interests. For this question, they had free text fields to write down all of their research areas.

The three questions covered study settings, tasks and motions.

On the last page of the survey were three multiple-choice questions. We ended up with three different questions to determine (1) usual settings in studies (e.g., whether a person was sitting or standing), (2) the tasks that are usually used and (3) motions that need to be used in order to complete tasks. This division was made to view the conditions

a user is in, the cognitive performance that is asked of them and their movement (and thus the noise that should be introduced) more distinctively.
In the questions, a sentence beginning was provided, for which participants had to choose the answers they felt to be most representative. The prompts were as follows:

Q1 - Setting "A typical study includes a participant ..."
01. Sitting in front of a laptop/computer | 02. Standing in front of a laptop/computer | 03. Using a phone | 04. Using pen and paper | 05. Using an iPad or such with a pen | 06. Wearing a VR headset | 07. Standing in a room | 08. Other

Q2 - Task "A participant may be asked to ..."
01. Write a text | 02. Draw something | 03. Read a text | 04. Interact with a touch screen | 05. Listen to music | 06. Play a video game | 07. Interact with a VR Headset | 08. Other

Q3 - Motions "A participant may be ..."
01. Clicking a mouse | 02. Moving a mouse | 03. Drawing with a mouse | 04. Typing on a keyboard | 05. Using a pen to draw/write | 06. Using a controller | 07. Talking | 08. Using hand gestures | 09. Moving their head | 10. Walking | 11. Other

All pages of the survey are also added in Appendix A.

VR and AR are omitted in our study.

Since VR and AR are part of our institute's research as well, they were included for representation. However, in the final study design we omit the usage of AR- and VR-glasses as the simultaneous usage of such glasses and the Muse was not feasible.

3.2.2 Pilot Survey

Before reaching out to researchers, a short pilot survey with one student assistant from our institute was conducted.

The phrasing of questions and layout of the survey were adjusted to be clearer.

Originally, the survey consisted of more pages. Additional text fields would be shown on a separate page if the answer option "Other" was previously chosen. However, this was perceived as more confusing instead of helpful. Based

on this feedback, we omitted the additional pages and collected all three questions on one single page. Instead of extra questions on separate pages, "Other" had a free-text answer field to directly add own ideas. The option could be chosen manually or automatically once an addition to the text-field was made.
Ready-made options were added to choose from in the home institution demographics. This was done to ease the indication of the home institution and provide a clearer structure. Instead of simple free form text fields, suggestions would appear when there was a match between answer options and the respective institution's name.

3.2.3 Participants

The survey was targeted at researchers in the field of HCI and research assistants, who have experience with HCI user studies. We reached out to 53 different researchers from eight German universities. Additionally, we sent a link to the survey in the institute's general Slack channel.

20 HCI researchers participated in the survey.

In total, 20 researchers of various research fields responded. The research fields were divisible into six major categories. These were AR/VR, accessibility, shape-changing interfaces, dark patterns, games and social research. Researchers of a total of seven German universities, including RWTH Aachen University, KIT, TU Darmstadt, Dresden University of Technology, Leibniz University Hannover, University of Würzburg, and University of Stuttgart took part in the survey. 15 of the researchers were doctoral students, and the other five were master's students. 11 participants reported having conducted between three to five studies, four had conducted a total of six to ten studies, three reported having conducted one to two studies and the remaining two reported having conducted more than a total of ten studies.

3.3 Survey Results

Sitting in front of a screen is the most used setting.

A majority of researchers (see Figure 3.1) reported that

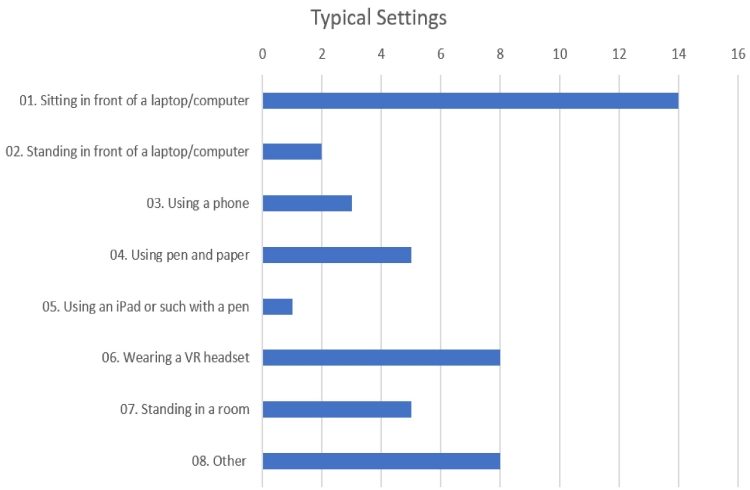


Figure 3.1: Result of "Q1 - Settings" in the Survey.

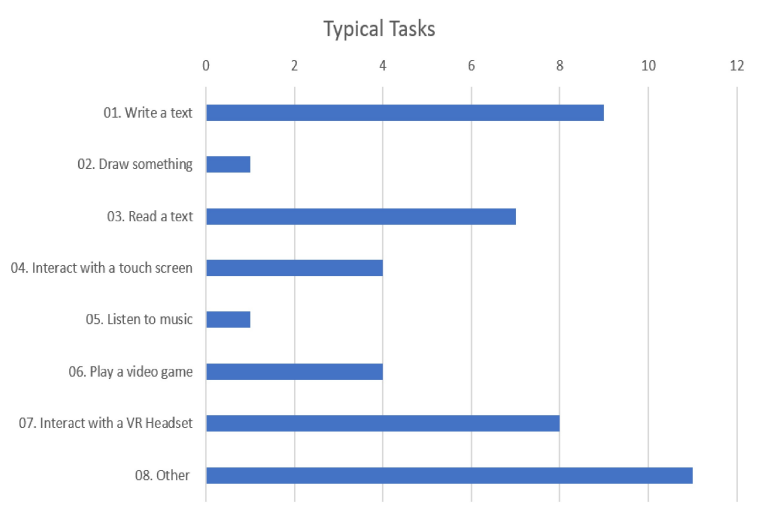


Figure 3.2: Result of "Q2 - Tasks" in the Survey.

they typically use settings where a participant is sitting in front of a laptop or computer. As this was by far the most chosen answer option, it was prioritized using this setting in the final design of the study. Thus, this setting was applied in more tasks. Additionally, settings where someone is wearing a VR headset, standing in a room or using pen and paper were part of the more utilized

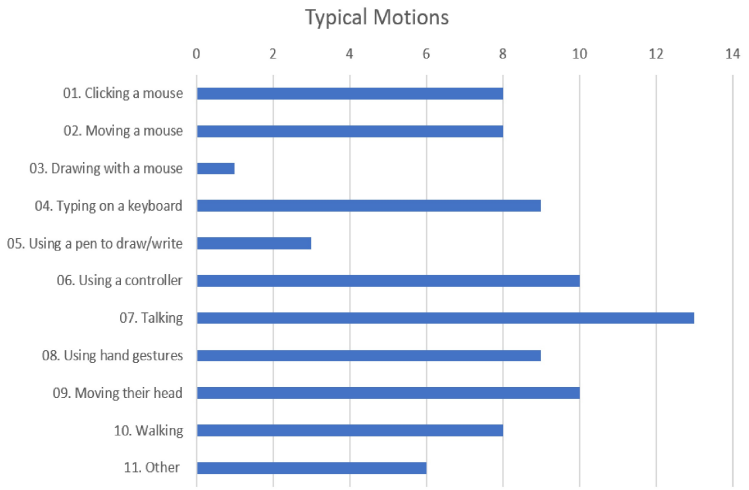


Figure 3.3: Result of "Q3 - Motions" in the Survey.

study settings. Lesser chosen settings include the usage of phones, iPads and the operation of laptops and computers while standing. Besides the provided answers, some other settings were remarked in the category "Other". Additional settings mentioned included AR headsets, the testing of prototypes, sitting in a room without a screen, smart watches and the addition of textile interfaces to the study setting.

The most common tasks were writing, reading and interacting with VR-headsets.

Answers for typical tasks were more broadly distributed. Most commonly, tasks included writing and reading texts, as well as VR-headset interactions, as seen in Figure 3.2. Additionally, talking and answering questions in interviews was mentioned by five researchers. This made talking part of the three most used tasks for HCI user studies in addition to the provided answer options. Other tasks mentioned once or twice as additional task, include the interaction with textile interfaces or 3D-printed objects, and the usage of AR headsets. Out of the provided answer options, interactions with touchscreens and video games were also used frequently according to four researchers. However, listening to music and drawing were chosen only once.

Motions mainly include talking, controllers and head movements.

For the last question regarding motions in user studies, researchers chose a variety of answer options. The most chosen movement was talking which was ticked a total

of 13 times, as seen in Figure 3.3. Additionally, controller handling, head movements, hand gestures, keyboard typing, mouse usage and walking were chosen between eight to ten times. Other motions mentioned included drinking and dancing. Furthermore, one researcher mentioned that talking in their studies usually included gesturing and head movement unconsciously which is why the focus was put on more conscious movement such as mouse and keyboard usage in addition to talking.

Chapter 4

User Study

This chapter starts with two research questions. Then, the methods used during the study are covered. This includes the program used to record data, the study process and the pilot study that was conducted before the beginning of the final study. This chapter will cover the design of the tasks and counterbalancing measures. Finally, the data of both the EEG recordings and questionnaires is analyzed.

4.1 Research Question

Our main goal with this study is to investigate, how robust EEG-frequency band measurements are with the Muse. Following that goal, we aimed to test the limits of the consumer-grade headset but also aimed to seek alternatives to ease the cleaning of EEG data as preprocessing methods are less reliable when using fewer electrodes [Gorjan et al., 2022; Lai et al., 2018]. Thus we formulated the following two research questions:

RQ2 How can we limit noise?

RQ3 To what extent can we allow noise?

We investigate movement reduction and the limits of preprocessing methods.

Noise in the context of RQ2 and RQ3 refers to movement-

induced artifacts. With RQ2 we aim to investigate how unwanted and unconscious movement during the recording process can be limited. RQ3 aims to investigate how much movement can be allow and still get meaningful information out of the data. Here the focus lies in the combination of physical and verbal noise restrictions and preprocessing methods. As a measure, the SNR data was analyzed.

4.2 Method

The study was approved by RWTH Aachen University's ethics committee.

Before the start of the study, the study was approved by the interfaculty ethics committee of RWTH Aachen University. This was required due to the recording and evaluation of EEG data.[1]

4.2.1 Recording Method

Experiment conditions can be created once and are reusable.

To record the EEG data, we used the program by Bangert [2026]. The interface provides the option to create and repeatedly use a study setting, enabling a data comparison of the CLI and ENG between all recorded sessions. In addition to that, it also ensures that the durations and naming of the different tasks remain the same throughout multiple study sessions. Once the experiment and its conditions are set up and the Muse was connected successfully, the program proceeded with a 30 second electrode check and a simultaneous baseline recording. As seen in Figure 4.1 the interface provides an explanation for the contact quality of the individual electrodes. Once these were considered satisfactory, the program proceeds to the recording of the conditions. The order in which the conditions are recorded and how often they are recorded are set manually.

Filtering, ICA and manual artifact rejection are supported as preprocessing methods.

When all conditions are recorded, it is proceeded to the analysis and preprocessing of the raw recording. The preprocessing methods that the program provides are filtering,

[1] https://www.rwth-aachen.de/cms/root/wir/einrichtungen/rektorat/~bezzxa/fakultaet ethikkommission, *last accessed September 30, 2026*



Figure 4.1: An example electrode check screen from the interface by Bangert [2026]. If the contact was not satisfactory, red and yellow text bubbles informed of the insufficient electrode contact. Then a rerecording of the baseline and a new electrode check could be done

ICA and manual artifact rejection. Once the preprocessing was finished, the data analysis was possible. This analysis includes the SNR, CLI and ENG among other values. An experiment wide comparison of the CLI and ENG for all study recordings was provided as well.

4.2.2 Final Study Process

After the pilot study, the final process of the study was set and is described in the following. The whole duration of one study was between 60 and 90 minutes.

First, participants were given an introduction.

Before the start of the recordings, the purpose and goal of the study were explained to the participants. It was also pointed out throughout the study that the performance of a participant would not be assessed or evaluated in any way. Instead, participants were told that the most important factor was their concentration.

Secondly, we prepared the skin and Muse with water.

Then we fitted the headset. For that, participants with longer hair were asked to tie their hair into a ponytail and if they had bangs, a hair clip was used to move the hair out of the way. After this, cotton pads drenched in water were used to wet the skin on the forehead and behind the ears. We did the same thing for the electrodes on the Muse to reduce impedance [Cannard et al., 2021].

An electrode check and two baseline recordings were made.

Once the Muse was fitted correctly, the study continued with an electrode check during which the baseline was recorded simultaneously. The electrode check lasted 30 seconds. During this time, participants were asked to sit in a

relaxed position with their eyes closed and not move. After the recording finished, the results of the electrode check were reviewed. Figure 4.1 shows an example for such an electrode check. If the contact of an electrode was shown to not be sufficient, the fitting of the Muse was adjusted and the baseline rerecorded. This was done until all electrodes were sitting adequately. The study then proceeded with a second baseline recording during which participants had their eyes open. The participants were asked to blink and move as little as possible during the 30 seconds the recording lasted. After all baseline recordings were completed successfully, the study continued with the actual tasks and noise settings.

4.2.3 Pilot Study

Before the beginning of the actual study, a pilot study with one participant was conducted. This lead to a few adjustments made to the process of the study and the tools used.

The headrest was reconstructed.

Originally the headrest consisted of a laser-cut box with a 3D-printed headrest put on top of it. However, the box was too bulky and made it uncomfortable to reach the keyboard. Thus, this was changed to the headrest being mounted to a construction of aluminum extrusion profiles. We made sure that the horizontal aluminum extrusion was long enough so that participants could comfortably place their hands in between the support extrusions. This construction allowed a more precise height adjustment as well.

We revised the phrasing for parts of the study.

Additionally, we observed that certain verbal restrictions and task explanations were not understood correctly. This is why the phrasing of these instructions was revised to be clearer and easier to understand.

Safety concerns were addressed.

Originally, alcohol wipes were used to clean the participants' skin. Cleaning with alcohol wipes was recommended for a better contact between the skin and the Muse [Cannard et al., 2021]. However, due to concerns regarding possible skin irritation, we decided to omit this step completely. Instead, we used water to lightly clean the forehead and behind the ears for every participant.

4.2.4 Task Design

Tasks were designed for specific power bands.

The design of our study was based on the interactions determined by the survey in Chapter 3. The tasks given to participants in the study were designed to address the most relevant power bands in typical studies (theta, alpha and beta). Delta and gamma frequencies were omitted due to their dominance in tasks irrelevant to this study [Long et al., 2021] or their high susceptibility to noise [Pattisapu and Ray, 2023]. For every task, a different type of noise was introduced. The task was then executed a total of three times under three different conditions. These conditions were:

No Noise (NN): Participants were instructed to move as little as possible while executing a task. In this condition, we did not introduce any noise. Instead, the no noise condition functions as a baseline for each task.

Noise (NZ): In this condition, we introduced noise. For each task a different movement was introduced as noise, based on the previous survey described in Chapter 3.

Reduced Noise (RD): For this condition, restrictions were combined with the introduced noise.

The reduced noise settings had two possible setups:

Setup 1: A combination of verbal and physical restraints was used. Physical restraints included a headrest and a footrest. The headrest (see Figure 4.2 (a) and (b)) was a construction combining a 3D-printed part and aluminum extrusions. The footrest (see Figure 4.2 (c)) was bought at IKEA.[2]
Setup 2: Only verbal instructions were used to restrict the participants. This setup was used for settings where a participant was standing and a foot- or headrest were not feasible.

This design resulted in one DV and one IV.

These setups and conditions resulted in one depen-

[2] www.ikea.com/de/de/p/dagotto-fussstuetze-schwarz-40240989/, *last accessed September 30, 2026*

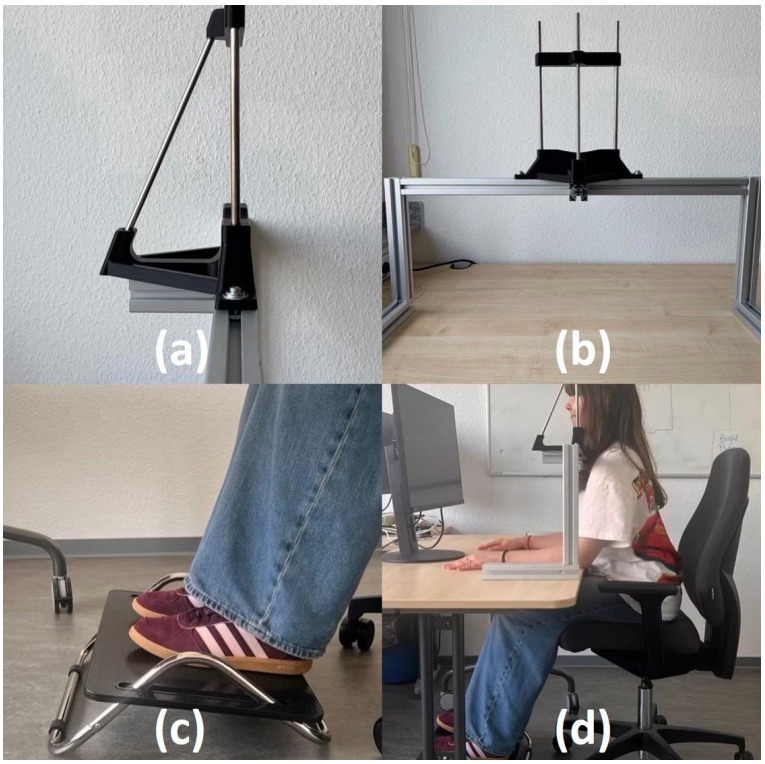


Figure 4.2: (a) A view of the headrest from the side. The support for the forehead sticks out a bit. (b) The whole headrest construction from the front. The horizontal aluminum profile was chosen longer to allow a more comfortable placement of the arms between the support profiles. (c) The footrest provided a set place for the feet. (d) A mock setup combining the footrest and headrest.

dent variable (DV) and one independent variable (IV) with multiple levels for each task. The DV describes the Signal-to-Noise Ratio (SNR) and the IV the movement and restrictions that were introduced. The different levels of the IV are the corresponding baseline (eyes open if the task was done with the eyes open or eyes closed if not), the no noise, noise and reduced noise condition.

A questionnaire to evaluate comfort during the study was used.

At the beginning of the study, participants were given an informed consent and a demographics form they were asked to fill out after they had received an introduction to the study. The questionnaire was available in both English

and German. Participants were free to choose between these two languages. After every noise and reduced noise setting, participants were given a short, multiple-choice forms to fill out. We omitted the form for the no noise settings since these functioned as a baseline and were not the main object of our research questions. The questionnaire consisted of five statements and an additional field for remarks. For each statement there were five answer options provided, ranging from strongly agree, agree, neutral, disagree and strongly disagree, following a likert scale [Lazar et al., 2017; Nemoto and Beglar, 2014]. The statements were as follows:

Q1 I felt uncomfortable due to the headset.

Q2 I felt too restricted in my movement.

Q3 I felt comfortable executing the task.

Q4 I could execute the task with my full attention.

Q5 My behavior felt natural.

The statements were used to evaluate how comfortable a participant felt with the Muse and possible restrictions. In the course of this, we also aimed to evaluate the feasibility of such restriction settings for future HCI user studies in case they yielded higher quality data. The investigation of the feasibility applied to both the setups as well as the usage of the Muse.

Use of the questions was to consider a participants comfort for evaluation as well.

In the following, each task and the corresponding introduced noise and restrictions are described. All tasks were created to use some of the three power bands (alpha, beta and theta) that are typically relevant in user studies.

Reading

Participants were given one text per condition, each an article taken from The Economist.[3]

[3] https://www.economist.com/, *last accessed September 30, 2026*

We introduced keyboard typing as noise. For this, we used a Logitech keyboard with a QWERTZ layout. Typing is classified as reduced-high-noise according to Kyriaki et al. [2024].

The recording lasted two minutes while a participant was reading a text.

We asked participants to read the articles that were provided with as much attention as possible for two minutes. During the noisy settings, participants had a split screen view. On the left side they saw one page of the article; on the other there was an empty text file. The task participants were given, was to read the text attentively while simultaneously typing it out in the text file. The instructions were to correct typos that were instantly noticed. Any other mistakes made earlier in the text were supposed to be disregarded. In the restricted version we used setup 1. Verbal instructions were to not move and keep the arms in a fixed position on the table. The only parts of the body that were supposed to move were the hands and fingers.

Beta power is targeted.

As beta band oscillation is associated with visual and more importantly attentional tasks, this task is primarily aimed at the beta band [Ronconi et al., 2020].

Memory

At the beginning the card deck was shown.

Before beginning to play memory, participants were shown a collage of all the cards in the deck of the memory. This was done to give the participant the same knowledge about the cards for each round played. In total, each participant played three rounds of memory each time with a different, randomized field. The memory code we used for this task was generated by Claude.

The noise we introduced for this task was the usage of a computer mouse. This covered both moving and clicking and is considered low-noise [Kyriaki et al., 2024].

We recorded 30 seconds while the participant was recreating a card field.

After this, they were given 30 seconds to remember a randomized field and another 30 seconds to replicate it. Recordings were only made during the replication time. For the no noise setting, participants were shown one card of the field for 5 seconds and had to actively recall where they would place it on the memory field. When using the mouse, they were able to replicate the field by dragging and dropping the cards on the field. For the restricted noise,

setup 1 was used. The verbal instructions were to keep as still as possible. Participants were also told to place their unused hand flat on the table and not move it.

This task is targeting the theta and beta power bands. An increase of theta power is commonly related to the amount of information retained in the memory. Beta power is involved for attentional tasks and other cognitive processes, such as working memory and motor processing [Chikhi et al., 2022].

Theta and beta power are targeted.

ABC

In this task, participants were asked to recite the ABCs for 90 seconds, using the language they were most comfortable with. The recitation was done with various adjustments while having their eyes closed. If they finished reciting the ABCs before the 90 seconds were up, they were instructed to loop the ABCs and keep on reciting.

Speaking was the noise we introduced in this task. The intensity can be classified as reduced-high-noise according to Kyriaki et al. [2024].

For each noise condition, the task differed slightly. During the no noise condition, participants were told to sit comfortably with their eyes closed and recite the ABCs backwards. When talking without any physical restrictions, the task was to recite the ABCs forwards while only saying every third letter out loud. The beginning, for instance, was "A _ _ D" and if Z was reached, they looped to the next letter C. Moreover, the instructions were to sit comfortably in the chair. In the restricted condition, we used setup 1. Verbal instructions were to keep still and lay both hands flatly on the table and not move them. Then, the task was to recite the ABCs backwards while only saying every second letter out loud (for instance, "Z _ X").

The task was to recite the ABCs with different tweaks.

In this task, we target alpha and beta power. Alpha power is dominant when a person closes their eyes [Rao, 2013]. Additionally, this task engages beta power related processes [Frelih et al., 2025].

This task targets alpha and beta power.

Mental Math

Participants were told a series of mathematical equations they had to solve.

For this task, participants were told a series of mathematical equations. They were told one equation, for which they had three to seven seconds (depending on the equation) to solve as well as possible, before getting a new one. The calculation was solely in the participant's head, with no confirmation of how much a participant had calculated.
As noise, walking was introduced in this task. According to Kyriaki et al. [2024] the extent of this movement can be considered elevated-high-noise.

Walking and standing while doing mental math were the two conditions.

When recording this task's baseline, the participant was sitting in a chair and asked to look at a fixed point on a black screen. Then, for 90 seconds, they heard one equation they had to solve as well as possible before the next one. During the noise condition, they got different equations while walking around in the room. Here, instructions were to keep walking continuously in a shape they were comfortable with during the 90 seconds. In the reduced noise condition, we used setup 2. We asked the participants to stand still in the room and look at a fixed point on the room's wall while doing the calculations.

Beta and theta power are targeted.

Here beta and theta power were targeted. Both power bands have been associated with mental arithmetic [Garakh et al., 2015].

The study uses a within-group design.

Every participant in our study executed every task under the three different conditions mentioned previously. Thus, we ended up with a within-group design for this study.

4.2.5 Counterbalancing

Latin Squares for the tasks and conditions were made.

To address the problem of possible order effects in our task design, we used counterbalancing for our tasks and conditions [Lazar et al., 2017]. For that, we created a 4x4 Latin square for the tasks. We double checked this with a Latin square generator provided in the toolkit by Schwind et al. [2023] and ended up with the square in Figure 4.3 (a). Additionally, we made a 3x3 Latin square for the individual conditions resulting in six different sequences as seen in

(a) Tasks

Subject	1	2	3	4
#1	Reading	Memory	ABC	Mental Math
#2	Memory	Mental Math	Reading	ABC
#3	Mental Math	ABC	Memory	Reading
#4	ABC	Reading	Mental Math	Memory

(b) Conditions

Subject	1	2	3
#1	No Noise	Noise	Reduced
#2	No Noise	Reduced	Noise
#3	Reduced	No Noise	Noise
#4	Reduced	Noise	No Noise
#5	Noise	Reduced	No Noise
#6	Noise	No Noise	Reduced

Figure 4.3: Two Latin Squares made with the generator by Schwind et al. [2023] for cross-checking the sequences of (a) the tasks and (b) the conditions.

Figure 4.3 (b). Each condition sequence of conditions was assigned the respective number from their row as identification.

The two results were combined in a latin rectangle.

We then had to combine the results of two squares. Due to the number mismatch of the two squares (eight sequences for the tasks and six sequences for the conditions) we used a latin rectangle of size 12x4 [McKay and Wanless, 2005]. For that, we assigned every task an index from 0 to 3. Then, we calculated the condition with the formula $((row + task_index) \mod 6) + 1$, assigning every task each condition exactly twice. Furthermore, every condition sequence in the table is only used once. A table of the final layout is shown in Appendix D. With this design we needed a minimum of 12 participants for our study, in order to cover all different settings.

4.2.6 Participants

Participants were recruited from the institute and via personal networks.

Participants were recruited via the institute's Slack channel and through personal networks using snowball sampling. To make appointments, we used DFN Terminplaner[4], a privacy-friendly scheduling tool for the German research community. Besides basic English proficiency to understand the texts, no other knowledge was required. The participation was voluntary and not compensated.

We recruited 25 participants, with one dataset being omitted in the final analysis due to electrode misplacement.

In total, n = 25 people from ages 19 to 59 (MD = 22.5; SD = 9.81) could be recruited. 13 participants identified as male, nine as female, two as non-binary and one participant preferred not to disclose their gender. Most participants (n = 16) had a high school diploma, four had a bachelor's degree, four a master's-level degree and one doctor of laws. A majority were students (n = 22) out of which five had student assistant jobs. The remaining three participants worked as a secretary assistant, civil engineer and lawyer. In the analysis only 24 of the recordings were considered, as in one recording the Muse headset got loose and falsified the data. However, the data from the questionnaire will still be evaluated normally.

4.3 Data Analysis

Data was preprocessed using all the tools we had available.

Before analyzing the data, every recording was preprocessed. This included filtering, ICA and manual artifact rejection. As tasks were targeted at alpha, beta and theta frequencies only, the fitting filter was used. Thus, the filter used were a high-pass filter of 4 Hz, a low-pass filter of 30 Hz and a bandpass filter from 4 to 30 Hz. Additionally, after filtering, every recording and the ICs were visually examined for abnormalities which were rejected manually from the recordings if needed. For the analysis, the decibel SNR of all electrodes was used. The individual channels were omitted, as the aim was to analyze the general usability of the recordings as a whole in noisy settings.

[4] https://terminplaner.dfn.de/, *last accessed September 30, 2026*

(a) Friedman Test

Main code

```
fit <- rstatix::friedman_test(df, y ~ condition |
subject)
fit
```

Effect sizes

```
rstatix::friedman_effsize(df, y ~ condition | subject)
```

(b) Post-hoc

Main code

```
rstatix::pairwise_wilcox_test(
  df,
  y ~ condition,
  paired = TRUE,
  p.adjust.method = "holm"
)
```

Effect sizes

```
# df_pair should only contain the two conditions of
one reported contrast
rstatix::wilcox_effsize(df_pair, y ~ condition, paired
= TRUE)
```

Figure 4.4: The code to calculate the values for (a) the Friedman test and (b) the pairwise Wilcoxon signed-rank test with Holm correction [Schwind et al., 2023].

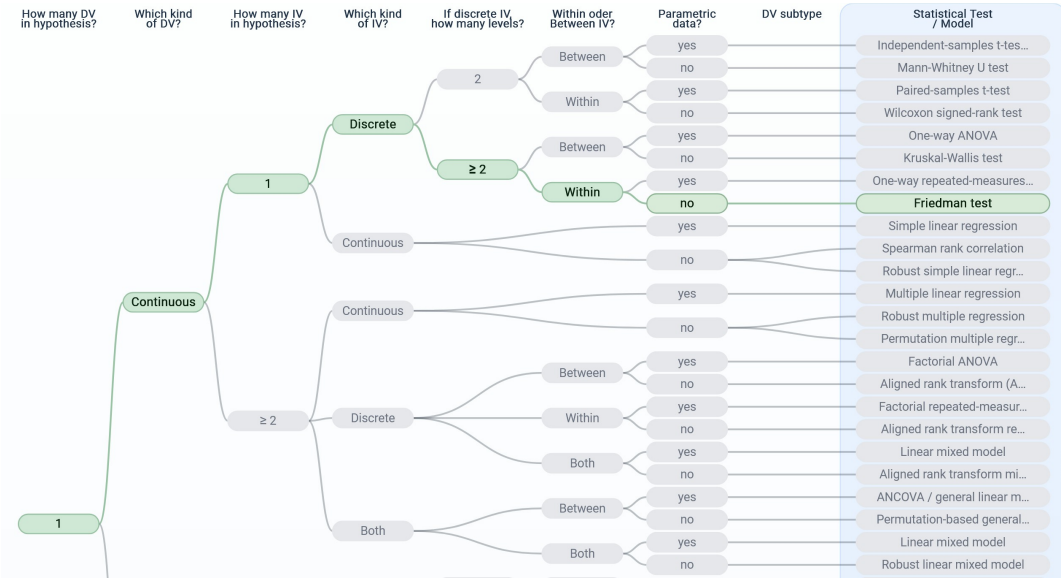


Figure 4.5: The flow of the statistical decision tree from the tool kit by Schwind et al. [2023]. The Questions (left to right) were: How many DVs are in the hypothesis? Which kind of DV? How many IV in hypothesis? Which kind of IV? If discrete IV how many levels? Within or Between IV? Parametric data? After answering all questions, the final statistical test suggested was the Friedman test, highlighted in green.

4.3.1 Evaluation Method

The evaluation of the data was done in the program R. The basis of the code to compute the results for the tests was the code provided by Schwind et al. [2023], also pictured in Figure 4.4.

Analysis was done in R with the Friedman test and with post-hoc analysis using pairwise Wilcoxon.

To evaluate the collected data, the Friedman test with the pairwise Wilcoxon signed-rank method with Holm p-value correction for post-hoc analysis if $p \leq 0.05$ was used. This value was chosen since it indicates a statistically significant difference [Pereira et al., 2015]. The Friedman test is a test designed for one DV and IV with multiple levels. It is used to analyze IVs of multiple levels for one task. If a statistical effect is observed, the pairwise Wilcoxon post-hoc test is used to compare each condition to one another [Okoye and Hosseini, 2024]. As p-value correction, the Holm correction was used. This method was chosen as it provides a strong, family-wise error correction and is less conservative than the Bonferroni method [Schwind et al., 2023].
These tests were chosen on the basis of the HCI user study toolkit created by Schwind et al. [2023], which also provides a statistical decision tree. After answering all questions in the decision tree, the final path ended with the Friedman test, as seen in Figure 4.5. To determine if the collected data was parametric or not, the Shapiro-Wilk test and Mauchly's test of sphericity were used. After visually examining the Q-Q plots and histograms, the conclusion was that the data was non-parametric.
The null hypothesis for this test was as follows:

H_0 = There is no statistical difference between all the different noise conditions per task and their respective baseline

This hypothesis would be rejected if the p-value was less than or equal to 0.05 and a post-hoc analysis would be conducted.

4.3.2 Results

The tests were applied to the tasks separately.

The Friedman test and if needed the pairwise Wilcoxon signed-rank test as post-hoc analysis were applied to all four tasks separately to analyze the decibel SNR among all recordings. Therefore, we had four different test results. For each task it holds that the degree of freedom $df = 3$, also meaning that four conditions were evaluated per task. These are the corresponding baseline and the three conditions that each task was executed in. For the tasks where a participant had their eyes open (memory, reading and mental math) the eyes open baseline recording was used and for the other task (ABC) the standard baseline where a participant had their eyes closed.
For the evaluation the cleaned and raw decibel SNRs were used. The program used in the study, computed the decibel SNR as follows:

$$20log_{10}(\frac{RMS_{Signal}}{RMS_{Noise}})$$

RMS describes the root-mean-square amplitude of either the signal or noise. The RMS is defined as

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}$$

where x_i describes either the signal or the noise [Rodrigues and Rodrigues, 2024]. In the code, signal was hardcoded and defined through a bandpass filter ranging from 1 to 30 Hz. Noise was calculated by subtracting the computed signal from the data.

Reading

The evaluation for this task resulted in a value for the Chi-square $\chi^2 = 27.8$ and a p-value that was $p < 0.01$. The value of Kendall W was 0.385 implying a moderate magnitude of the effects between the different recordings. As $p < 0.05$ a post-hoc analysis was conducted.

Effect of unreduced noise on the cleaned SNR in comparison to the task baseline was significantly large.

The post-hoc analysis showed that there was a large

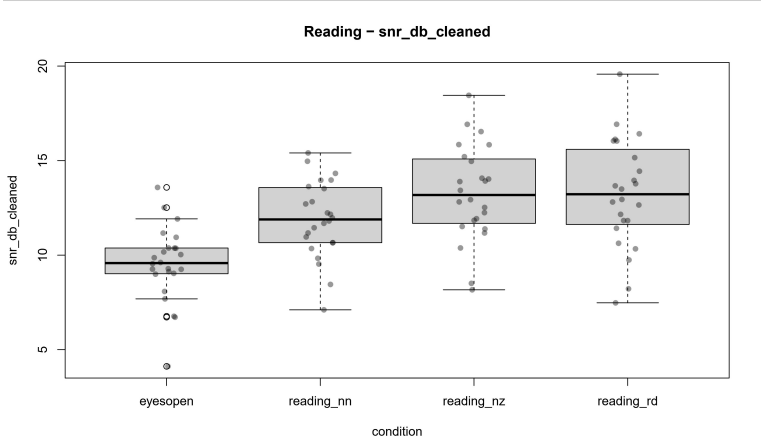


Figure 4.6: The boxplot for the reading task with the cleaned SNR. The x-axis values are (left to right): eyes open baseline, no noise, noise, reduced. The y-axis indicates the SNR cleaned value.

($r = 0.682$), statistically significant ($p_{adjusted} < 0.01$) effect on the SNR between the eyes open baseline and the no noise recording. This also held for the noisy ($r = 0.793$, $p_{adjusted} < 0.001$) and reduced ($r = 0.846$, $p_{adjusted} < 0.001$) recordings in comparison to the eyes open baseline. As seen in Figure 4.6 the median value for the SNR of the eyes open baseline was lower than the other conditions. Furthermore, the matched rank-biserial value for the difference between the no noise and noisy setting was $r = 0.525$ suggesting a large effect of the typing on the SNR value. With an adjusted p-value $p_{adjusted} < 0.05$ this difference was considered statistically significant as well. Between the no noise and reduced conditions there was a medium effect ($r = 0.443$) on the SNR. Between the different noisy conditions (noise and reduced) there was a negligible ($r = 0.035$) effect on the quality. However, both of these matched biserial-rank values were statistically not seen as significant as both adjusted p-values were larger than 0.05.

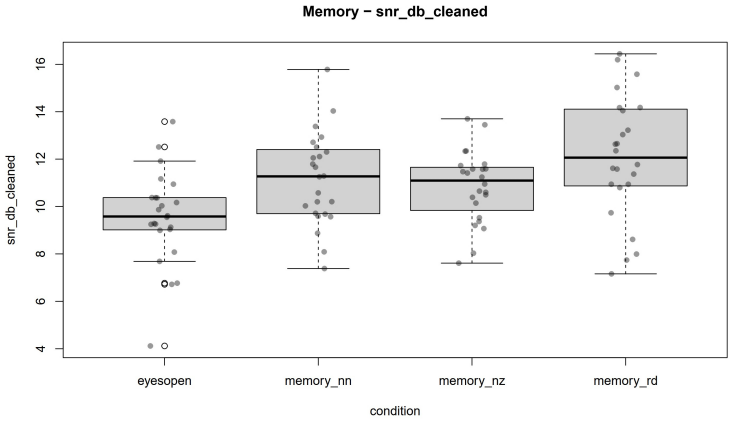


Figure 4.7: The boxplot for the memory with the cleaned SNR. The x-axis values are (left to right): eyes open baseline, no noise, noise, reduced. The y-axis indicates the SNR cleaned value.

Memory

It held that $\chi^2 = 16.4$ and $p < 0.01$, signifying that H_0 was to be rejected and a post-hoc analysis was needed. Furthermore, the effect size equaled to 0.227 implying a small magnitude of the effect according to the Kendall W [Schwind et al., 2023].

There was a significant and large effect between the two noise conditions.

The post-hoc analysis found that again, there was a large, significant ($p_{adjusted} = 0.0326$) effect on the signal quality between the general eyes open baseline and the no noise task recording with a matched rank-biserial value of $r = 0.519$. A large effect was also seen for the noise ($r = 0.537$, $p_{adjusted} < 0.05$) and reduced ($r = 0.694$, $p_{adjusted} < 0.01$) setting. This is also visible in the boxplot in Figure 4.7, where the median for the eyes open baseline was the lowest. In comparison to the no noise recording, the effect of the noisy task ($r = 0.227$) was small, whereas the reduced task ($r = 0.332$) had a medium impact. However, as both adjusted p-values were bigger than 0.05 they were statistically not significant. With a $p_{adjusted} = 0.0326$ the difference in the SNR between only the noisy recordings was significant and had a large effect on the SNR ($r = 0.542$). This

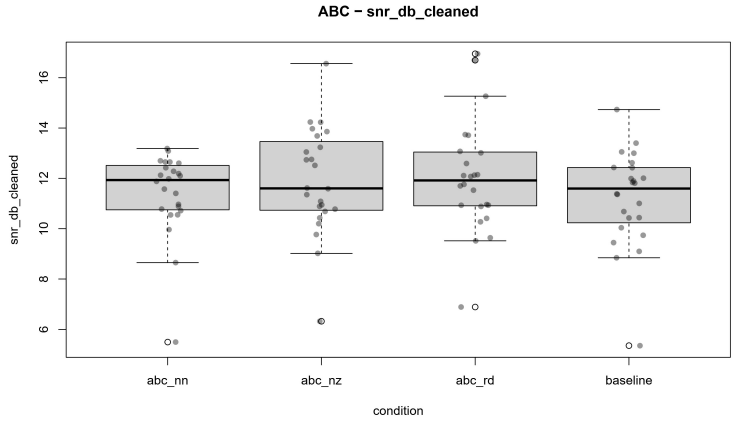


Figure 4.8: The boxplot for the ABC task with the cleaned SNR. The x-axis values are (left to right): no noise, noise, reduced, baseline. The y-axis indicates the SNR cleaned value.

is visible in Figure 4.7 as well where the reduces memory yielded a higher median SNR than the other conditions.

ABC

The Friedman test found no statistical significant difference across all four conditions

For this task it holds that $\chi^2 = 1.65$ with a p-value of $p = 0.648$. Since $p > 0.05$ there was not enough evidence to reject H_0 and a post-hoc analysis was omitted. According to the Kendall W = 0.0229, the effect size implied a small impact on the SNR between all conditions. In the boxplot in Figure 4.8 it is visible as well that the median for all tasks were in a similar range.

Mental Math

After applying the Friedman test we had the values of $\chi^2 = 29.2$ and $p < 0.01$ implying a strong incompatibility with H_0, requiring a post-hoc analysis. Besides that, the Kendall W value indicated a moderate (effsize = 0.406) difference among all four conditions.

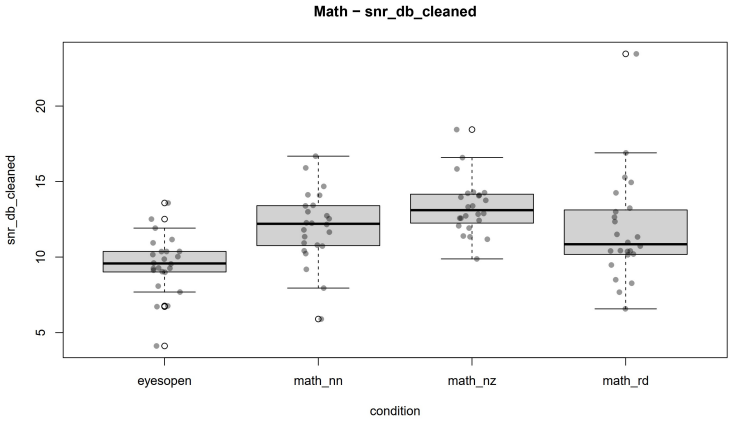


Figure 4.9: The boxplot for the mental math task with the cleaned SNR. The x-axis values are (left to right): eyes open baseline, no noise, noise, reduced. The y-axis indicates the SNR cleaned value.

There was a large, significant effect between the two noisy recordings.

There was a significantly ($p_{adjusted} = 1.61e-3$) large ($r = 0.688$) effect on the SNR between the eyes open and no noise recordings. This also held for the noisy ($r = 0.851$, $p_{adjusted} < 0.01$) and reduced ($r = 0.566$, $p_{adjusted} < 0.001$) recordings. Again, this is visible in the boxplot in Figure 4.9 where the median SNR for the eyes open recordings was lower than the other recordings. The largest ($r = 0.566$) and statistically significant ($p_{adjusted} = 1.73e-2$) effect on the quality was between the noisy (walking) and reduced (standing) math settings. Among the different task settings, this was the biggest and most significant effect observed. The difference between the no noise recording and the noisy condition was of medium magnitude ($r = 0.42$) and the effect from no noise to reduced was of small magnitude ($r = 0.181$). However, both of these comparisons also had an adjusted p-value larger than 0.05 implying that the statistical significance was disregardable.

SNR - Raw

Analysis of the raw SNR was done as an addition due to a contradictory observation in the values of the cleaned SNR.

Since it was observed that the baselines with eyes closed and eyes opened had constantly the lowest SNR values

according to the boxplots, the Friedman test with our post-hoc analysis was applied to the raw SNR. This measure was taken to double check whether the data had been over-cleaned during the preprocessing and thus the resulting SNRs were not reliable. The null-hypothesis remained the same, except this time it focused on the raw values.

Reading: This analysis with the Friedman test yielded a value for Chi-squared of 47 and a p-value $p = 3.56e - 10$. The effect size was large with $effsize = 0.652$. Since the p-value signified a strong incompatibility with H_0 a post-hoc analysis was done.

The effect among all task condition was significant, with the noise - reduced pair having the lowest impact.

The difference between the eyes open baseline and no noise was of a moderate magnitude ($r = 0.461$). In addition, the differences to the eyes open baseline and noisy recordings had both a large effect on the raw SNR (noise: $r = 0.875$, reduced: $r = 0.828$). The differences between the no noise to the noise and reduced condition were both times similarly large. Unreduced noise resulted in $r = 0.799$ and $r = 0.729$ for the reduced version of keyboard typing. The effect difference between the two noisy conditions was of medium magnitude with $r = 0.426$. All these matched rank-biserial r-values were statistically considered significant as the adjusted p-values were all less than 0.05. Especially significant were the differences for the eyes open and the two noisy settings and the no noise task recordings in comparison to both noisy settings, where the adjusted p-values were less than 0.001.

Memory: With the Friedman test, we calculated the values $\chi^2 = 19.4$ and $p < 0.001$. In course of this, post-hoc analysis was applied. For the Kendall W it holds the the effect size was of small magnitude with a value of 0.27.

The effects of the task conditions were not significant.

The pairwise comparison between the eyes open recording and the three task condition resulted in a large difference in the raw SNR, all of statistical significance. The memory no noise task compared to the noisy and reduced version resulted for both settings in an r-value indicating a medium effect on the raw SNR ($r_{nz.nn} = 0.408$, $r_{rd.nn} = 0.321$). Additionally, the difference between the noise and reduced noise was small ($r = 0.117$). It has to be noted that the task comparisons all yielded adjusted p-values larger than 0.05,

indicating statistical insignificance.

ABC: Analysis resulted in a Chi-square value of $\chi^2 = 35.6$ and a p-value of $p < 0.001$. Thus, in contrast to the cleaned SNR, a post-hoc analysis was required. The effect size was of moderate magnitude ($effsize = 0.495$).

The effect of the no noise ABC task in comparison to the baseline was moderate ($r = 0.49$, $p_{adjusted} < 0.05$). The difference between the baseline and the noisy ($r = 0.828$, $p_{adjusted} < 0.001$) and the reduced ($r = 0.787$, $p_{adjusted} < 0.001$) condition was of a larger magnitude. The ABC no noise task yielded for both noise settings an r-value larger than 0.5 implying a large effect on the SNR. The difference between ABC - noise and ABC - reduced resulted in a small impact with a value of $r = 0.262$. For this pair it also held that $p_{adjusted} > 0.05$.

The difference between the two noisy task settings were not significant.

Math: Here the effect size was large with a value of 0.705. Additionally, $\chi^2 = 50.8$ and $p < 0.001$, thus requiring a post-hoc analysis.

The impact on the raw SNR between the task conditions and eyes open baseline yielded r-values larger than 0.5 suggesting a large impact of the tasks. All adjusted p-values were less than 0.001. Moreover, the noisy condition had a large ($r = 0.875$, $p_{adjusted} < 0.001$) impact. Similarly, there was a large difference between the two noise setting ($r = 0.875$, $p_{adjusted} < 0.001$). The difference between the no noise task and the reduced version resulted in a small difference of $r = 0.157$ (however $p_{adjusted} = 4.56e - 1$).

The difference between the noisy and reduced values was significant and large.

In addition to these results, it is visible in Figure 4.10, that the raw SNR of the baselines had the lowest median in all tasks. The second lowest median had the no noise condition across all tasks. In contrast to that, the median of unrestricted, noisy tasks such as typing and walking yielded the highest median raw SNR in their respective tasks. In the ABC task the highest median was calculated for the reduced recording. In the memory task there was no big visible difference between the median values of the noise settings.

The baselines have the lowest median SNR in all tasks.

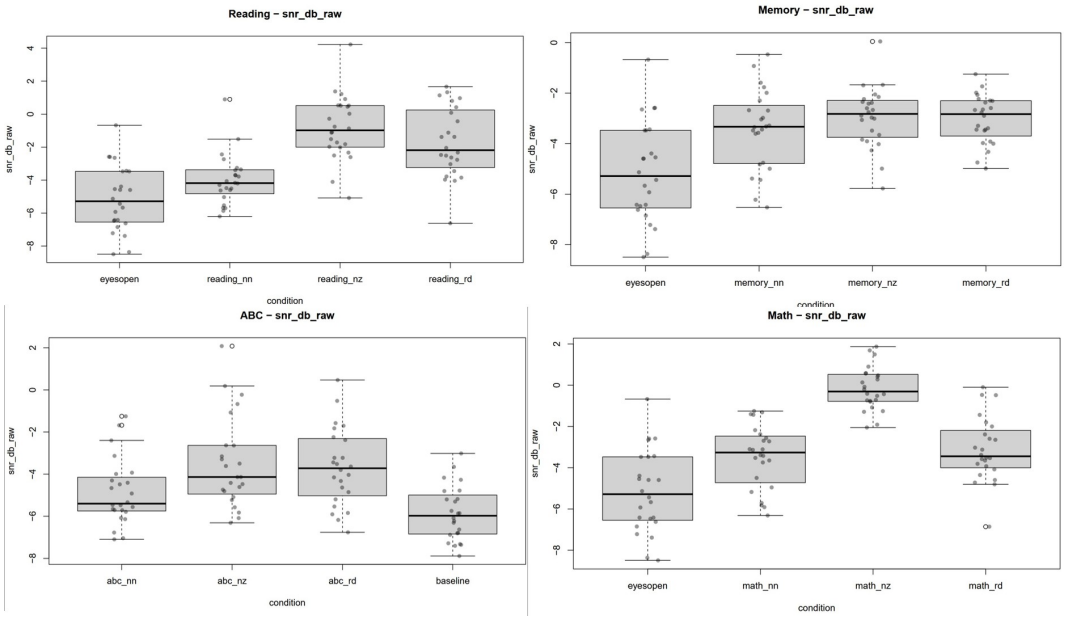


Figure 4.10: Collection of all boxplots according to the Friedman test for the raw SNR of every task. The order of the task boxplots is (left to right, top to bottom): reading, memory, ABC and math. The x- and y- axis for each task remain the same to their ealier describe values.

4.3.3 Questionnaire

In the following, the results of the questionnaire for the different conditions of the tasks are presented.

Q1 was generally disagreed upon in all tasks.

Q1 - I felt uncomfortable due to the headset: The median response was that participants disagreed to this statement for all conditions. However, a slight increase in agreement was observable in the conditions using physical restraints. For example, in the ABC - reduced task, six people agreed with the statement, whereas in the noisy version of this task only two people agreed. The same effect was also observed for the memory and reading tasks. Additionally, five participants pointed out explicitly, that particularly towards the end of the study they felt the muse getting more and more uncomfortable. One participant also pointed out that their "ear [...] hurts from the headset" (P24) due to the prolonged wearing time of the Muse.

Q2 - I felt too restricted in my movement: Typically, the answer to this question was "Disagree" for all settings not using noise restrictions excluding the mental math task. For the restricted memory and ABC task, there was a slight shift from disagreeing with the statement to the "Neutral" category. However, the most shift could be observed in the restricted reading setting. Here, the median of the answers was "Agree". This can be attributed to two main factors. Firstly, some participants pointed out that typing with the given keyboard was complicated by the keyboard that was provided which they were not used to. Especially one participant pointed out that they usually use a QWERTY layout as opposed to the provided QWERTZ keyboard. Secondly, the forehead support of the headrest that sticks out, also shown previously in Figure 4.2 (a), obstructed the view on the keyboard partially. Thus, it was reported by some that they could not see the keyboard as comfortably as they would have wanted to.
For the mental math task, participants strongly disagreed to Q2 when they were walking, and disagreed when they were told to stand still.

When the headrest blocked an essential part of the field of view, the physical restrictions were considered more restrictive.

Q3 - I felt comfortable executing the task: Participants agreed to this statement for all tasks not using physical restraints (setup 1). On average, participants felt more comfortable executing the math task when they were stationary, in contrast to walking around in the room. This can be seen by a slight increase of answers agreeing or strongly agreeing to the statement (15 when walking agreed, 16 when standing still) and a slightly bigger decrease in responses of disagreeing to strongly disagreeing. In total, seven participants responded negatively when walking, whereas only three disagreed or strongly disagreed when they were stationary. Additionally, one participant pointed out that they felt the walking around in the room to be distracting during their calculations.

Walking was perceived to be less comfortable than standing still.

In the other tasks using the noisy condition with no restrictions, participants agreed to Q3 overall. When using restriction setup 1, the comfort decreased slightly. Again, the decrease was bigger when participants had to use the keyboard and headrest simultaneously, with 13 participants answering negatively as opposed to none during the noisy reading task. In the memory and ABC task, the

Comfort decreased when using physical restraints.

increase in negative responses was smaller. During the ABC the disagreement rose from two negative responses to seven in the reduced setting, and from one negative response to six during the memory. Thus, there was a shift from agreeing in the non-restricted settings to overall neutrality when restriction setup 1 was used. Furthermore, it was observable that the discomfort during the ABC was generally slightly higher compared to the memory. One participant pointed out that they felt primarily uncomfortable due to the given task related to the ABC, and not as much due to the restrictions that were used. This should be considered as well, as a negative perception of a task may affect the overall perception of the restraints used.

Overall, participants agreed that they were using their full attention.

Q4 - I could execute the given task with my full attention: The median answer for all tasks and all settings was agreement. However, when considering the mean as well, it was observed that the agreement with this statement shrinks slightly with the introduction of physical restraints. This is important to note, as participants have pointed out not being able to view the keyboard comfortably which could impact the overall potential the tasks were executed with as well. Additionally, two participants reported that they found keeping their head and body still to be distracting from the original task they were supposed to work on.

Typing with the headrest felt the least natural to participants.

Q5 - My behaviour felt natural: This statement was generally agreed on for the unrestricted memory and reading settings. For the restricted reading tasks, participants generally disagreed to Q5. In addition to that, the overall sentiments for the rest of the tasks (i.e., both mental math and ABC conditions and the reduced memory) ended up being neutral. Other reasons, beside the physical restrictions used, can include comments of participants. Mainly comments such as that walking while doing mathematical calculations was distracting, keeping still for a certain amount of time was considered difficult and finding certain tasks uncomfortable were additional reasons for a less agreeing result.

There was mixed feedback for the headrest.

Additional comments: Generally, the headrest got mixed feedback, most of it being negative. This includes

that talking while having the head placed on the headrest was described as weird. Moreover, one participant "felt like Frankenstein [...] with no movement possible" (P07) while executing tasks with the headrest. Another participant pointed out, that the headrest at first was a bit confusing. Besides that, they pointed out that the headrest was "a bit uncomfortable but [they] quickly ignored it" (P11). It was also pointed out by another participant that they felt they moved their head when looking around on the screen, for instance. This happened even though they were using the headrest. However, one participant perceived the headrest in a more positive light, describing that it was "nice to have something hold my head" (P25).

The footrest was perceived positively.

In contrast to the headrest, the used footrest was generally seen in a positive light. Four participants in total mentioned the footrest positively, describing it as "comfortable" (P25) or "nice" (P27). In addition to that, we observed that during the studies, participants had to be reminded to take their feet off the footrest implying a certain level of comfort and it fading into an unconscious support.

Chapter 5

Discussion

This chapter returns to the three research questions. The findings are summed up under consideration of these questions. Then, the implications of these findings are covered. Lastly, this chapter considers possible limitations in this work that need to be taken into account in connection with the presented results.

5.1 Findings

RQ1 - What noise does typically occur in HCI user studies?

Talking and using computer hardware were commonly used movements.

The survey found that most commonly researchers have their participants talking in their studies. In addition to that, motions using computer hardware such as computer mice and keyboards were commonly needed. Controller usage and body gestures, such as hand or head movement and walking, were included as common motions as well. These movements can also be associated with the usage of AR and VR headsets [Kim et al., 2025].
We also found in our survey that a majority of researchers mostly use settings where a participant is sitting in front of a laptop or computer screen. The second most used setting according to our survey is the usage of VR-headsets.

The most common tasks we found were related to text comprehension and creation, as well as interactions in a VR environment. Furthermore, interviews were an additional task we found outside of our provided answer options.

RQ2 - How can we limit noise

The effects of the limitations cannot be evaluated accurately.

We utilized a combination of physical and verbal restraints in our study. These restraints had different effects on the data we recorded. However, the analysis of the data with the Friedman test showed contradictory results to common literature [Chong et al., 2023; Semmlow, 2018]. Noisy SNR values showed a median SNR value higher than for the recordings made under stationary conditions. This can be seen especially for the math task where for both the raw and the cleaned SNR the median value was the highest when walking (see Figure 4.10 and 4.9). Thus, considering that the computation of the SNR appeared to be unreliable, this question cannot be answered definitively.

If, in a revised analysis, restrictions result in higher quality data, the usage of the physical restriction may not be feasible.

Although there cannot be made any certain statements regarding the impact of the restrictions used in this study, one conclusion is that these restrictions generally come at the cost of a loss of comfort. Thus, if the data were to be evaluated again with a valid computation and yielded data of higher quality in comparison to the unrestricted conditions, these restrictions may not be feasible for typical HCI user studies.

RQ3 - To what extent can we allow noise?

Based on the data and analysis in this thesis, this question cannot be answered definitely.

As mentioned earlier, the computation of the SNR was not reliable enough to answer the question accurately. The Friedman test yielded a high median SNR value for both raw and cleaned SNR data of the walking math task in comparison to the other conditions, suggesting an inaccurate calculation for the SNR. Previous literature suggests walking should reduce signal quality as opposed to the here observed increase [Kyriaki et al., 2024].

5.2 Implications

The SNR computation may be inadequate for the collected data.

Overall, our data evaluation and analysis could not provide a clear and reliable answer to RQ2 and RQ3. This is due to the contradicting behavior of the median SNR values and the conditions they are associated with, for instance the walking math condition. The data, in connection with literature, suggests that the computation of the SNR was not adequate for evaluating data quality in movement heavy data recordings. A possible cause of this may be the definition of signal and noise during the computation. Signal was defined with the bandpower of 1 to 30 Hz. The rest of the data was considered noise. This appears to not be reliable as the bandpass filter targets a wider range of frequencies and may falsely disregard or include noise. When considering the math task, for instance, the signal included theta and beta. However, with the used bandpass filter delta and alpha frequencies are included as well.

5.3 Limitations

To further contextualize the presented findings, it is important to consider limitations and possible shortcomings of our approach.

Our preceded study has a limited amount of participants.

Firstly, our preceded survey found typical settings, tasks and motions in HCI user studies. Although a survey was conducted beforehand to make our final study design more representative of typical HCI user studies, the survey had a total of only 20 participants, all from German universities. This limits the overall informative value we could gain from the survey.

Omitting AR/VR-headsets made our study design less representative.

Secondly, the usage of AR and VR headsets, although a considerable number of researchers indicated the usage of such systems, was omitted. The exclusion was due to the incompatibility of a simultaneous use of such headsets and the Muse. However, this also means that the overall design of our study was less representative, as we excluded an important part of the survey.

The EEG we used is older and has fewer electrodes than other commercial-grade options.

The third limitation we encountered is, that we used

an older, commercial-grade EEG with only four dry electrodes. Having more electrodes to record, especially in highly noisy settings, can improve the quality of the data and better the effect of preprocessing methods like ICA [Downey and Ferris, 2023]. There are other consumer-grade headsets on the market besides the Muse that provide more electrodes. One such example is the Emotiv Epoc X headset by Emotiv[1] which provides 14 channels. Generally, it is recommended to have EEGs with more electrodes. Furthermore, the electrodes on the forehead of the Muse (AF7 and AF8) are especially susceptible to unconscious forehead movement such as frowning, possibly impacting the connection and data recordings [Fu et al., 2006]. Due to this reason we also had to exclude one recording, as unconscious frowning impacted the electrodes' contact to the skin.

The study is of a smaller scale.

Besides this, our study had 25 participants with only 24 datasets being evaluated. Lazar et al. [2017] recommend that for robust Chi-squared values a minimum of 20 participants should be recruited. Although we do have more samples than that, the number is only slightly above the recommendation (four samples more than recommended). This limits the reliability of total statistical effects that are observable.

The context we introduced talking as noise may be too unnatural.

During the ABC task, we introduced talking as a noise. However, with the adjustments we made participants would say individual letters with breaks in between. While this was intended to mimic speech during a mentally demanding task, the talking may be more cut-up and miss the natural flow during speech. Participants also only said the letters to themselves as opposed to an interactive conversation. This may have impacted the voice volume while speaking and lessened the effect of the movement.

5.3.1 Excursus - Alternative Data Analysis

The SNR was calculated manually for three examples.

As the data showed unexpected tendencies of the SNRs through the median values based on the Friedman test, an additional manual calculation of the SNR was done for one example task with three random samples. For this calcula-

[1] https://www.emotiv.com/de/, *last accessed September 30, 2026*

tion, the absolute power of the individual power bands was used. Since this calculation used absolute power bands the formula that was used was:

$$SNR_{dB} = 10log_{10}(\frac{\theta + \beta}{\frac{\alpha+\delta+\gamma}{3}})$$

$\theta + \beta$ described the signal, $\frac{\alpha+\delta+\gamma}{3}$ was used to describe the noise [Meigen and Bach, 1999]. As the mental math task was aimed at theta and beta, these two bands were considered signal, whereas the rest were considered to be noise. The values were computed using numpy. This formula was applied to three randomly chosen recordings for the mental math task. The values computed with the previous formula, also noted in Table 5.1, differed from the values computed with the program, seen in Figure 5.2.

	R-27	**R-25**	**R-02**	**Average**
Eyes Open	11.76	21.08	15.6	16.15
No Noise	14.45	19.03	14.43	15.97
Walking	14.02	21.98	14.19	16.73
Standing	14.64	27.47	16.16	19.42

Table 5.1: Computation of the SNR dB for the math task. Three samples were used as reference.

Manual recalculation barely showed an effect.

In two recordings (R-27 and R-02), a slight shift was observable. In the original computation of the program showed higher values for the walking condition in comparison to the less movement-hevy recordings. With the new computation the SNR of the walking condition sank to be the lowest SNR (without the eyes open baseline). However, this effect did not hold in R-25 where the SNR during walking still resulted in a value higher than the no noise version. On average, the highest SNR for the manual computation was the reduced condition, and the lowest the no noise version of the math task (see Figure 5.1). The provided SNR calculation yielded a higher average SNR for the walking condition and the lowest for the no noise and eyes open baseline.

A further evaluation regarding the definition and computation of noise and signal is needed.

This additional calculation suggests that further research and considerations need to be made while calculating the

	R-27	R-25	R-02	Average
Eyes Open	11.17	9.87	13.06	11.37
No Noise	16.68	12.27	10.42	13.12
Walking	16.59	11.92	11.18	13.23
Standing	15.28	11.51	12.65	13.15

Table 5.2: Computation from the program of the SNR dB for the math task. Three samples were used as reference.

SNR. A possible issue may lie with the definition of signal and noise in both computations. In our manual computation, noise was defined as the mean of the sum of all powers that were not targeted by the task. However, this definition of noise and signal, which in a similar sense was applied during the program calculation (only with a set bandpass filter not a customized one), may not be reliable to identify actual noise and signals [Suarez-Perez et al., 2018].

Chapter 6

Summary and Future Work

This chapter provides an overview of this thesis's contribution. Based on our observations we conclude this thesis with possible future work.

6.1 Summary and Contributions

We conducted an experimental study with various forms of movement.

In this thesis we aimed to investigate the robustness of EEG-frequency band measurements to movement-induced noise with a commercial-grade headset. We aimed to analyze the EEG in HCI relevant study settings. To investigate this, we preceded with a survey to determine typical user studies employed in HCI. Then, an experimental study with four different tasks under four different movement-intensities based on the survey was conducted. Both the cleaned and raw Signal-to-Noise Ratio (SNR) of our recorded data was evaluated with the Friedman test and pairwise Wilcoxon signed-rank method for possible post-hoc evaluation.

The computation of the SNR is flawed.

Based on the computed SNR values, we found that that unrestricted typing had a significant impact in both the raw and cleaned SNR. Mouse usage resulted in non-

significant differences between data quality in comparison to the task baseline. Talking was found to be not significant accross all conditions compared for the cleaned data. In the raw data, there were significant effects observed when introducing the noise. Lastly, walking had a significant impact on raw data quality compared to the reduced version. The impact was more significant in the raw data. However, we also found disagreements with the computed median SNR-values in this study to the typical interpretation and indication [Semmlow, 2018].

This thesis contributes EEG data recordings which can be used in future evaluation.

A contribution made in this thesis was the recording of EEG data with a consumer-grade EEG headset. We recorded, processed and evaluated 25 EEG data recordings in four different movement-intense conditions with regard to the Signal-to-Noise Ratio. Based on the evaluation, we found that the computation of the SNR appeared to be unreliable for this use and may benefit from further consideration.

The effect on comfort of physical restraints was still interpretable.

Although it was not possible to reliably assess the effect of physical restraints on data quality, we found that physical restraints that were aimed to reduce unconscious movement, had a negative impact on the comfort for study participants. In general, physically restricted settings appear to not be feasible for HCI user studies in regard to user comfort.
However, due to the limitations of the computation of the SNR, we cannot provide answers to RQ2 and RQ3.

6.2 Future Work

There are multiple possibilities for future work building up on the findings of our study. These will be presented in the following.

A larger scale study with a more diverse group of participants is recommended.

As this study had a limited number of participants the findings are limited by that as well. Thus, we recommend conducting a larger scale study that is more representative. This also applies to the demographics. Although we aimed to recruit a diverse group of participants, a majority were recruited through being acquaintances

from the same field of study. The median age also lay at 22.5 years with a standard deviation of 9.81 years in an age range from 19 to 59 year old participants.

Studies with commercial-grade EEGs providing more electrodes can be conducted.

Although this study focused on a consumer-grade EEG, the Muse has a limited number of electrodes in comparison to other available off-the-shelf, consumer-grade EEG headsets. A larger number of electrodes and possibly the usage of wet electrodes instead of dry electrodes could achieve a better quality for the recordings [Kleeva et al., 2024]. Conducting a study investigating consumer-grade EEGs that provide such hardware and comparing them to the recorded data from this study is worth considering.

Future studies with a focus on the CLI and ENG are suggested.

This study focused on the SNR values and the achievable overall quality. Future studies can also consider the evaluation of the CLI and ENG for certain tasks in either noisy or non-noisy settings. The evaluation can be done with a comparison of values recorded via an EEG and conventional cognitive load estimation methods. Moreover, EEG recordings are individual and differ greatly for every person. Investigating CLI and ENG under consideration of this is an advisable step before further considering to introduce EEGs as evaluation tool.

Individual definition of noise and signal per conditions should be provided.

Most importantly, there were indications of posssible issues with the program during evaluation of the recorded data. These concern especially the computation of the SNR. For one, the signal is not always definable with just one bandpass filter. A dual bandpass filter could be implemented. A single bandpass filter cannot account for filtering for two data frequencies that are not in a single, continuous range, for instance beta and theta. A dual bandpass filter can account for such limitations better. In the previous excursus, we tried to address this limitation. Furthermore, the SNR should be definable for each task individually. The tasks in our study were focused on different power bands. In the course of this, the computation of the SNR needs to be adjustable for each condition individually as the definition of signal and noise shift.

A reconsideration for the definition of noise and signal may be beneficial.

A general reconsideration of the definition of noise and signal may be required. The formulas used in this thesis defined certain frequencies as noise. In the program, frequencies outside a 1 to 30 Hz range were defined as

noise. In the manual check, alpha, gamma and delta were defined as noise. However, both of these calculations may falsely label frequencies as noise. This is the case especially for gamma power [Muthukumaraswamy, 2013]. A possible alternative approach to define noise may be the definition via ICA.

Including computed values in the saved data can ease the analysis.

While evaluating the data we also found, that the program used during the study saved only certain computed values. These did not include values such as the SNR which were the main measure this study aimed to evaluate. Generally, the SNR is an essential value to consider while analyzing EEG data, as it gives insight into the quality of data [Chong et al., 2023]. Thus, saving these values directly in addition to the recorded EEG samples can ease the evaluation greatly.

Appendix A

Survey Design

A printed version of our survey conducted on SoSciSurvey.
The survey consisted of three pages.

Page 01

A113

1. Informed Consent

Thank you for taking your time to answer this survey.
Before you participate, we need your informed consent. Please read the following information before continuing.

Purpose of Survey
The findings of this survey will form the basis for my study, which I will be conducting during my bachelor's thesis at RWTH Aachen University. In the study, we will be investigating how electroencephalograms (EEGs), a popular method for measuring brain activity, can be introduced as an additional information source for HCI studies.
EEGs are extremely noise-sensitive, which means that even the smallest movement of a participant will impact the recordings. We want to investigate how we can reduce such noise in typical HCI studies while imposing minimal restrictions on the participant.
To make our study design as accurate as possible, we are asking HCI researchers to share their experiences from their studies.

Principal Investigator: Wiebke Merle Neumeuer, wiebke.neumeuer@rwth-aachen.de, RWTH Aachen University

Target Audience
This study is aimed at researchers or students who have a background in HCI and have conducted at least one study in this field.
The study will take approximately 10 minutes.

Data Protection and Privacy:
Your response will be anonymous. We do not collect identifying data such as name, IP address, or other device information. The data will be stored securely on the SoSci Survey servers and is only accessible to the research team. The data will be used for scientific purposes only. Anonymized data will form the basis of and may be published in the final bachelor's thesis.
The data will be stored until 30th November 2026 and then deleted permanently.

Voluntariness and Right to Withdraw:
The participation is entirely voluntary. You may withdraw from the survey at any point without providing a reason and without facing negative consequences.
Please note that deleting your data after finishing the survey will not be possible due to the anonymization of the data.

For data-protection questions, participants may contact the Data Protection Officer of RWTH Aachen University:
Stabsstelle Datenschutz der RWTH Aachen University:
Templergraben 83, 52062 Aachen, Germany
RWTH Aachen University
E-Mail: dsb@rwth-aachen.de

Tel: +49 241 80 94114

If you have any further questions concerning this study please contact the principal investigator.

By clicking “I agree,” you confirm that you have read, understood, and agree to the information provided above.

◯ I agree

A110

2. What is your current ocupation?

- ◯ Bachelor’s student
- ◯ Master’s student
- ◯ Doctoral student
- ◯ Postdoctoral researcher
- ◯ Industry professional
- ◯ Other (please specify):

A111

3. Please select how many studies you have conducted in total:

- ◯ 1 – 2
- ◯ 3 – 5
- ◯ 6 – 10
- ◯ More than 10

A115

4. Please indicate your home university or institution.

Write “None” if not applicable.

A102

5. Please indicate your research interest(s):

Please write one research interest per line.

A106

6. Please select the settings that you typically use in your HCI study.

A typical study includes a participant ...

(You can select multiple answers.)

- [ ] 01. Sitting in front of a laptop/computer
- [ ] 02. Standing in front of a laptop/computer
- [ ] 03. Using a phone
- [ ] 04. Using pen and paper
- [ ] 05. Using an iPad or such with a pen
- [ ] 06. Wearing a VR headset
- [ ] 07. Standing in a room
- [ ] 08. Other (Please elaborate. If you have multiple ideas, please separate them with a semicolon):

A107

7. Please select the tasks that you typically give to participants in your study.

A participant may be asked to ...

(You can select multiple answers.)

- [ ] 01. Write a text
- [ ] 02. Draw something
- [ ] 03. Read a text
- [ ] 04. Interact with a touch screen
- [ ] 05. Listen to music
- [ ] 06. Play a video game
- [ ] 07. Interact with a VR Headset
- [ ] 08. Other (Please elaborate. If you have multiple ideas, please separate them with a semicolon):

A108

8. Please select motions that a participant usually uses in your studies.

A participant may be ...

(You can select multiple answers.)

- [ ] 01. Clicking a mouse
- [ ] 02. Moving a mouse
- [ ] 03. Drawing with a mouse
- [ ] 04. Typing on a keyboard
- [ ] 05. Using a pen to draw/write
- [ ] 06. Using a controller
- [ ] 07. Talking
- [ ] 08. Using hand gestures
- [ ] 09. Moving their head
- [ ] 10. Walking
- [ ] 11. Other (Please elaborate. If you have multiple ideas, please separate them with a semicolon):

Last Page

Thank you for participating in this survey!

Your answers were submitted, you may close the browser window or tab now.

Wiebke Neumeuer, RWTH Aachen – 2026

Appendix B

Study Questionnaire

The questionnaire used for the study including the study information, the consent form, the demographics and a form that was provided after the two noise conditions.

Study Information

Investigating the Robustness of EEG-Frequency Band Measurements to Movement-Induced Noise

Principal Investigator: Wiebke Merle Neumeuer
E-Mail: wiebke.neumeuer@rwth-aachen.de
Lehrstuhl Informatik i10, 52056 Aachen, Germany
RWTH Aachen University

Study Purpose: This study seeks to investigate how to minimize the noise in EEGs when they are used as additional information source in studies. The aim is to minimize noise through conditioning individuals while also ensuring as much comfort during the study as possible.

Procedure: During the study you will be asked to execute various tasks, such as typing texts etc., as instructed by your principal investigator. These tasks will be executed multiple times under varying circumstances. While doing so, you will wear the Muse headset which records your brain activity in the process. After each iteration you will get a questionnaire, in which you can report your level of comfort during the different iterations.
The study may be conducted in either English or German. After completing all the tasks, you may be asked to participate in a short interview about your thoughts and opinions.

Risks and Discomfort: There are no risks associated with the use of the Muse headband. If, however, you experience any discomfort, you may pause or terminate your participation at any time without any negative consequences.

Costs and Compensation: There are no costs associated with participating in this study. Snacks and drinks may be provided during breaks or after the session. No financial compensation will be offered.

Data Protection and Contact: All the data that is collected during this study is treated confidentially and is stored under an anonymous ID. Any publications or records will not contain identifying information. Video and audio recordings, if collected, will be used solely for analysis purposes and deleted after transcription and evaluation. For data-protection questions, participants may contact the Data Protection Officer of RWTH Aachen University:

Stabsstelle Datenschutz der RWTH Aachen University:

Templergraben 83, 52062 Aachen, Germany
RWTH Aachen University
E-Mail: dsb@rwth-aachen.de
Tel: +49 241 80 94114

Management

Katharina Saric
E-Mail: katharina.saric@zhv.rwth-aachen.de
Tel: +49 241 80-94114

Teilnehmerinformation

Untersuchung der Robustheit von EEG-Frequenzbandmessungen gegenüber bewegungsinduziertem Rauschen

Studienleiterin Wiebke Merle Neumeuer
E-Mail: wiebke.neumeuer@rwth-aachen.de
Lehrstuhl Informatik i10, 52056 Aachen, Germany
RWTH Aachen University

Zweck der Studie: Diese Studie soll untersuchen, wie sich das Rauschen durch Bewegungen in EEGs minimieren lässt, wenn diese als zusätzliche Informationsquelle in Studien verwendet werden. Ziel ist es, das Rauschen durch die Konditionierung der Proband*innen zu minimieren und gleichzeitig den größtmöglichen Komfort während der Studie zu gewährleisten.

Ablauf: Während der Studie werden Sie gebeten, verschiedene Aufgaben auszuführen, wie z. B. Texte zu tippen oder zu lesen, gemäß den Anweisungen Ihrer Studienleiterin. Diese Aufgaben werden mehrmals unter unterschiedlichen Bedingungen durchgeführt. Dabei tragen Sie das Muse-Headset, das Ihre Gehirnaktivität während des Vorgangs aufzeichnet. Nach jeder Durchführung erhalten Sie einen Fragebogen, in dem Sie Ihren Komfort während der verschiedenen Durchführungen bewerten können.
Die Studie kann entweder auf Englisch oder auf Deutsch durchgeführt werden. Nach Abschluss aller Aufgaben werden Sie möglicherweise gebeten, an einem kurzen Interview teilzunehmen, in welchem Sie weitere Anmerkungen und Meinungen äußern können.

Risiken und Unannehmlichkeiten: Mit der Verwendung des Muse-Stirnbandes sind keine Risiken verbunden. Sollten Sie jedoch Unannehmlichkeiten verspüren, können Sie Ihre Teilnahme an der Studie jederzeit ohne negative Konsequenzen unterbrechen oder beenden.

Kosten und Vergütung: Mit der Teilnahme an dieser Studie sind keine Kosten verbunden. Snacks und Getränke können während Pausen oder nach der Sitzung bereitgestellt werden. Eine finanzielle Vergütung wird nicht angeboten.

Datenschutz und Kontakt: Alle im Rahmen dieser Studie erhobenen Daten werden vertraulich behandelt und unter einer anonymisierten Kennung gespeichert. Veröffentlichungen oder Aufzeichnungen enthalten keine personenbezogenen Daten. Video- und Audioaufnahmen werden, sofern sie erstellt werden, ausschließlich zu Analysezwecken verwendet und nach der Transkription und Auswertung gelöscht. Bei Fragen zum Datenschutz können sich die Teilnehmer an den Datenschutzbeauftragten der RWTH Aachen wenden:

Stabsstelle Datenschutz der RWTH Aachen University

Templergraben 83, 52062 Aachen, Germany
RWTH Aachen University
E-Mail: dsb@rwth-aachen.de
Tel: +49 241 80 94114

Leitung:

Katharina Saric
E-Mail: katharina.saric@zhv.rwth-aachen.de
Tel: +49 241 80-94114

Consent Form

Investigating the Robustness of EEG-Frequency Band Measurements to Movement-Induced Noise

Principal Investigator: Wiebke Merle Neumeuer
wiebke.neumeuer@rwth-aachen.de
Lehrstuhl Informatik i10, 52056 Aachen, Germany
RWTH Aachen University

By signing this form, I confirm that:

- ☐ I have received and read the study information document in the language I selected (English or German) and fully understand its contents.
- ☐ I understand the purpose and procedure of the study, including that I will be asked to follow certain restrictions.
- ☐ I understand that my participation is voluntary, and that I may stop the experiment or withdraw from the study at any time, without giving a reason and without any negative consequences.
- ☐ I understand that all data collected during the study will be treated confidentially, that I will be identified only by an anonymous ID and that no identifying information will be included in any publications or reports.
- ☐ I have been informed of the responsible researcher(s) and know whom to contact if I have questions about the study or my rights as a participant.

Name – Participant	Signature – Participant	Date

	Signature - Investigator	Date

Einverständniserklärung

Untersuchung der Robustheit von EEG-Frequenzbandmessungen gegenüber bewegungsinduziertem Rauschen

Studienleitung: Wiebke Merle Neumeuer
wiebke.neumeuer@rwth-aachen.de
Lehrstuhl Informatik i10, 52056 Aachen, Germany
RWTH Aachen University

Mit meiner Unterschrift bestätige ich:

- ☐ Ich habe die Studieninformationsunterlagen in der von mir gewählten Sprache (Englisch oder Deutsch) erhalten und gelesen und habe deren Inhalt vollständig verstanden.
- ☐ Ich verstehe den Zweck und den Ablauf der Studie, einschließlich der Tatsache, dass ich bestimmte Einschränkungen einhalten muss.
- ☐ Ich verstehe, dass meine Teilnahme freiwillig ist und dass ich das Experiment jederzeit ohne Angabe von Gründen und ohne negative Folgen abbrechen oder aus der Studie ausscheiden kann.
- ☐ Ich verstehe, dass alle während der Studie erhobenen Daten vertraulich behandelt werden, dass ich nur durch eine anonyme ID identifiziert werde und dass keine identifizierenden Informationen in Veröffentlichungen oder Berichten enthalten sein werden.
- ☐ Ich wurde über die verantwortlichen Forscher informiert und weiß, an wen ich mich wenden kann, wenn ich Fragen zur Studie oder zu meinen Rechten als Teilnehmer habe.

Name – Teilnehmer*in	Unterschrift	Datum
	Unterschrift Studienleitung	Datum

Case: ID:

Questionnaire

Investigating the Robustness of EEG-Frequency Band Measurements to Movement-Induced Noise

Demographics

Age:

Gender:
- ☐ Female
- ☐ Male
- ☐ Non-Binary
- ☐ Prefer not to say

Occupation:

Highest education:

Case: ID:

Fragebogen

Untersuchung der Robustheit von EEG-Frequenzbandmessungen gegenüber bewegungsinduziertem Rauschen

Demografie

Alter:

Geschlecht:
- ☐ Weiblich
- ☐ Männlich
- ☐ Nicht-binär
- ☐ Keine Angabe

Beruf:

Höchste Bildung:

Case: ID:

Questionnaire

I felt uncomfortable due to the headset.

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

I felt too restricted in my movement.

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

I felt comfortable executing the tasks.

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

I could execute the given task with my full attention.

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

My behavior felt natural.

☐ Strongly agree ☐ Agree ☐ Neutral ☐ Disagree ☐ Strongly disagree

Further remarks that were not covered in previous questions:

Case: ID:

Umfrage

Ich habe mich aufgrund des Headsets unwohl gefühlt.

☐ Ja ☐ Eher ja ☐ Neutral ☐ Eher nein ☐ Nein

Ich habe mich in meiner Bewegung eingeschränkt gefühlt.

☐ Ja ☐ Eher ja ☐ Neutral ☐ Eher nein ☐ Nein

Ich habe mich beim Ausführen meiner Aufgabe wohl gefühlt.

☐ Ja ☐ Eher ja ☐ Neutral ☐ Eher nein ☐ Nein

Ich konnte meine Aufgabe mit voller Aufmerksamkeit bearbeiten.

☐ Ja ☐ Eher ja ☐ Neutral ☐ Eher nein ☐ Nein

Mein Verhalten hat sich natürlich angefühlt.

☐ Ja ☐ Eher ja ☐ Neutral ☐ Eher nein ☐ Nein

Weitere Anmerkungen:

Appendix C

Mental Math Tasks

Participants had either 3 seconds for tasks of the form X x XX or 7 seconds for tasks of the Form XX x XX. The tasks were as follows:

	No Noise	Noise	Reduced
1	8*17	7*19	9*14
2	6*14	8*15	7*17
3	8*16	7*18	9*16
4	15 * 19	12 * 19	18 * 19
5	6*17	9*17	8*14
6	14 * 18	13 * 16	17 * 19
7	9*13	6*18	7*14
8	21 * 24	21 * 25	21 * 23
9	12 * 13	14 * 19	14 * 15
10	16 * 12	12 * 15	12 * 18
11	22 * 23	16 * 18	13 * 15
12	18 * 13	22 * 24	12 * 14
13	14 * 16	15 * 18	22 * 25
14	17 * 18	16 * 19	16 * 17
15	12 * 17	15 * 17	17 * 14
16	15 * 16	13 * 19	14 * 13
17	23 * 25	23 * 24	24 * 25

Table C.1: The math tasks that each participant was given.

Appendix D

Final Study Layout

For the calculation, each task was assigned an offset (Reading - 0; Memory - 1; Math - 2; ABC - 3). The counting of the rows starts at zero in order for the formula to work.

	Task 1	**Task 2**	**Task 3**	**Task 4**
0	Reading - 1	Memory - 2	ABC - 4	Math - 3
1	Memory - 3	Math - 4	Reading - 2	ABC - 5
2	Math - 5	ABC - 6	Memory - 4	Reading - 3
3	ABC - 1	Reading - 4	Math - 6	Memory - 5
4	Reading - 5	Memory - 6	Math - 1	ABC - 2
5	Memory - 1	Math - 2	ABC - 3	Reading - 6
6	ABC - 4	Reading - 1	Memory - 2	Math - 3
7	Math - 4	ABC - 5	Reading - 2	Memory - 3
8	Reading - 3	Memory - 4	ABC - 6	Math - 5
9	Memory - 5	Math - 6	Reading - 4	ABC - 1
10	Math - 1	ABC - 2	Memory - 6	Reading - 5
11	ABC - 3	Reading - 6	Math - 2	Memory - 1

Table D.1: The table of the final calculation for all counterbalanced conditions.

Bibliography

[1] Pavlo Antonenko, Fred Paas, Roland Grabner, and Tamara Van Gog. Using electroencephalography to measure cognitive load. *Educational psychology review*, 22(4):425–438, 2010.

[2] Alyssa Arjoonsingh, Benjamin C Jamal, Latha Ganti, and Alyssa D Arjoonsingh. History and evolution of the electroencephalogram. *Cureus*, 16(8), 2024.

[3] Vanessa Julia Bangert. Design and Implementation of an Interactive Interface for EEG Data Analysis. Bachelor Thesis, RWTH Aachen University, 2026.

[4] Cédric Cannard, Helané Wahbeh, and Arnaud Delorme. Validating the wearable MUSE headset for EEG spectral analysis and Frontal Alpha Asymmetry. In *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 3603–3610. IEEE, 2021.

[5] Won-Du Chang. Electrooculograms for human–computer interaction: A review. *Sensors*, 19(12):2690, 2019.

[6] Samy Chikhi, Nadine Matton, and Sophie Blanchet. EEG power spectral measures of cognitive workload: A meta-analysis. *Psychophysiology*, 59(6):e14009, 2022.

[7] Samy Chikhi, Nadine Matton, and Sophie Blanchet. EEG power spectral measures of cognitive workload: A meta-analysis. *Psychophysiology*, 59(6):e14009, 2022.

[8] Shurui Chong, Xiangfei Hong, and Shanbao Tong. Comparing the Signal-to-Noise Ratio Estimates for Event-Related Potentials: A Simulation Study. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 1–4, 2023. doi.org/10.1109/EMBC40787.2023.10340887.

[9] Stefania Coelli, Roberta Sclocco, Riccardo Barbieri, Gianluigi Reni, Claudio Zucca, and Anna Maria Bianchi. EEG-based index for engagement level monitoring during sustained attention. In *2015 37th Annual International Conference*

of the IEEE Engineering in Medicine and Biology Society (EMBC), pages 1512–1515, 2015. doi.org/10.1109/EMBC.2015.7318658.

[10] Ryan J Downey and Daniel P Ferris. iCanClean removes motion, muscle, eye, and line-noise artifacts from phantom EEG. *Sensors*, 23(19):8214, 2023.

[11] Tisa Frelih, Andraž Matkovič, Tjaša Mlinarič, Jurij Bon, and Grega Repovš. Modulation of aperiodic EEG activity provides sensitive index of cognitive state changes during working memory task. *Elife*, 13:RP101071, 2025.

[12] Jérémy Frey, Christian Mühl, Fabien Lotte, and Martin Hachet. Review of the use of electroencephalography as an evaluation method for human-computer interaction. *arXiv preprint arXiv:1311.2222*, 2013.

[13] Michael J Fu, Janis J Daly, and M Cenk Cavusoglu. A detection scheme for frontalis and temporalis muscle EMG contamination of EEG data. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 4514–4518. IEEE, 2006.

[14] Zhanna Garakh, Yuliya Zaytseva, Alexandra Kapranova, Ondrej Fiala, Jiri Horacek, Alexander Shmukler, Isaac Ya Gurovich, and Valeria B Strelets. EEG correlates of a mental arithmetic task in patients with first episode schizophrenia and schizoaffective disorder. *Clinical Neurophysiology*, 126(11):2090–2098, 2015.

[15] Evgenia Gkintoni and Constantinos Halkiopoulos. Mapping EEG Metrics to Human Affective and Cognitive Models: An Interdisciplinary Scoping Review from a Cognitive Neuroscience Perspective. *Biomimetics*, 10(11), 2025. doi.org/10.3390/biomimetics10110730.

[16] Dasa Gorjan, Klaus Gramann, Kevin De Pauw, and Uros Marusic. Removal of movement-induced EEG artifacts: current state of the art and guidelines. *Journal of neural engineering*, 19(1):011004, 2022.

[17] Md Kafiul Islam, Amir Rastegarnia, and Zhi Yang. Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology*, 46(4-5):287–305, 2016.

[18] Leqi Jia. Low-Cost Wearable EEG in Non-medical HCI: A Systematic Review of Applications, ML Pipelines, and Deployment Challenges. In *International Conference on Human-Computer Interaction*, pages 127–144. Springer, 2026.

[19] Rozemun Khan, Johannes Vernooij, Daniela Salvatori, and Beerend P Hierck. Assessing cognitive load using EEG and eye-tracking in 3-D learning environments: A systematic review. *Multimodal Technologies and Interaction*, 9(9):99, 2025.

[20] Hyejin Kim, Sukwon Lee, and Changgu Kang. From controllers to multimodal input: A chronological review of XR interaction across device generations. *Sensors*, 26(1):196, 2025.

[21] Daria Kleeva, Ivan Ninenko, and Mikhail A. Lebedev. Resting-state EEG recorded with gel-based vs. consumer dry electrodes: spectral characteristics and across-device correlations. *Frontiers in Neuroscience*, Volume 18 - 2024, 2024. doi.org/10.3389/fnins.2024.1326139.

[22] Thomas Kosch, Jakob Karolus, Johannes Zagermann, Harald Reiterer, Albrecht Schmidt, and Paweł W. Woźniak. A Survey on Measuring Cognitive Workload in Human-Computer Interaction. *ACM Comput. Surv.*, 55(13s), July 2023. doi.org/10.1145/3582272.

[23] Olav Krigolson, Mathew Hammerstrom, Wande Abimbola, Robert Trska, Bruce Wright, Kent Hecker, and Gordon Binsted. Using Muse: Rapid Mobile Assessment of Brain Performance. *Frontiers in Neuroscience*, 15, 01 2021. doi.org/10.3389/fnins.2021.634147.

[24] Naveen Kumar and Jyoti Kumar. Measurement of cognitive load in HCI systems using EEG power spectrum: an experimental study. *Procedia Computer Science*, 84:70–78, 2016.

[25] Konstantina Kyriaki, Dimitrios Koukopoulos, and Christos A. Fidas. A Comprehensive Survey of EEG Preprocessing Methods for Cognitive Load Assessment. *IEEE Access*, 12:23466–23489, 2024. doi.org/10.1109/ACCESS.2024.3360328.

[26] Chi Qin Lai, Haidi Ibrahim, Mohd Zaid Abdullah, Jafri Malin Abdullah, Shahrel Azmin Suandi, and Azlinda Azman. Artifacts and noise removal for electroencephalogram (EEG): A literature review. In *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, pages 326–332. IEEE, 2018.

[27] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. Chapter 2 - Experimental research. In *Research Methods in Human Computer Interaction (Second Edition)*, pages 25–44. Morgan Kaufmann, Boston, second edition edition, 2017. doi.org/10.1016/B978-0-12-805390-4.00002-9.

[28] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. Chapter 4 - Statistical analysis. In *Research Methods in Human Computer Interaction (Second Edition)*, pages 71–104. Morgan Kaufmann, Boston, second edition edition, 2017. doi.org/10.1016/B978-0-12-805390-4.00004-2.

[29] 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*, pages 81–90, 2006.

[30] Siyu Long, Rui Ding, Junce Wang, Yue Yu, Jing Lu, and Dezhong Yao. Sleep quality and electroencephalogram delta power. *Frontiers in Neuroscience*, 15: 803507, 2021.

[31] Kelsie L Lopez, Alexa D Monachino, Santiago Morales, Stephanie C Leach, Maureen E Bowers, and Laurel J Gabard-Durnam. HAPPILEE: HAPPE In Low Electrode Electroencephalography, a standardized pre-processing software for lower density recordings. *NeuroImage*, 260:119390, 2022.

[32] Brendan D McKay and Ian M Wanless. On the number of Latin squares. *Annals of combinatorics*, 9(3):335–344, 2005.

[33] Thomas Meigen and Michael Bach. On the statistical significance of electrophysiological steady-state responses. *Documenta Ophthalmologica*, 98(3):207–232, 1999.

[34] Gernot R Müller-Putz, René Riedl, Selina C Wriessnegger, and others. Electroencephalography (EEG) as a research tool in the information systems discipline: Foundations, measurement, and applications. *Communications of the Association for Information Systems*, 37(1):46, 2015.

[35] Suresh D Muthukumaraswamy. High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. *Frontiers in human neuroscience*, 7:138, 2013.

[36] Tomoko Nemoto and David Beglar. Likert-scale questionnaires. In *JALT 2013 conference proceedings*, volume 108, pages 1–6, 2014.

[37] Jennifer J Newson and Tara C Thiagarajan. EEG frequency bands in psychiatric disorders: a review of resting state studies. *Frontiers in human neuroscience*, 12:521, 2019.

[38] Kingsley Okoye and Samira Hosseini. Wilcoxon statistics in R: Signed-rank test and rank-sum test. In *R Programming: Statistical data analysis in research*, pages 279–303. Springer, 2024.

[39] Robert Oostenveld and Peter Praamstra. The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, 112(4): 713–719, 2001. doi.org/10.1016/S1388-2457(00)00527-7.

[40] Yoselyn Nohemi Ortega-Gijon and Carmen Mezura-Godoy. Usability evaluation of brain computer interfaces: analysis of methods and tools. In *2019 IEEE International Conference on Engineering Veracruz (ICEV)*, volume 1, pages 1–8. IEEE, 2019.

[41] Yoselyn Nohemi Ortega-Gijon and Carmen Mezura-Godoy. Usability evaluation of brain computer interfaces: analysis of methods and tools. In *2019 IEEE*

International Conference on Engineering Veracruz (ICEV), volume 1, pages 1–8. IEEE, 2019.

[42] Srividya Pattisapu and Supratim Ray. Stimulus-induced narrow-band gamma oscillations in humans can be recorded using open-hardware low-cost EEG amplifier. *PloS one*, 18(1):e0279881, 2023.

[43] Weiwei Peng. EEG preprocessing and denoising. In *EEG signal processing and feature extraction*, pages 71–87. Springer, 2019.

[44] Dulce G Pereira, Anabela Afonso, and Fátima Melo Medeiros. Overview of Friedman's test and post-hoc analysis. *Communications in Statistics-Simulation and Computation*, 44(10):2636–2653, 2015.

[45] Alan T Pope, Edward H Bogart, and Debbie S Bartolome. Biocybernetic system evaluates indices of operator engagement in automated task. *Biological psychology*, 40(1-2):187–195, 1995.

[46] Felix Putze, Susanne Putze, Merle Sagehorn, Christopher Micek, and Erin T Solovey. Understanding hci practices and challenges of experiment reporting with brain signals: Towards reproducibility and reuse. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 29(4):1–43, 2022.

[47] Rajesh P.N. Rao. *Brain-Computer Interfacing: An Introduction*. Cambridge University Press, USA, 2013. ISBN 0521769418.

[48] Bujar Raufi and Luca Longo. An evaluation of the EEG alpha-to-theta and theta-to-alpha band ratios as indexes of mental workload. *Frontiers in Neuroinformatics*, 16:861967, 2022.

[49] René Riedl, Adriane B Randolph, Jan vom Brocke, Pierre-Majorique Léger, and Angelika Dimoka. The potential of neuroscience for human-computer interaction research. 2010.

[50] Sérgio Daniel Rodrigues and Pedro Miguel Rodrigues. Electroencephalogram-based time-frequency analysis for Alzheimer's disease detection using machine learning. *Journal of Biological Methods*, 12(1):e99010042, 2024.

[51] Luca Ronconi, Andrea Vitale, Alessandra Federici, Elisa Pini, Massimo Molteni, and Luca Casartelli. Altered neural oscillations and connectivity in the beta band underlie detail-oriented visual processing in autism. *NeuroImage: Clinical*, 28:102484, 2020.

[52] Aktham Sawan and Mohammad Awad. Assessing the Stress-Reducing Effects of Quranic Listening: An EEG-Based Classification Study. In *2023 24th International Arab Conference on Information Technology (ACIT)*, pages 1–9, 2023. doi.org/10.1109/ACIT58888.2023.10453846.

[53] Valentin Schwind, Stefan Resch, and Jessica Sehrt. The HCI user studies toolkit: Supporting study designing and planning for undergraduates and novice researchers in human-computer interaction. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–7, 2023.

[54] John L Semmlow. *Circuits, signals and Systems for Bioengineers*. Elsevier, 2018.

[55] Mahsa Soufineyestani, Dale Dowling, and Arshia Khan. Electroencephalography (EEG) technology applications and available devices. *Applied Sciences*, 10 (21):7453, 2020.

[56] Alex Suarez-Perez, Gemma Gabriel, Beatriz Rebollo, Xavi Illa, Anton Guimerà-Brunet, Javier Hernández-Ferrer, Maria Teresa Martínez, Rosa Villa, and Maria V. Sanchez-Vives. Quantification of Signal-to-Noise Ratio in Cerebral Cortex Recordings Using Flexible MEAs With Co-localized Platinum Black, Carbon Nanotubes, and Gold Electrodes. *Frontiers in Neuroscience*, Volume 12 - 2018, 2018. doi.org/10.3389/fnins.2018.00862.

[57] Nazmi Sofian Suhaimi, James Mountstephens, and Jason Teo. EEG-based emotion recognition: a state-of-the-art review of current trends and opportunities. *Computational intelligence and neuroscience*, 2020(1):8875426, 2020.

[58] Gabriella Tamburro, Pierpaolo Croce, Filippo Zappasodi, and Silvia Comani. Is Brain Dynamics Preserved in the EEG After Automated Artifact Removal? A Validation of the Fingerprint Method and the Automatic Removal of Cardiac Interference Approach Based on Microstate Analysis. *Frontiers in Neuroscience*, Volume 14 - 2020, 2021. doi.org/10.3389/fnins.2020.577160.

[59] Ronca Vincenzo, Cecchetti Marianna, Capotorto Rossella, Di Flumeri Gianluca, Giorgi Andrea, Germano Daniele, Borghini Gianluca, Babiloni Fabio, and Aricò Pietro. Beyond the lab: real-world benchmarking of wearable EEGs for passive brain-computer interfaces. *Brain Informatics*, 13(1):3, 2026.

Index

AF7 . . . . . 6, 50
AF8 . . . . . 6, 50
amplitude 12, 35
artifacts 2, 10

BCI . . . . . *see* Brain-Computer Interfaces
Brain-Computer Interface 2, 9

CLI . . . . . *see* Cognitive Load Index
Cognitive Load Index 8
counterbalancing 30

dB 12, 32, 35
decibles . . . . . *see* dB
dependent variable 26
DV . . . . . *see* dependent variable

electrode check 23
ENG . . . . . *see* Engagement Index
Engagement Index 8
ethics committee 22
Excursus 50

Friedman test 34
Future Work 54

Holm p-value correction 34

Implications 49
independent variable 26
IV . . . . . *see* independent variable

Latin square 30
likert scale 27
limitations 49

movement-induced 3, 11, 53

null-hypothesis 40

pairwise Wilcoxon signed-rank 34

Q1 . . . . . 27, 42
Q2 . . . . . 27, 43
Q3 . . . . . 27, 43, 44
Q4 27
Q5 . . . . . 27, 44

RMS 12, 35
root mean square . . . . . *see* RMS
RQ1 . . . . . 13, 47
RQ2 . . . . . 21, 48
RQ3 . . . . . 21, 48

Setup 1 25
Setup 2 25

TP10 6
TP9 6

within-group 30

Typeset September 30, 2026