Methodological considerations and advancements of Mobile Brain/Body Imaging data analysis

vorgelegt von

M. Sc. Marius Sebastian Klug

ORCID: 0000-0001-8667-3457

an der Fakultät V – Verkehrs- und Maschinensysteme
der Technischen Universität Berlin
zur Erlangung des akademischen Grades

Doktor der Naturwissenschaften
– Dr. rer. nat. –

Promotionsausschuss:

Vorsitzender: Prof. Dr. Markus Feufel
Gutachter: Prof. Dr. Klaus Gramann
Gutachter: Dr. Michael S. Cohen

Tag der wissenschaftlichen Aussprache: 11. November 2022

Berlin 2023

brain and body imaging, cyberpunk

– DALL•E 2

AI-generated image, https://openai.com/dall-e-2

It does not matter that we will never reach our ultimate goal. The effort yields its own rewards.

– Lt. Cmdr. Data

Echevarria, R. (Writer), & Frakes, J. (Director). (1990, March 12). The Offspring (Season 3, Episode 16) [TV series episode]. In R. Berman (Executive Producer), *Star Trek: The Next Generation*. Paramount Domestic Television.

It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.

– Sherlock Holmes

Doyle, A. C. (1999). *A scandal in Bohemia book/CD pack.* Longman.

I like to move it, move it. We like to – move it.

– King Julien

Darnell, E., & McGrath, T. (Directors). (2005). *Madagascar* [Film]. DreamWorks Animation, PDI/DreamWorks.

Contents

Summary

Recent technological advancements in both instrumentation and analysis methods of human brain imaging data such as electroencephalography (EEG) increasingly allow the measurement of mobile participants that interact with their environment. The new field of Mobile Brain/Body Imaging (MoBI; Gramann et al., 2011; Makeig, Gramann, Jung, Sejnowski, & Poizner, 2009) combines these measurements with imaging methods regarding the body, such as motion or eye tracking, and analyzes the multimodal data in order to investigate natural cognition in action. These analyses require the synchronized import of all data streams, options to process body data modalities, reliable preprocessing of EEG data in light of the elevated amount of non-cortical contributions in mobile settings, and the combined functional analysis of all modalities. Here, especially the sound preprocessing of EEG data from mobile settings suffers from a lack of information regarding a number of parameters that can be adjusted during the cleaning. Additionally, a comprehensive toolbox that addresses all four of these aspects is missing to date, but could be highly beneficial to the field. Two overarching goals were thus formulated in this dissertation: The first is to increase the reliability of MoBI data analysis, on the one hand by investigating the effect of different steps during the EEG processing, and on the other hand by standardizing the analysis of MoBI data. The second is to increase the usability of MoBI data analysis methods, focusing on the employment of easy-to-use and transparent automated processing tools.

To realize these goals, this dissertation presents two studies that investigate the parameters mobility, channel density, high-pass filter, and time-domain cleaning on their effect on the decomposition of EEG using independent component analysis (ICA). These studies lead to a set of best practices that can be employed when decomposing EEG data with ICA. Additionally, two automated toolboxes for the analysis of MoBI data are presented: The first, Zapline-plus, allows the removal of frequency specific artifacts from EEG data while minimizing the impact on non-artifactual elements of the data. The second, the BeMoBIL Pipeline, is a comprehensive pipeline for the analysis of MoBI data that addresses the four formulated requirements by making use of Zapline-plus and the information collected in the two studies, and augmenting them with a variety of wrapper functions that can be used automatically with minimal setup time. It emphasizes the use of robust methods to increase the reproducibility of the analysis, and provides documentations of all processing milestones and performed steps. The presented works are finally discussed with regards to their contributions to MoBI as a research method, in particular addressing the use of automation when processing MoBI data. Realizing the two formulated goals, this dissertation seeks to increase the applicability of MoBI in general by consolidating the use of mobile EEG and body data as a highly effective imaging method, and to increase the accessibility of MoBI as a tool for researchers from other fields.

Zusammenfassung

Jüngste Entwicklungen hinsichtlich der Analysemethoden für bildgebende Verfahren des menschlichen Gehirns, als auch der hierfür notwendigen Messinstrumente—beispielsweise Elektroenzephalographie (EEG)—ermöglichen zunehmend, Versuchspersonen in Bewegung und in direkter Interaktion mit ihrer Umwelt zu untersuchen. Das ebenfalls noch junge Feld des Mobile Brain/Body Imaging (MoBI; Gramann et al., 2011; Makeig, Gramann, Jung, Sejnowski, & Poizner, 2009) nutzt diese Messmethoden und kombiniert sie mit körperbezogenen bildgebende Verfahren wie Motion- und Eye Tracking, um anhand der analysierten Daten menschliche Kognition in Aktion untersuchen zu können. Die Voraussetzungen für die Analyse bilden dabei: der synchronisierte Import dieser Daten aus verschiedenen Quellen, Optionen zur Verarbeitung der verschiedenen körperbezogenen Daten, eine verlässliche Vorverarbeitung von EEG Daten unter Berücksichtigung eines erhöhten Anteils nicht-kortikaler Artefakte durch die Mobilität der Versuchspersonen sowie eine abschließende funktionale Analyse der Daten in ihrer Gesamtheit. In diesem Prozess lässt vor allem die Vorverarbeitung der EEG Daten Informationen darüber vermissen, wie bestimmte Parameter während der inkludierten Bereinigung der Daten angepasst werden sollten. Darüber hinaus mangelt es an einer umfassenden Toolbox, welche die vier angeführten Aspekte berücksichtigt. Ein solches Tool könnte jedoch einen relevanten Beitrag zum Forschungsfeld Mobile Brain/Body Imaging leisten.

Darauf basierend lassen sich für die vorliegende Dissertation zwei übergreifende Ziele formulieren: Zum einen soll eine erhöhte Reliabilität der MoBI Datenanalyse erreicht werden – einerseits durch die Untersuchung von möglichen Auswirkungen aller Schritte der EEG-Verarbeitung, andererseits durch eine Standardisierung ebendieser Analyse. Zum anderen soll die Anwendbarkeit der für die Analyse notwendigen Methoden erleichtert werden, wobei der Fokus auf einer leicht zu bedienenden, dabei jedoch gut nachvollziehbaren Automatisierung des Prozesses liegt. In Hinblick auf ebendiese Ziele werden in der vorliegenden Dissertation zwei Studien vorgestellt, welche die Parameter Mobilität, Kanaldichte, Hochpassfilter und Bereinigung in der Zeitdomäne hinsichtlich ihres Einflusses auf die Zerlegung der EEG Daten in ihre Quellen unter Verwendung von Independent Component Analysis (ICA) untersuchen. Aus diesen Studien werden eine Reihe von Best Practices abgeleitet, die für die Zerlegung von EEG Daten mittels ICA angewendet werden können. Darüber hinaus werden zwei automatisierte Toolboxen für die Analyse von MoBI Daten vorgestellt: Zapline-plus, welche das Entfernen frequenzspezifischer Artefakte aus EEG Daten bei minimaler Auswirkung auf die anderen, relevanten, Signale der Daten ermöglicht. Sowie die BeMoBIL-Pipeline, welche ein umfassendes Tool zur Analyse von MoBI Daten darstellt und die vier zuvor formulierten Anforderungen an den Prozess abdeckt, indem sie Zapline-plus und die Erkenntnisse zur Anwendung der ICA nutzt und die Vorverarbeitung von EEG Daten um eine Vielzahl von Wrapper-Funktionen erweitert. Die BeMoBIL-Pipeline kann nicht nur nach minimaler Einrichtungszeit vollautomatisch verwendet werden, sie setzt auch auf robuste Methoden, die die Reproduzierbarkeit der Datenanalyse erhöhen und bietet eine umfangreiche Dokumentation aller Verarbeitungsschritte und Meilensteine. Die vorgestellten Arbeiten werden abschließend in Hinblick auf ihren Beitrag zu MoBI als Forschungsmethode diskutiert, wobei besonders der Einsatz von Automatisierungen bei der Verarbeitung von MoBI-Daten fokussiert wird. Durch die Umsetzung der beiden zuvor formulierten Ziele wird im Rahmen dieser Dissertation dank der Konsolidierung der Verwendung von EEG- und körperbezogenen

Daten als hocheffektives Bildgebendes Verfahren sowohl die Anwendbarkeit von MoBI im Allgemeinen erhöht, als auch MoBI für Forschende aus anderen Feldern leichter zugänglich gemacht.

List of publications

Chapters 1 to 3 in this dissertation correspond to the following published or submitted manuscripts (chapter 2 includes two thematically connected studies):

Chapter 1:

Klug, M., & Kloosterman, N. A. (2022). Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG. *Human Brain Mapping*, 43(9), 2743–2758. https://doi.org/10.1002/hbm.25832.

Chapter 2:

Klug, M., & Gramann, K. (2021). Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *The European Journal of Neuroscience*, 54(12), 8406–8420. https://doi.org/10.1111/ejn.14992.

and

*Klug, M., *Berg, T., & Gramann, K. (submitted). No need for extensive artifact rejection for ICA - A multi-study evaluation on stationary and mobile EEG datasets.

Chapter 3:

Klug, M., Jeung, S., Wunderlich, A., Gehrke, L., Protzak, J., Djebbara, Z., Argubi-Wollesen, A., Wollesen, B., & Gramann, K. (submitted). The BeMoBIL Pipeline for automated analyses of multimodal mobile brain and body imaging data.

** authors contributed equally*

Acknowledgements

First of all, I want to express my deepest gratitude to Prof. Dr. Klaus Gramann, who was my supervisor in the last six years at TU Berlin. Klaus, throughout the time I was in your group, you always gave me both the freedom to find my own way and the guidance to do so without losing sight. Your inspiration and the work I was able to do at BPN means more to me than I can express—it gave me a goal in life and rekindled my enthusiasm. I could not have wished for a better Doktorvater.

Then, of course, there is the amazing team of colleagues at the BPN: Dr. Anna Wunderlich, Lukas Gehrke, and Dr. Janna Protzak, who accompanied me from the start and were always ready for great discussions about MoBI research or life in general, I will always remember the time we spent together. Friederike Hohlefeld, Sein Jeung, Chris Hilton, Sheng Wang, and Madeleine Fricke, who joined and became an enriching part of the team. Yiru Chen and Timo Berg, who were the backbone of the lab and the studies we conducted, together with Benjamin Paulisch and Reinhard Gilles, our IT and infrastructure masterminds, and Carolin Engbertz, who brought order into our lives and lifted great weights from our shoulders. The many guests that visited us and shared their insights, making the group the international hub it is, all in front Dr. Zakaria Djebbara, who never lost against me in foosball. And the many students who came and went and always brought a breath of fresh air into the mix. This team was a true inspiration to my life and I can only wish any doctoral student to experience such a level of solidarity and team spirit.

I would also like to thank Prof. Dr. Thorsten Zander, who inspired me greatly with his view on BCI, Dr. Laurens Krol, and Lena Andreessen for enriching my life in Berlin with great conversations and fun times.

My gratitude also goes to Prof. Dr. Mike X. Cohen, who kindly agreed to assess this work. His book *Analyzing Neural Time Series Data: Theory and Practice* and teachings on Udemy have greatly enhanced my understanding of the methdology of EEG analysis, and my own university course has benefited a lot from his permission to use parts of his MATLAB lecture available online.

I thank my family for always being supportive and just great all around. My mother, Carmen Klug, took me on adventures of the mind, and my father, Jörg Wolfinger, showed me the awe and beauty of nature. I thank them both for my life. And Roswitha Herbst, who became another parent to me, persistently taught me the importance of persistence, which also made this work possible. My brothers Fabian Klug and Kilian Eder have been, and always will be there, which I thoroughly enjoy.

My friends from Tübingen are missed dearly here in Berlin, and I especially thank Johannes Stein for accompanying me for almost half my life. Victor Kossak gave me a haven in the big city life for which I will always be grateful.

Last, but most assuredly not least, I want to thank my fiancée Laura Grimm. Laura, your presence in my life over the last years gave me so much it is impossible to overstate. You are my sparring partner in all sorts of conversations, my source of energy, inspiration, emotion, reflection, fun, and comfort. You have my back, and I have yours. I love you.

Introduction

What goes on in our minds when we do something? This semantically simple question has been debated for centuries by researchers and thinkers from various fields, yet it still defies a comprehensive answer. Everyone can have a first attempt by turning their attention inwards and learning to observe the workings of their mind, thus finding their own, subjective, experiences. But as tempting and intuitive as this approach may seem at first glance, its weaknesses quickly reveal themselves: First and foremost is the lack of objectivity, as any answer obtained in such a manner cannot be verified by other observers. In their attempts to gather objectively measurable insights about the human (and animal) psyche, researchers of the early twentieth century then turned towards not dealing with the mind at all, focusing exclusively on the "doing something" in their behaviorist approach (Skinner, 2011; Watson, 1913). It was only with the end of behaviorism and the advent of cognitive science in the 1960's and 70's (e.g. Neisser, 2014, first published in 1967) that scientists began to look inward again. They started to investigate not only human behavior but also, and perhaps even more so, the cognition that drives it—thus shifting the focus back towards the mind. This time, however, they did not rely on subjective measures and introspection, but wanted to investigate the substrate that provides the cognitive abilities, the brain, in the assumption that insights about the brain would equate to insights about the mind.

To this end, neuroscientists used a new array of technologies that image the human brain non-invasively. These methods allowed them, for example, to examine anatomical structures of the brain (magnetic resonance imaging, MRI), measure metabolic changes that go hand in hand with solving cognitive tasks (functional MRI, fMRI, and positron emission tomography, PET), or electrical or magnetic field changes stemming from the synchronized firing of neurons in the brain (electroencephalography, EEG, and magnetoencephalography, MEG). Neuroscience thus built a solid foundation that showed the brain's reaction to different stimuli or the brain's functioning during cognitive processes. Yet, while this foundation has greatly advanced scientists' understanding of cognition, it still cannot answer the initially posed question satisfactorily, as now behavior ("doing something") was largely omitted from these studies. This had two main reasons: On the one hand, it was necessary first to have a basis to stand on before venturing deeper into the complexities of dynamic interaction. On the other hand, the used experimental apparatuses typically restrict participants' movements, either due to the size and weight of the devices or in fear of artifacts in the acquired data. Yet, answering the initially posed question by imaging the human in motion, both the brain and the body, can be considered of utmost importance if one strives to understand the human mind holistically, as the idea of *embodied cognition* (further introduced below) suggests.

The field of Mobile Brain/Body Imaging (MoBI—discussed in depth below) was introduced about a decade ago (Gramann et al., 2011; Makeig et al., 2009) and presents itself as a possible means to this undertaking. MoBI studies make use of recent advancements in brain imaging technology that allow the collection of data from participants in motion, focusing mostly on EEG. This brain imaging data is then analyzed in conjunction with additionally acquired body imaging data such as motion capture or eye tracking. While MoBI stands on firm grounds and has already led to a number of new and important insights considering the human mind and body, it is still in an early stage and faces methodological challenges during data analysis, as discussed further below. These challenges can be summarized as, on the one hand, a lack of in-

formation on the specific needs of mobile EEG analysis, and, on the other hand, a reliance on time-consuming manual data processing and lack of tools for the combined MoBI analysis. The latter is in parts a consequence of the former, as reliable methods for automated EEG processing were largely unavailable and manual cleaning was thus regarded as the safer option (Cohen, 2017a, Gramann, personal communication), but especially regarding the processing of body data, MoBI was simply too young to provide comprehensive options. As these methodological challenges can hinder the wide-spread use of MoBI as a research method, this dissertation is thus focused on two main goals: The first is to increase the reliability of MoBI data analysis, on the one hand by investigating the effect of different steps during the EEG processing, and on the other hand by standardizing the analysis of MoBI data. The second is to increase the usability of MoBI data analysis methods, focusing on the employment of easy-to-use and transparent automated processing tools. In doing so, it seeks to increase the applicability of MoBI in general by consolidating the use of mobile EEG and body data as a highly effective imaging method, and to increase the accessibility of MoBI as a tool for researchers from other fields. These goals are realized by providing a toolbox to remove frequency-specific artifacts from EEG data (chapter 1), by presenting two studies that investigate important parameters during the cleaning of MoBI data using independent component analysis (ICA, chapter 2), and by providing a toolbox for the comprehensive automated processing of MoBI data including both EEG and body dynamics that can be used with minimal setup time and provides important feedback to keep the automation transparent (the BeMoBIL Pipeline, chapter 3).

Embodied Cognition

As embodied cognition can be considered to lie at the heart of MoBI, the following section will give a brief overview of the origins and implications of this idea. As a foundation, the so-called mind–body–problem is concerned with questions about the interaction between the mind (the subjective experience) and the body (the matter of the physical world) and whether they are essentially the same or entirely different entities (Revonsuo, 2010). The *Hard Problem* in this question is the fact that science and philosophy currently have no clear idea on how physical systems could possibly produce subjective experience, although bridging this *Explanatory Gap* is attempted with many different approaches[1].

One answer to this question has been offered by the concept of embodiment, or embodied cognition, which is the idea that consciousness—or the mind—is not born from the abstract mental thought only, but lies explicitly within the interaction of a moving body and its surroundings. It is this having a body that constitutes the crucial feature of consciousness and the existence of an *I* that has the capacity to experience (Blanke & Metzinger, 2009). Following the notion of embodiment will thus result in the realization that if one strives to understand consciousness, one must not search for individual neural correlates of it but instead should focus on larger associations of neural dynamics in conjunction with the body and sensorimotor processes (Thompson & Varela, 2001). In this view, consciousness and the subjective experience are not located within the brain only, but "cut across the brain–body–world divisions" (Thompson & Varela, 2001, p. 425)[2]. However, although principally rooted in philosophy, cognitive science is not primarily concerned with consciousness. From a practical perspective,

[1] For a more comprehensive discussion of these topics, see the book *Consciousness—The Science of Subjectivity* by Antti Revonsuo (2010)

[2] This view is not always met with praise within the field of philosophy of the mind, see Revonsuo (2010), p. 194.

cognition—independently of the first-person subjective experience—can be dependent on the body and one's surroundings, as summarized by Margaret Wilson:

> (1) cognition is situated; (2) cognition is time-pressured; (3) we off-load cognitive work onto the environment; (4) the environment is part of the cognitive system; (5) cognition is for action; (6) off-line cognition is body based (M. Wilson, 2002, p. 625).

Rephrasing this concept as *situated cognition*, this has later been condensed to the three aspects of the embodied, embedded, and extended mind (Robbins & Aydede, 2008). Here, "embodied" refers to psychological findings in which, for example, different bodily actions or states affected the participant (e.g. Cacioppo, Priester, & Berntson, 1993; Chen & Bargh, 1999), "embedded" means the off-loading of workload and cognitive effort onto the environment (e.g. Kirsh, 1995), and "extended" refers to the dynamical system of the human and its environment that only in combination constitutes the full cognitive capacity (which, as discussed above, might even go so far as to describe the entire mind as being "within the system", Clark & Chalmers, 1998).

While being focused on general psychological research and theories at first, this view has major implications on neuroscientific studies as well. If embodied or situated cognition is taken as the foundational model of the human mind, the classic neuroscientific experimental paradigms describing a linear development of stimulus–processing–response are not ideal when one wants to understand natural human behavior and cognition. Of course, processing a stimulus can indeed lead to an output, which in classic experiments is often the press of a button as a (literally) digitized decision. A button press, however, is hardly comparable to the outputs we generate as we live our daily life, and, importantly, the proprioceptive and kinesthetic sensory inputs generated from movement are essential to our understanding of the world (Buzsaki, 2006; Buzsáki, 2019; Feinberg, 1978). It is this indefinite cycle of perception and action that allows us to exist in, and make sense of the world by permanently actively infering its state (Parr, Pezzulo, & Friston, 2022).

In summary, generalizing from traditional stationary brain imaging experiments to applicable insights about the human mind and behavior can be problematic. Nevertheless, until recently, studies of the human brain in its natural habitat, namely a freely moving participant in active interaction and engagement with their environment, were as rare as they were necessary.

Mobile Brain/Body Imaging

A new paradigm for experiments to investigate cognitive functioning in human subjects was needed. This was presented as the field of Mobile Brain/Body Imaging (Gramann, Ferris, Gwin, & Makeig, 2014; Gramann et al., 2011; Makeig et al., 2009), defined as "a method to record and analyze brain dynamics and motor behavior under naturalistic conditions" (Jungnickel, Gehrke, Klug, & Gramann, 2019, p.59). In this new paradigm, the dynamics of natural embodied cognition are the core feature of interest, and these dynamics are investigated in human participants that actively move in, as well as interact and engage with the world around them. These investigations include measurements of the brain, such as mobile EEG, in combined analysis with measurements of the body, such as motion or eye tracking, as this contextual information is the precise element that sheds light on the embodied, embedded, and extended mind. This approach to human imaging enables cognitive science to look beyond the

horizon of stationary experiments and take a more action-orientated stance to bridge the gap between traditional brain imaging modalities and our everyday lives (Engel, Maye, Kurthen, & König, 2013; Ladouce, Donaldson, Dudchenko, & Ietswaart, 2017). MoBI is a method not limited to one research field, and its approaches have been applied in various directions. To lay the groundwork for mobile EEG applications, classic paradigms such as the oddball task were replicated in moving participants, measuring visually (Gramann, Gwin, Bigdely-Shamlo, Ferris, & Makeig, 2010) or auditory (Debener, Minow, Emkes, Gandras, & de Vos, 2012) evoked brain responses. This established a foundation showing that the core idea of MoBI was feasible.

As physically navigating spaces became possible in mobile experiments, spatial cognition is a core interest in the MoBI community, for example by measuring EEG and movements of participants exploring a virtual room (Snider, Plank, Lynch, Halgren, & Poizner, 2013) or hippocampal local field potentials during navigation (Bohbot, Copara, Gotman, & Ekstrom, 2017). Other studies investigated, for example, the exploration of an invisible virtual maze (Gehrke, Iversen, Makeig, & Gramann, 2018), how physical and virtual rotations lead to different electrocortical responses (Gramann, Hohlefeld, Gehrke, & Klug, 2021), or how instructions affect memory when navigating through a city (Wunderlich & Gramann, 2021).

But while the act of orientating oneself and navigating the environment is undoubtedly complex, the mere ability to walk is no triviality either. Walking as such presents a task for the brain to solve, and this task is increasingly studied, for example, by investigating gait specifically (e.g. Castermans, Duvinage, Cheron, & Dutoit, 2014; Gwin, Gramann, Makeig, & Ferris, 2010; Nordin, Hairston, & Ferris, 2019; Seeber, Scherer, Wagner, Solis-Escalante, & Müller-Putz, 2015; Wagner, Makeig, Gola, Neuper, & Muller-Putz, 2016; Wagner, Martínez-Cancino, & Makeig, 2019; Wagner et al., 2012) or with the use of dual task paradigms (e.g. De Sanctis, Butler, Malcolm, & Foxe, 2014; Nenna, Do, Protzak, & Gramann, 2021; Protzak, Wiczorek, & Gramann, 2021; Reiser, Wascher, & Arnau, 2019).

Another important application of the MoBI approach is in neuroergonomics, the study of the brain at work and in everyday life (Ayaz & Dehais, 2018; Mehta & Parasuraman, 2013; Parasuraman & Rizzo, 2007). Although studies of workload, attention, and engagement are often conducted in seated settings (e.g. Gerjets, Walter, Rosenstiel, Bogdan, & Zander, 2014; Gevins & Smith, 2003; Pope, Bogart, & Bartolome, 1995; G. F. Wilson & Russell, 2007), they can include eye gaze (e.g. Matthews, Reinerman-Jones, Barber, & Abich, 2015), be applied in real-world tasks or realistic settings (e.g. Aricò et al., 2016; Miklody, Moessmer, Dettmann, Klinkenberg, & Blankertz, 2017), and also in real life scenarios such as driving (e.g. Kohlmorgen et al., 2007; Zander et al., 2017), navigation (e.g. McKendrick et al., 2016), air traffic control (e.g. Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014; Sebastiani et al., 2020; Yeo et al., 2017), or even real flights (e.g. Dehais et al., 2019; Dehais, Roy, Durantin, Gateau, & Callan, 2018; Gateau, Ayaz, & Dehais, 2018).

Last, MoBI attracts interest also from fields not traditionally concerned with neuroscientific investigations, such as architecture (Banaei, Hatami, Yazdanfar, & Gramann, 2017; Djebbara, Fich, Petrini, & Gramann, 2019) or dance (Barnstaple, Protzak, DeSouza, & Gramann, 2021).

Instrumentation of MoBI measurements

In order to enable the comprehensive measurement of mobile brain and body dynamics, several requirements have to be met (Gramann et al., 2014, 2011; Jungnickel et

al., 2019; Makeig et al., 2009). First and foremost, as the participants will move in and interact with the environment, the imaging method of the brain has to be mobile, resistant to movement artifacts, and fast enough to capture the brain dynamics supporting behavior and natural cognition. Second, the tracking of body measures such as motion and eye gaze data is a natural necessity to provide the contextual body information for the analysis of the acquired brain imaging data. Third, even though MoBI experiments are directed towards the investigation of natural cognition, in order to provide reliable and repeatable experimental paradigms it may still be relevant to have control over the participant stimulation. Fourth and last, for a meaningful interpretation, the temporal context of the different data modalities needs to be preserved. The following section will give an overview of the challenges and solutions to each of these requirements.

Mobile brain imaging excludes methods that rely on heavy machinery, such as fMRI or PET scanners, as these are immobile and require the participants to lie supine. Additionally, these methods measure hemodynamic responses, which means that their temporal resolution is too low to cover the small, sudden, and highly dynamic changes in brain activity that are required for motor control or real-world interactions (Gramann et al., 2014; Mehta & Parasuraman, 2013). MEG scanners in turn allow for a sitting posture during the measurement and can include arm reaching movements (Waldert et al., 2008). Efforts have also been made to increase their mobility (Boto et al., 2018) and the high temporal resolution (Mehta & Parasuraman, 2013) would make MEG a viable choice for MoBI measurements in principle. Yet, their size and weight renders them unsuitable for MoBI without further technological advancements.

In contrast to these large and heavy devices, EEG and functional near-infrared spectroscopy (fNIRS) present viable techniques of measuring mobile participants, as both are lightweight enough to be carried but allow functional interpretations (Gramann et al., 2014; Mehta & Parasuraman, 2013). Similar to fMRI, fNIRS also measures hemodynamic activity, but does so using near-infrared light that traverses the brain (Curtin & Ayaz, 2018). Hence, although fNIRS devices allow for ambulatory measurements of brain activity (e.g. McKendrick et al., 2016), they are still limited by the low temporal resolution of the hemodynamic response, leaving EEG as the main brain imaging method in MoBI.

EEG uses electrodes to measure the electrical field changes on the scalp of the participants that are the result of large populations of neurons firing in the neocortex, specifically, aligned pyramidal cells that generate transient dipoles within the brain (Cohen, 2017b). These miniscule signals require amplification to be recorded, and advancements in this technology resulted in lightweight and portable options such as the BrainProducts LiveAmp[3], the mBrainTrain Smarting[4], or the ANT eego sports system[5]. These new devices also allow the recording of EEG with the high channel density that is required for the analysis of the data on the source level as discussed further below, thus increasing the viability of EEG as a brain imaging method. Additionally, EEG has a high temporal resolution and can be measured with sampling rates of up to several kilohertz, making it the ideal candidate for MoBI research (e.g. Gramann et al., 2014; Jungnickel & Gramann, 2016; Mehta & Parasuraman, 2013). An overview of addressing the challenges of movement in EEG experiments will be given in the section below.

A wide array of sensor options is available to measure the body: Most notably mo-

[3] https://www.brainproducts.com/solutions/liveamp
[4] https://mbraintrain.com/smarting-pro-line
[5] https://www.ant-neuro.com/products/eego_sports

tion capture, for example using optical systems such as Vicon[6] or PhaseSpace[7]. These optical systems are stationary but allow unrestrained movement in the range of several meters. While this enables a wide range of MoBI studies, investigating farther mobility requires other systems based on inertial measurement units (IMUs) such as the XSens[8]. The second important body measure is eye tracking, the measure of gaze direction (which can be analyzed both in terms of input and output of the participant) and pupil size (e.g. as a proxy of workload or stress; Mitz, Chacko, Putnam, Rudebeck, & Murray, 2017; Wahn, Ferris, Hairston, & König, 2016). Developments in this technology led to a small and lightweight form factor and eye trackers can now be integrated in glasses such as the Tobii Pro[9] or the Pupil Labs Invisible[10].

While MoBI experiments are focused on naturalistic behavior, they can still require presenting the participant with stimuli, and recording button presses (e.g. when investigating oddball paradigms; Debener et al., 2012; Ladouce, Donaldson, Dudchenko, & Ietswaart, 2019). Stimuli can be presented auditorily (e.g. Wunderlich & Gramann, 2021) or visually (e.g. Protzak et al., 2021) and participant responses can be recorded, for example, by giving verbal feedback (e.g. McKendrick et al., 2016) or using wireless computer mouses or controllers (e.g. Protzak et al., 2021). Recently, Virtual Reality (VR), the immersive presentation of a virtual environment is increasingly used in MoBI. In VR, the user wears a head-mounted display (HMD) that displays a 3D environment and is tracked constantly in order to update the virtual view in congruence with the physical movement, often in combination with motion controllers or additionally tracked body parts[11]. VR is a useful tool as participants can be mobile and highly immersed in a virtual world that allows control over the elements presented, including visual, auditory, haptic, and other stimulation (e.g. Gehrke, Akman, et al., 2019), thus bridging the gap between a strictly controlled seated study and research "in the wild" (e.g. Wunderlich & Gramann, 2021). In addition, VR includes motion capture and some devices even include eye tracking (e.g. the HTC Vive Pro Eye[12], or the HP Reverb G2 Omnicept[13]) or heart rate sensors (e.g. the HP Reverb G2 Omnicept).

With the large variety of data modalities in MoBI, preserving the temporal context by keeping the resultant recorded data synchronized is essential for any analysis (Gramann et al., 2014, 2011; Jungnickel et al., 2019). While other options such as trigger pulses or photodiodes exist, Lab Streaming Layer (LSL)[14] is an excellent option for MoBI. It is an open-source data streaming protocol that allows the unified storage of any kind of time-series data, even when recorded on a different computer. This is achieved by adding a timestamp to each sample as soon as it is acquired and before it is sent. The final recording of all streams is then stored in the Extensible Data Format (XDF) as a single file and includes metadata such as the sampling rates and channel names of each modality. LSL is increasingly supported by many manufacturers, including a variety of devices such as EEG, fNIRS, eye tracking, motion capture, and human interface devices[15].

[6] https://www.vicon.com/hardware

[7] https://www.phasespace.com

[8] https://www.xsens.com/motion-capture

[9] https://www.tobii.com/products/eye-trackers/wearables/tobii-pro-glasses-3

[10] https://pupil-labs.com/products/invisible/

[11] Other systems can use a room in which the virtual environment is projected to the walls (Cruz-Neira, Sandin, DeFanti, Kenyon, & Hart, 1992).

[12] https://www.vive.com/de/product/vive-pro-eye/overview

[13] https://www.hp.com/de-de/vr/reverb-g2-vr-headset-omnicept-edition.html

[14] https://github.com/sccn/labstreaminglayer

[15] https://labstreaminglayer.readthedocs.io/info/supported_devices.html

Analysis of MoBI measurements

Even with the correct instrumentation in place, the analysis of MoBI data is a complex task. As the studies presented in this dissertation are concerned only with the analysis of EEG data as the brain imaging modality, the analysis of fNIRS will not be discussed. The processing and analysis of MoBI data can be organized into the following aspects (Gramann et al., 2014, 2011; Jungnickel et al., 2019; Makeig et al., 2009): i) ensuring and maintaining a synchronized dataset, ii) preprocessing and analysis of body data to provide contextual information, iii) robust preprocessing of EEG data including the separation of cortical signals from other contributions, iv) the generation of the final brain measures in light of the body dynamics and presented stimuli[16]. Each of these points will be introduced in the following section as these lay the foundation for the subsequent chapters.

First of all, the recorded data needs to be imported with preserved temporal information. The different streams can have different and inconsistent sampling rates, yet they need to be in an identical temporal structure for their final analysis to provide the respective context. One solution to this is the MOBILAB toolbox (Ojeda, Bigdely-Shamlo, & Makeig, 2014), a plugin to EEGLAB (Delorme & Makeig, 2004; Delorme et al., 2011) that allows the import and visual inspection of multimodal data. It also includes body signal processing options such as computing derivatives and extracting events from motion data. When finally exporting the data to EEGLAB for further analysis, it combines all streams of an XDF file into a single set by cutting and interpolating them where necessary. MOBILAB is to date the main option to address aspects i) and ii) of the MoBI data processing.

With an imported synchronized dataset, preprocessing and cleaning the EEG data is essential to MoBI data analysis (e.g. Cohen, 2014, 2017a; Gramann et al., 2014; Makeig et al., 2009). A key issue here is that EEG does not only measure brain activity, but picks up any electrical signal at the electrodes. Thus, recordings on the scalp carry information from different sources in the head and the environment. Here, electrical current source signals stemming from the synchronized firing of neurons are transmitted to the scalp through volume conduction, thus spreading out and leading to a mix of signals of different brain regions that are recorded at the electrodes on the scalp. Additionally, muscle activity also generates electrical signals (Raez, Hussain, & Mohd-Yasin, 2006), thus the activity of muscles in the neck and face will further contribute to the EEG recording. The eyes are another contributor, as their corneo-retinal dipolar structure generates electrooculographic (EOG) signals when moving them or when blinking (Dimigen, 2020; Gramann et al., 2014). Other contributions can be electrical artifacts from technical equipment, or mechanical artifacts, for example, from cable sway or electrode shifts on the scalp (Gwin et al., 2010). As a consequence, the EEG recordings contain a variety of different signals and artifacts that need to be removed or disentangled for a proper analysis, which is aspect iii) of MoBI data processing.

This data cleaning is typically achieved first by removing noisy channels and bad time-segments from the dataset, either by manual inspection (Cohen, 2017a) or using automatic tools such as the *clean_rawdata* plugin to EEGLAB, and by applying spectral filters (Widmann, Schröger, & Maess, 2015). However, especially the removal of time-segments that contain signals from non-brain sources is not a straightforward option for MoBI data analysis, as MoBI experiments can include activities that lead to these contributions throughout the recording, and even specifically during periods

[16] This aspect can be split into a manifold of subaspects and is usually the main focus of the analysis, but as this work is concerned primarily with the preprocessing of MoBI data, it is condensed into one combined aspect here.

of interest (e.g. Jungnickel & Gramann, 2016). Another option to clean the data is independent component analysis (ICA; Bell & Sejnowski, 1995; Hyvärinen, Karhunen, & Oja, 2001), a blind source separation algorithm that leverages the spatial distribution of the EEG channels to unmix the summed recordings on the scalp in order to retrieve the estimated original source activities. Finding this unmixing matrix, however, is an ill-posed problem, as the amount of sources greatly exceed the amount of channels, which is why heuristics are used to obtain this matrix. The basic idea behind ICA algorithms is to reconstruct components such that their activity is minimally normally distributed, rendering them statistically independent according to the central limit theorem (Fischer, 2011). These ICs can then be interpreted on their origin, such as brain, muscle, eye, line noise, or other, less clear sources (Chaumon, Bishop, & Busch, 2015; Pion-Tonachini, Kreutz-Delgado, & Makeig, 2019; Snyder, Kline, Huang, & Ferris, 2015). Additionally, dipole models can be fitted based on a physiological head model with an accuracy of 1-2cm for brain ICs (Acar & Makeig, 2010; Delorme et al., 2011), and the residual variance of this dipole model can be interpreted as its physiological plausibility (Delorme, Palmer, Onton, Oostenveld, & Makeig, 2012).

The final aspect iv) of MoBI data processing marks the functional interpretation of the data for example by analyzing event-related potentials (ERPs, e.g. Luck, 2014), or using time-frequency analysis (event-related spectral perturbations, ERSPs; e.g. Cohen, 2017a; Gramann, Onton, et al., 2010; Onton & Makeig, 2006). This analysis can be done for scalp channels or for brain regions of interest, represented by ICs that are located in the respective brain region (e.g. Lin, Chiu, & Gramann, 2015). To enable the latter, it must first be established which ICs belong together across subjects, which is commonly done with a k-means clustering approach. Here, all ICs of a study can be clustered based on their dipole location, scalp topographies, spectra, ERP, or ERSP, thus allowing for a group-level analysis of EEG source data. This approach to the source-level analysis of EEG data has been increasingly applied in the last years (e.g. Gehrke, Akman, et al., 2019; Gramann, Gwin, et al., 2010; Gramann et al., 2021; Jungnickel & Gramann, 2016; Nenna et al., 2021; Nordin et al., 2019; Protzak & Gramann, 2021; Wagner et al., 2012) and is an important step towards the interpretation of EEG data with respect to its origin in the brain.

In addition to these established approaches, new advancements allow for different analysis techniques that can be highly effective for MoBI data analysis. One approach involves spatial filtering methods that take additional information into account, in order to fuse the different modalities together at an earlier stage in the processing, for example joint decorrelation (de Cheveigné & Arzounian, 2015; de Cheveigné & Parra, 2014), Source Power Comodulation (SPoC; Dähne et al., 2014), which allows the separation of data subspaces that comodulate with a given target value such as the motion of a body part (Gehrke, Guerdan, & Gramann, 2019), Canonical Correlation Analysis (CCA), which can be used to extract subspaces between different high-dimensional datasets (e.g. Biebmann et al., 2010), or Opticat, an ICA technique that makes additional use of eye tracking data to optimize the decomposition (Dimigen, 2020). Another important technique addresses the temporal overlaps between stimuli and motor behavior in ecologically valid experiments. To disentangle these overlaps, the Unfoldtoolbox for regression-based ERP analyses was created (Ehinger & Dimigen, 2019), making use of massive univariate linear models with integrations to deconvolve overlapping events and allowing, for example, to untwine the effect of eye movements from cognitive processes (Dimigen & Ehinger, 2021) or to regress out residual artifacts stemming from gait (Wunderlich & Gramann, 2021).

Challenges of MoBI data processing addressed in this work

As laid out before, the main goals addressed in this dissertation are to make MoBI data analysis more reliable and more usable. Although the analysis of MoBI data can rely on the discussed established processing approaches from classic stationary experimental setups, MoBI requires new developments in data processing to meet its particular requirements and increase its applicability and accessibility. This work is thus concerned with the preprocessing of MoBI data, focusing on aspect iii), the preprocessing of EEG data, but providing tools for all four aspects discussed above. The individual challenges will be laid out in the following.

The first challenge (regarding aspect iii) is the contamination of EEG data by electrical artifacts, addressed in chapter 1. Removing such artifacts, for example stemming from the 50/60 Hz power grid or laboratory equipment, is a basic step in many EEG analysis pipelines, as these contributions can be in frequency ranges of interest (for example the EEG gamma band ranges from 30 to 100 Hz; Hoogenboom, Schoffelen, Oostenveld, Parkes, & Fries, 2006). Thus, if such frequencies are to be investigated, a spectral filter cannot be used to remove the artifacts. Using a spatial filter like ICA alone has the major limitation that it assumes stationarity of the noise, but especially in MoBI data, electrical artifacts can be highly variable. For example, the artifact can change if the participant is equipped with devices in different conditions, or simply by walking through variable electrical fields. Additionally, as MoBI can make use of a wide range of equipment, this equipment can pollute the EEG data with unpredictable artifacts, possibly in frequency ranges of interest. As will be shown, the introduced *Zapline-plus* toolbox solves these issues by finding peaks in the spectrum, separating the data into chunks to alleviate the non-stationarity limitation, and then applying the Zapline (de Cheveigné, 2020) noise removal tool automatically, which combines spatial and spectral filters and alleviates their individual limitations.

The second challenge (also regarding aspect iii) is the decomposition of mobile EEG data using ICA. Although this is a core feature that is applied in many MoBI studies and might even be considered a main factor that enabled the conception of the field (Gramann et al., 2014, 2011; Makeig et al., 2009), the prerequisites for an optimal decomposition are not systematically investigated. In particular in the context of increased mobility in MoBI studies, the impact of different parameters such as the number of EEG channels, the applied spectral high-pass filter, and the rejection of artifactual samples before computing ICA is unclear. Chapter 2 presents two studies that investigate these effects systematically. First, the effect of EEG channel density is investigated in a mobile and a stationary condition of the same dataset. Then, taking more datasets with an increased variety in movement into account, the effect of artifact cleaning in the time-domain (i.e. the rejection of bad samples) is examined. Both studies use the amount of ICs classified as stemming from brain, muscle, eye, and other origin, their dipolarity, and an exemplary signal-to-noise ratio from ERPs as quality measures. The findings from these studies can be summarized into a set of best practices that can be applied when analyzing EEG data from both stationary and mobile experiments with ICA.

The third challenge addressed in this dissertation is the need for a comprehensive toolbox for MoBI data analysis, covering all four aspects of MoBI data processing introduced above. The MOBILAB plugin for EEGLAB (Ojeda et al., 2014) has earlier been suggested as a prime candidate to facilitate MoBI data analysis (Gramann et al., 2011; Makeig et al., 2009). However, it is not supported anymore and its functionalities are limited. This toolbox is mostly centered around a visual inspection tool and the syn-

chronized export of the data to EEGLAB, as its analysis of body data includes only motion with a basic event extraction, thus limiting aspect ii) of MoBI data processing. The main limitation of MOBILAB, however, is the fact that it is an enclosed environment, and once the data is exported to EEGLAB it cannot be transferred back to MOBILAB again, meaning that the benefits of MOBILAB are not applicable anymore. Thus, even if there are routines for body data processing and event extraction available in MOBILAB, their use requires full knowledge of the downstream analysis plans, which may be the case in an ideal world, but is rarely so in the messy reality of MoBI data analysis. In this regard, MOBILAB solves the initial synchronization of aspect i), but it cannot maintain a synchronized analysis of body and EEG data, further limiting aspect ii). It would thus be preferable to have the tools for the analysis of body data in the same environment as the tools for the analysis of EEG data, and the two modalities available in parallel. Regarding aspect iii), the preprocessing of mobile EEG data, a particularly important limitation is that it can be time-consuming and intimidating, and automated methods are not available or not as reliable as desired. Regarding aspect iv), a limitation is that the k-means clustering of ICs does not result in stable results (Gramann et al., 2021), further necessitating manual inspection or other assessments (Ehinger et al., 2014). As already briefly introduced and will be discussed in depth in chapter 5, a robust automated processing pipeline would be beneficial for MoBI, in particular because MoBI is an imaging modality that attracts interest from other scientific fields—and therefore may be used by researchers who are not trained in EEG analysis. However, thus far, such automated tools are not readily available, especially in light of the requirements of mobile settings, and automation in EEG data cleaning is not universally accepted, mostly in reference to the lack of reliable tools and the preference of human expert assessment of the data (Cohen, 2017a). Some efforts have been made to make this processing more standardized and reliable (e.g. Bigdely-Shamlo, Mullen, Kothe, Su, & Robbins, 2015; Gabard-Durnam, Leal, Wilkinson, & Levin, 2018; Pernet, Martinez-Cancino, Truong, Makeig, & Delorme, 2021), but they are not tuned to mobile EEG and they do not address the combined requirements of MoBI data analysis of aspects i) to iv). To address this challenge, the *BeMoBIL Pipeline*[17] was developed—an automated, open-source, and easy-to-use pipeline in MATLAB that supports the time-synchronized handling and comprehensive analysis of multimodal MoBI data. As will be shown, it solves the limitations of MOBILAB identified here. It also makes use of the Zapline-plus toolbox presented in chapter 1 and applies insights gained in the studies presented in chapter 2. By using a variety of wrapper functions to process EEG and other data in a robust and transparent manner, it aims to address the reservations regarding automated EEG data processing, and enhance the reliability and usability of MoBI as a research method.

[17] https://github.com/BeMoBIL/bemobil-pipeline

1 Zapline-plus

Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG

Klug, M., & Kloosterman, N. A. (2022). Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG. *Human Brain Mapping*, 43(9), 2743–2758. https://doi.org/10.1002/hbm.25832.

Received: 18 October 2021 | Revised: 17 February 2022 | Accepted: 1 March 2022

DOI: 10.1002/hbm.25832

TECHNICAL NOTE

WILEY

Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG

Marius Klug[1] | Niels A. Kloosterman[2,3]

[1]Biopsychology and Neuroergonomics, Institute of Psychology and Ergonomics, Technische Universität Berlin, Berlin, Germany

[2]Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany

[3]Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

Correspondence
Marius Klug, Biopsychology and Neuroergonomics, Technische Universität Berlin, Berlin 10623, Germany.
Email: marius.klug@tu-berlin.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: GR2627/8-1; U.S. Air Force, Grant/Award Number: ONR 10024807

Abstract

Removing power line noise and other frequency-specific artifacts from electrophysiological data without affecting neural signals remains a challenging task. Recently, an approach was introduced that combines spectral and spatial filtering to effectively remove line noise: Zapline. This algorithm, however, requires manual selection of the noise frequency and the number of spatial components to remove during spatial filtering. Moreover, it assumes that noise frequency and spatial topography are stable over time, which is often not warranted. To overcome these issues, we introduce Zapline-plus, which allows adaptive and automatic removal of frequency-specific noise artifacts from M/electroencephalography (EEG) and LFP data. To achieve this, our extension first segments the data into periods (chunks) in which the noise is spatially stable. Then, for each chunk, it searches for peaks in the power spectrum, and finally applies Zapline. The exact noise frequency around the found target frequency is also determined separately for every chunk to allow fluctuations of the peak noise frequency over time. The number of to-be-removed components by Zapline is automatically determined using an outlier detection algorithm. Finally, the frequency spectrum after cleaning is analyzed for suboptimal cleaning, and parameters are adapted accordingly if necessary before re-running the process. The software creates a detailed plot for monitoring the cleaning. We highlight the efficacy of the different features of our algorithm by applying it to four openly available data sets, two EEG sets containing both stationary and mobile task conditions, and two magnetoencephalography sets containing strong line noise.

KEYWORDS
artifact removal, filter, gamma oscillations, line noise, M/EEG, preprocessing, principal component analysis, signal processing, spectral analysis

1 | INTRODUCTION

The task paradigm is well thought out. The experiment set up, the electroencephalography (EEG) recording goes well, 30 data sets and more. A masterpiece, really. Finally, you have time to plot your first power spectra. Then, peaks in your spectra, particularly at 50 or 60 Hz, but also in other frequencies, right where you want to analyze your data.

Removing frequency-specific noise artifacts from electrophysiological data is a key issue in any EEG or magnetoencephalography (MEG) experiment. Modern laboratories contain many different electrical devices that all need power, and with great power comes great line noise. However, noise is not only limited to the 50/60 Hz power line artifact, but may also arise from many different sources. Recently, a novel algorithm, Zapline, was introduced that combines spectral and

spatial filters to isolate and remove the power line noise (de Cheveigné, 2020). In this article, we present an adaptive wrapper software for Zapline to enable the fully automatic removal of frequency-specific noise artifacts, including the selection of noise frequencies, chunking the data into segments in which the noise is spatially stable, automatically selecting the number of principal components to remove with Zapline, as well as a comprehensive analysis and visualization of the cleaning and its impact on the data.

1.1 | EEG noise removal is especially difficult in mobile experiments

Mobile EEG studies require specific treatment to remove noise stemming from muscles and other sources, and often independent component analysis (ICA) can be used for this (Klug & Gramann, 2020). Finding the right way to remove frequency-specific noise from the data, however, is a difficult task, especially since it is not necessarily spatially stable and thus can have a strong negative impact on ICA. Shielding the laboratory, finding the sources, and eliminating them before recording the data help to alleviate the issue, but this is not always feasible, and sometimes the noise goes unnoticed at first. As recent developments in EEG experimental paradigms show a trend toward measuring the human in its natural habitat, the world (Gramann, Ferris, Gwin, & Makeig, 2014), it can become increasingly difficult or impossible to control noise sources. The fields of mobile brain/body imaging (Gramann et al., 2011; Jungnickel, Gehrke, Klug, & Gramann, 2019; Makeig, Gramann, Jung, Sejnowski, & Poizner, 2009) and neuroergonomics (Dehais, Lafont, Roy, & Fairclough, 2020; Raja & Matthew, 2009) use devices like virtual reality head mounted displays, motion tracking, eye tracking, treadmills, flight simulators, or actual airplanes, and more. In these experiments, participants move around and interact with the world, including for example navigating through a city (Wunderlich & Gramann, 2018), a virtual maze (Gehrke & Gramann, 2021), or flying an airplane (Dehais et al., 2019). These data sets are almost always riddled with frequency-specific noise, not only stemming from the power line but also other devices, and often it is just accepted that recordings contain noise. Removing this noise during processing is especially important when comparing different conditions like seated versus mobile experiments, as different noise sources may be nearby for the different conditions, and untreated noise can be wrongfully interpreted as an effect of the conditions.

1.2 | Line noise artifacts are particularly strong in MEG

MEG is a technique closely related to EEG, in which rather than the electrical activity itself, its concurrent magnetic fields are recorded (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). Compared to EEG, MEG allows for better spatial specificity of (superficial) sources of neural activity in the brain. Moreover, it does not require extended subject preparation and electrode gel, which makes MEG more feasible for clinical populations as well as children. Magnetic fields are less distorted by the skull than electrical activity, which makes MEG better suited for investigating high-frequency neural activity in the so-called gamma band (although gamma is investigated in EEG as well, for example, Kloosterman et al., 2019). However, the gamma band ranges from roughly ~30 to 100 Hz (Hoogenboom, Schoffelen, Oostenveld, Parkes, & Fries, 2006), which encompasses the 50 or 60 Hz line noise (and possibly it is first harmonic), to which MEG is highly sensitive and which can outweigh neural activity by several orders of magnitude. This noise is often removed using strong filters (see next section), which come at the cost of completely removing true neural activity in this range as well. This approach hampers in-depth investigation of the function of gamma activity in neural processing.

1.3 | Noise can be removed with spectral filters, regression, or spatial filters

Taken together, removing frequency-specific noise is a vital part of data processing.

Several methods are available to remove this noise, but these all come with individual drawbacks. Three main approaches can be distinguished:

1. *Spectral filters*: Filtering the data with a simple low-pass or notch filter is the most conventional approach. However, a low-pass filter may reduce the quality of decomposing the data using ICA (Dimigen, 2020; Hyvarinen, 1997) and a notch filter must have a steep roll-off to keep the notch small, which comes with the potential of ringing artifacts (Widmann, Schröger, & Maess, 2015). Additionally, both options remove all information in (or even above) the noise range and will make analysis of these frequencies impossible. An approach related to notch filtering is interpolation of the data in the frequency domain between directly neighboring frequencies that are unaffected by the noise (e.g., 48 to 52 Hz), followed by transformation of the data back into the temporal domain (Leske & Dalal, 2019). This approach indeed does not introduce a deep notch in the data at the line noise frequency, but nevertheless, all information at the line noise frequency is destroyed, rendering further analysis impossible.
2. *Regression-based approaches*: Regressing a target signal out of the data is another often used tool. Examples are the CleanLine plugin of EEGLAB (Delorme & Makeig, 2004), which uses a frequency-domain regression to remove sinusoidal artifacts from the data, or TSPCA, which uses a provided reference signal (de Cheveigné & Simon, 2007). These approaches depend on either a provided reference or a successful generation of a target signal in a given frequency. Here, some noise may be left in the data, especially fluctuations in amplitude or phase of the noise can be difficult to remove.
3. *Spatial filters*: Spatial filter options like ICA or joint diagonalization (de Cheveigné & Parra, 2014) are widely used and reduce noise by generating their own noise reference signal from a linear combination of all channels.

However, noise is not always linearly separable from neural activity, and thus removing noise components can inadvertently remove brain signals too. These methods are also vulnerable to nonstationary of noise, which can be particularly problematic in mobile EEG experiments. Finally, removing noise components from the data with a spatial filter relying on linear algebra always reduces the algebraic rank of the data matrix and can thus limit further analyses (Cohen, 2021). In sum, all of the above options come with drawbacks.

1.4 | Zapline is a promising tool

Recently, a promising new method that combines the spectral and spatial filtering approaches to overcome some of these issues has been introduced: Zapline (de Cheveigné, 2020). Zapline first uses a notch filter and its complementary counterpart to split the data into the clean and the noisy part, where summing them together would result in the original data. Then, the noisy part is decomposed using joint decorrelation (de Cheveigné & Parra, 2014), and the components that carry most of the noise are removed from the noisy data. Last, the now cleaned, previously noisy, data and the clean data are summed together to form the final cleaned data set. This approach has the advantage of (in principle) not leaving a notch in the spectrum while also not reducing the rank of the data matrix. Additionally, since Zapline removes noise using a spatial filter, it is not sensitive to variations in power of the noise over time. This is in contrast to regression-based methods such as CleanLine, which removes an oscillation of fixed amplitude from short data segments.

1.5 | Challenges of Zapline

However, some issues remain. On the one hand, as Zapline makes use of a spatial filter, it assumes a stable spatial topography of the noise over time. However, especially in mobile and task-based experiments, the spatial distribution of the noise can change (proximity changes of devices, orientation changes of the participant, touching cables, etc.). When comparing different conditions, it may even be the case that some noise artifacts are entirely absent in parts of the recording. This issue can lead to insufficient cleaning in some, too much cleaning in other parts of the data, or the need to remove many components, which can distort the data. Furthermore, a key challenge of Zapline is that it needs to be manually tuned to each data set. Specifically, the following issues can be discerned:

1. Finding out the correct number of components to remove. This is not straightforward—recommendations range from two to four (de Cheveigné, 2020), but in individual cases, as many as 25 components have been reported to be removed (Miyakoshi, Schmitt, Erickson, Sweeney, & Pedapati, 2021). Presumably, the number of components depends on the noise structure and number of sensors or electrodes. In our tests with high-density EEG and MEG data, removing 10–15 components was usually necessary to contain the noise.
2. The noise frequency needs to be chosen. In most cases, choosing the power line frequency is sufficient, but sometimes additional frequencies can be found, like a 90 Hz oscillation of a virtual reality head-mounted display, or other frequencies due to additional devices in the lab. Moreover, in some of our tests, Zapline proved to be sensitive to even small changes in the target frequency in the range of 0.1 Hz, which are hard to know in advance, especially if the frequency shifts during the recording.

Taken together, Zapline is a powerful tool but requires manual parameter selection, and using Zapline in an automated analysis pipeline is difficult due to this process of fine-tuning.

1.6 | Zapline-plus aims to overcome Zapline's manual tuning issues

We created Zapline-plus—an adaptive wrapper software for Zapline that allows fully automatic use without parameter tuning. The software searches for outlier peaks in the spectrum and applies Zapline to remove these. To alleviate the stationarity issue, the data is adaptively segmented into chunks in which the frequency-specific noise is relatively constant, as determined by the covariance structure of the data. Within each chunk, the individual chunk noise peak frequency is detected, and Zapline is applied at this frequency. An adaptive component detector then removes only the strongest noise components. Finally, a check of the cleaning is performed and the detection process is adjusted accordingly and the procedure is repeated if necessary. All used parameters and several performance indicators are stored to enable an understanding and easy replication of the cleaning, and a detailed plot is created to allow inspection of the cleaning performance. We tested the software on two open EEG and two open MEG data sets with promising results. We discuss limitations and implications for automated processing pipelines. The MATLAB source code of the software is available for download at https://github.com/MariusKlug/zapline-plus.

1.7 | The software package

In this section, we describe the different aspects of the adaptive algorithm, the processing flow, as well as the produced plots, and the optional parameters in case the default values are suboptimal.

2 | ALGORITHM

Zapline-plus contains several components that are discussed in the following.

The processing steps include:

1. the detection of noise frequencies,
2. adaptive segmentation of the time series in chunks based on the stability of the noise topography,

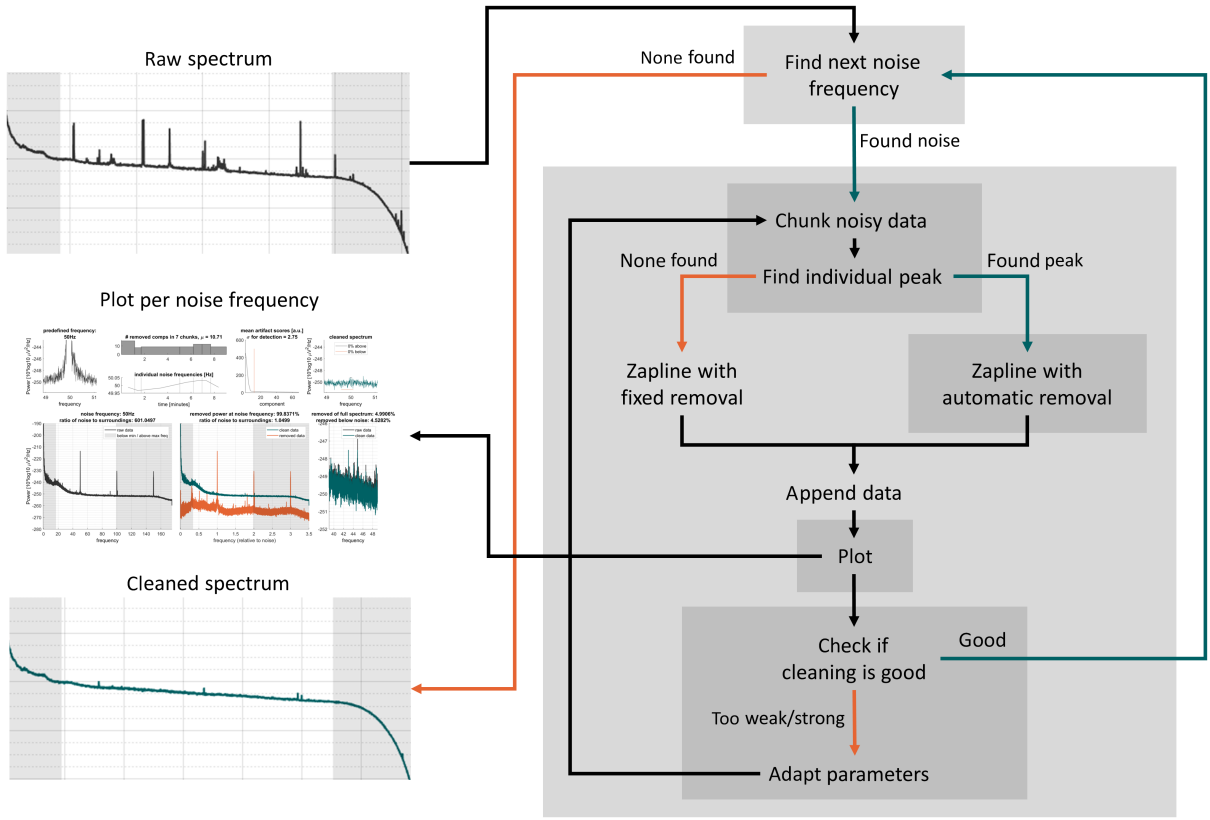


FIGURE 1 Processing flow of the Zapline-plus algorithm. Please refer to the text for details about the individual steps

3. applying Zapline on each segment at the detected frequency,
4. automatic detection and removal of noise components, and
5. adaptively changing and repeating the processing to prevent too weak or too strong cleaning.

The processing workflow is visualized in Figure 1.

2.1 | Noise frequency detection

Noise frequencies are defined as frequencies having abnormally large power compared to the neighboring frequencies, as determined by spectral density estimation using Welch's method (Welch, 1967). We used a hanning window because it resulted in less noisy spectra than the default hamming window for some data sets. The computed power spectral density (PSD) values are then log-transformed ($10\log_{10}$) and the mean over channels is computed (corresponding to a geometric mean of the spectra that is less outlier-driven). We chose this approach because, in our experience, the individual channel spectra are not always normally distributed, especially if there are a few very noisy outlier channels. In these cases, they mask the efficacy of Zapline and hide details of the overall spectrum. Importantly, the resulting geometric mean PSD is always greater than equal to the log of the arithmetic mean PSD. Subsequently, the first outlier frequency within a minimum (17 Hz) and maximum (99 Hz) frequency is searched with a 6 Hz moving window. If a frequency has a difference >4 of log PSD to the center log PSD (mean of left and right thirds around the current frequency), it is found to be an outlier and the search is stopped. As the input is in $10\log_{10}$ space, a difference of 4 corresponds to a 2.5-fold increase of the outlier power over the center.

2.2 | Adaptive time series segmentation into chunks for cleaning

Zapline detects noise components in the data using spatial principal components, and thus works on the assumption of a spatial noise distribution that is stable over time. However, this is not always guaranteed. Even small shifts in head orientation or a relocation of the participant due to the experimental paradigm can lead to slightly different noise topography or entirely new noise sources. To alleviate this issue, we implemented an adaptive method that segments the data into chunks with relatively fixed noise topography. Specifically, we apply the following steps:

1. Narrowband filter the continuous data around the detected noise frequency ±3 Hz.
2. Compute the channel-by-channel (i.e., sensors or electrodes) covariance matrix within data epochs of 1 s duration.

3. Compute the distance between pairs of channels in successive covariance matrices. This yields a measure of the change in covariance over time. A small distance indicates that the noise is roughly constant, whereas a large distance indicates a change in noise topography.
4. Determine segments (chunks) of stable noise topography by detecting peaks in the covariance stationarity.

We found that this method reliably detected segments in which the noise was spatially constant. However, we chose a minimum segment duration of 30 s to enable sufficient data for the spatial decomposition employed by Zapline. Applying Zapline separately to each chunk does not only allow different linear decompositions per chunk, but also allows fine-tuning of the target frequency to the peak in this chunk, further improving Zapline's effectiveness. Finally, this adaptive segmentation might help noise removal in cases where a change in noise topography is related to an external event in task-related data that cause subjects to move, such as a trial onset or the start of a short break in the experiment during which the recording continues.

2.3 | Application of Zapline

To detect the chunk's noise peak we first search for the peak frequency within a ±0.05 Hz range around the previously detected target frequency. We then determine a fine-grained threshold to define oscillations being present or absent in that chunk: The mean of the two lower 5% log PSD quantiles of the first and last third in a 6 Hz area around the target frequency is computed, and the difference to the center power (mean of left and right third log PSDs around the target frequency) is taken as a measure of deviation from the mean. (On a side note, both the *SD* and the median absolute deviation did not lead to good results, as they can be driven by outliers to the top.) Finally, the threshold is defined as the center power + 2 × deviation measure, and if the log PSD of the found peak frequency is above this threshold, the chunk is found to have a noise artifact.

In the next step, cleaning is performed on a per-chunk basis using the original Zapline algorithm, using either the found frequency peak and adaptive removal settings (starting with 3 *SD*, see Section 2.4, adaptive, see Section 2.5), or the original noise peak of the full data set and a fixed number of components to remove (starting at 1, adaptive, see Section 2.5). We chose to remove a minimum number even when no artifact was found, to make sure even missed artifacts are removed while also making sure not too many components are removed in case no artifact is actually present in the chunk at that frequency.

2.4 | Detection of noise components

One essential parameter of Zapline is the number of to-be-removed components after sorting components based on the amount of explained variance. So far, this had to be chosen manually, based on visual inspection of the "elbow" in the sorted components (i.e., transition from a sharp to shallow drop-off). We adapted the function to include a detector for outliers in the computed JD scores that represent to what extent the components load on the noise. To this end, an iterative approach based on a standard mean + *SD* threshold is used. In each iteration, the detector removes outliers and then recomputes mean and *SD* across all components, and repeats this procedure until none are left. The number of removed outliers is then taken as the number of components to remove in Zapline. We found this iterative approach to be more robust than an approach based on the median absolute deviation in this scenario. In a final step, if the number of found outliers is less than the entered fixed removal, the latter is being used, and, to prevent removing an unreasonable amount of components, the number is capped at one-fifth of the components. We found a value of 3 *SD*s to work well in most cases, but sometimes even this automatic detector removes too many or too few components, which is why the *SD* parameter is adapted in the next step.

2.5 | Adaptive changes of the cleaning procedure

After each chunk has been cleaned, the chunks are concatenated again and the cleaned spectrum is computed as in Section 2.4. Although the software already contains several steps to find an optimal noise reduction, the cleaning can still be too weak or too strong. We implemented a check for suboptimal cleaning by using the same fine-grained threshold as in Section 2.3. This check is now applied to search for introduced notches or remaining peaks in the power spectrum, indicating that the cleaning was too strong or too weak, respectively. Specifically, if there are 0.5% of samples of the spectrum in the range of ±0.05 Hz around the noise frequency above the threshold of center power + 2 × deviation measure, the cleaning is found to be too weak. If there are 0.5% samples of the spectrum in the range of −0.4 to +0.1 Hz around the noise frequency below the threshold of center power − 2 × deviation measure, the cleaning is found to be too strong. If the cleaning was too weak, the *SD* for the number of noise components is reduced by 0.25, up to a minimum of 2.5, and the fixed number of removed components (for chunks where no noise was detected) is increased by 1. If the cleaning was too strong, the *SD* for step "Noise component detection" is increased by 0.25, up to a maximum of 4, and the fixed number of removed components (for chunks where no noise was detected) is decreased by 1, up to a minimum of the initial fixed removal of 1. Too strong cleaning always takes precedence over too weak cleaning, and if the cleaning was once found to be too strong, it can never become stronger again even after it was weakened and is now found to be too weak.

Using these new values, the entire cleaning process of this noise frequency is re-run and re-evaluated. This leads to a maximally reduced noise artifact while ensuring minimal impact on any other frequencies. If no further adaptation of the cleaning needs to be performed, this noise frequency is assumed to be cleaned, and the next noise frequency is searched (Section 2.4) using the current noise

frequency +0.05 Hz as the new minimum frequency. If no other noise frequency is found, the cleaning completes.

2.6 | Output figures

For every frequency-specific noise artifact that is removed, a figure is generated. Example plots can be seen in Figures 4 and 5. Importantly, the plot per frequency is being overwritten in case the parameters are adapted, so the final plots only show the final values. These plots contain all information that is necessary to determine the success of the cleaning in a colorblind-friendly color scheme. The top row of the figure contains visualizations of the cleaning process, the bottom row contains the final spectra and analytics information (Figure 2).

In the top row, first, the noise frequency of this iteration is shown in a zoomed-in spectrum to ±1.1 Hz around the frequency (Figure 3a). The threshold that led to the detection of this frequency is shown in addition (red line), unless the detection is disabled. Next, the cleaning of the individual chunks is visualized in two ways: The number of removed components per chunk (Figure 3b), and the individual noise frequency detected for each chunk (Figure 3c). Additionally, chunks in which no noise was detected are marked as such and the mean number of removed components is denoted in the title of the plot. As each chunk contains a set of components and accompanying artifact scores, this is too much to be visualized without cluttering the plot, so we chose to only plot the mean artifact scores overall chunks next (Figure 3d). This plot also contains the mean number of removed components (red vertical line). Ideally, this line should cross the scores around the "elbow" of the curve, which indicates that the outliers (i.e., the components which carry most of the noise) were detected correctly. The abscissa is cut to one-third of the number of components to allow the visualization of the knee point. This is independent of the nkeep parameter that can be set (Section 2.7). The *SD* value that was used for the detector is denoted in the title of this plot. To finalize the visualization of the cleaning process, the zoomed-in spectrum of the cleaned data is shown alongside the thresholds that determine if the cleaning was too strong or too weak with respective horizontal lines (Figure 3e). The same *y*-axis is used as in Figure 3a to allow a comparison of pre- versus postcleaning. The legend of this plot also contains the proportion of frequency samples that are below or above these thresholds, which determines whether the cleaning needs to be adapted. It may happen that values exceeding these

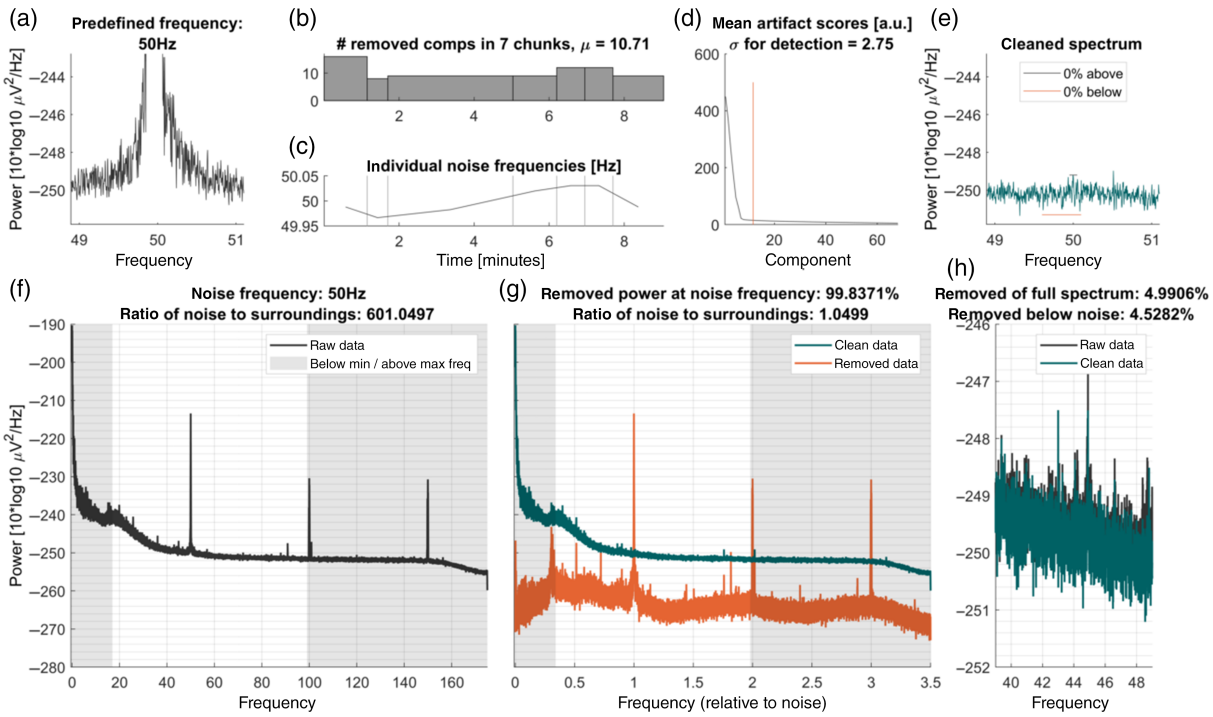


FIGURE 2 Example output plots produced by Zapline-plus for 50 Hz line noise. Shown is a 9 min MEG data set from MEG study I (Section 3.1), with 50 Hz predefined as the noise to remove. For a detailed explanation of the individual subplots, see section 2.6. (a) Power spectrum centered around the noise frequency. (b) Number of components removed by Zapline for each chunk. Chunks were defined as periods in which the noise was spatially stable. (c) Specific noise frequencies are detected within each chunk. (d) Component scores, sorted in descending strength. *Red line*: threshold for rejection based on outlier detection. (e) Same as (a), but after removal of the noise components. (f) Full power spectrum, depicting both the line noise and (sub-)harmonics. (g) Same as (e), but showing clean and noise data separately. The *x*-axis expresses frequency relative to the removed noise frequency, where 1 indicates the noise frequency. (h) Power spectrum of 10 Hz range below the noise frequency, indicating to what extent non-noise frequencies were affected by the cleaning

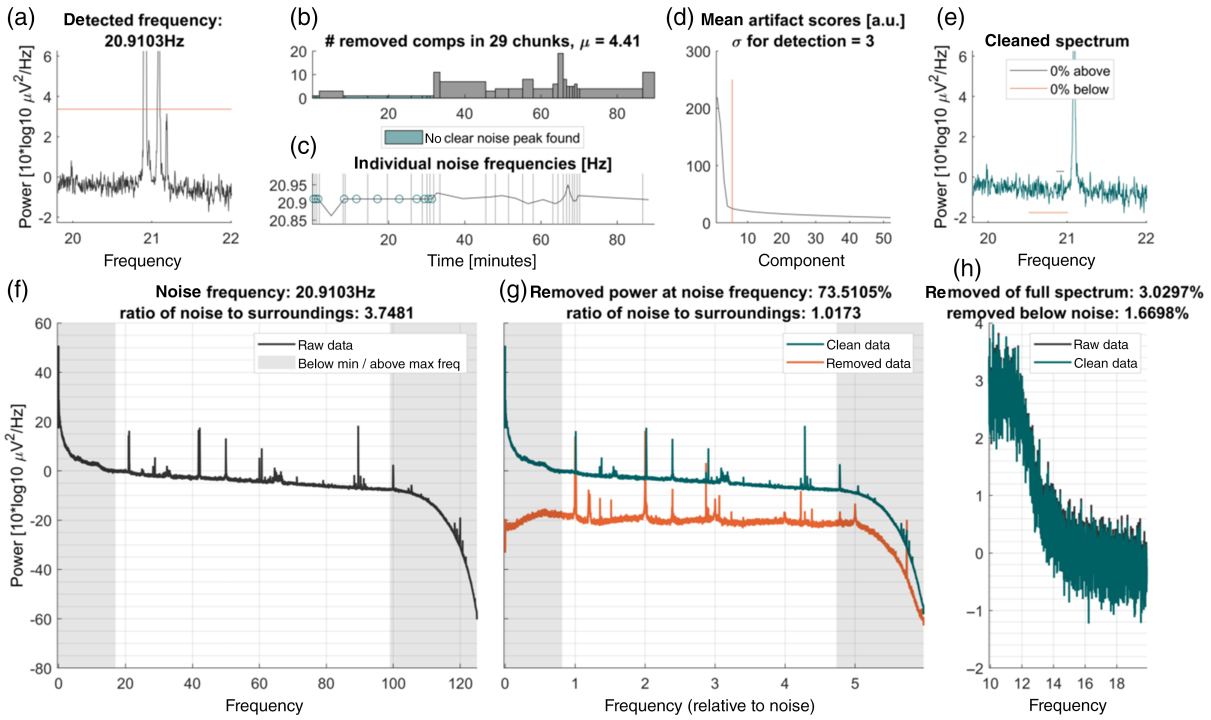


FIGURE 3 Example output plots produced by Zapline-plus for 21 Hz noise. Shown is an 87 min EEG data set from EEG study II (Section 3.1) containing a mobile and a stationary condition. This noise artifact was present only in the first part of the data. For an explanation of the individual subplots, refer to the "Plots" section. For a detailed explanation of the individual subplots, refer to section 2.6. Conventions as in Figure 2

thresholds remain, which can be either due to the minimum or maximum *SD* level being reached or due to the fact that the cleaning would to too strong if set to a stronger level.

Figure 3f shows the raw spectrum as the mean of the log-transformed channel spectra. Vertical shaded areas denote the minimal and maximal frequency to be checked by the detector, as this can be useful to know in case a spectral peak is present in this area and thus goes undetected. In Figure 3g, the spectra of the cleaned (green), as well as the removed data (red), are plotted. The abscissa in this plot is relative to the noise frequency which facilitates distinguishing removed harmonics from other frequencies. Last, as it was shown that Zapline can have undesirable effects on the spectrum below the noise frequency (Miyakoshi et al., 2021), Figure 3h shows the spectra of the raw and cleaned data again zoomed in to the part 10 Hz below the noise frequency to determine if this was the case. In the title of Figure 3g,h, we also denote several analytics: the proportion of removed power (computed on log-transformed data, corresponding to the geometric mean) of the complete spectrum, of the power ±0.05 Hz around the noise frequency, and of the power −11 Hz to −1 Hz below the noise frequency, as well as the ratios of power ±0.05 Hz around the noise frequency to the center power before and after cleaning. These plots facilitate both, an understanding of the data set itself, as well as the functioning of the cleaning. Although the algorithm is adaptive in many ways and should work "as is," it is naturally possible that the noise has properties that make cleaning with Zapline-plus difficult or impossible. Hence, these plots should always be inspected to determine if the cleaning was successful.

2.7 | Parameters and outputs

Although we strive to provide a fully automatic solution with no need for parameter tweaking, we still would like to provide options for all relevant aspects of the algorithm, including switching adaptations off in case they do not work as intended. Here, we describe the parameters, our reasoning for the default values and reasonable ranges, as well as the output of the cleaning and additional thoughts. The data and sampling rate are required inputs, all additional parameters can be entered either in key-value pairs or as a single struct:

- *noisefreqs* (default = empty): Vector with one or more noise frequencies to be removed. If empty or missing, noise frequencies will be detected automatically. Individual chunk peak detection will still be applied if set.
- *minfreq* (default = 17): Minimum frequency to be considered as noise when searching for noise frequencies automatically. We chose this default as it is well above the potentially problematic

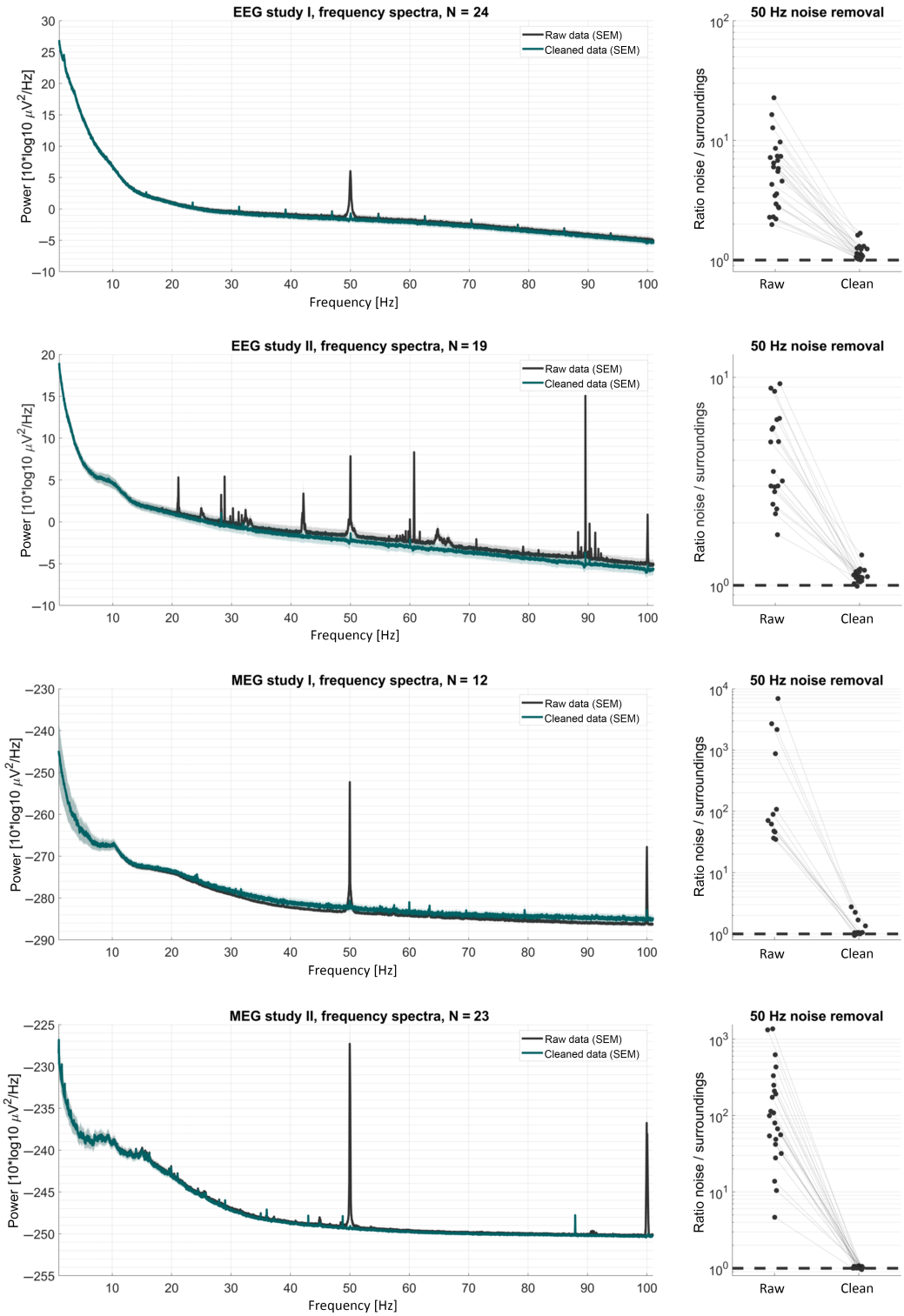


FIGURE 4 Frequency spectra and 50 Hz noise removal results of the example data sets. Rows results for the four M/EEG data sets. *Left panels*: frequency spectra before and after applying Zapline-plus. *Right panels*: ratio of power at noise/surrounding frequency for raw and cleaned data. A ratio of 1 (i.e., 10^0) indicates the absence of any remaining noise artifact in the power spectrum

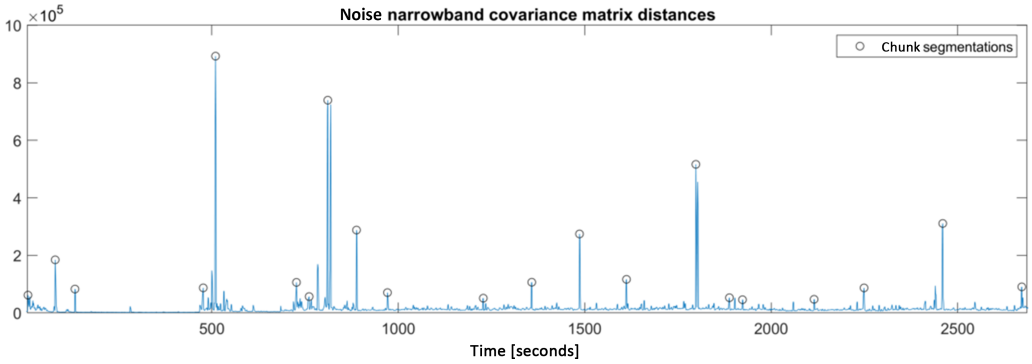


FIGURE 5 Example of the automatic data segmentation into chunks. Plotted are the distances of the narrowband-filtered covariance matrices of 1 s chunks of an example data set. Large distances correspond to a shift of the spatial distribution of the noise, which is a violation of the stability assumption of a spatial filter. The adaptive segmentation finds these peaks and uses them to initiate a new chunk, thus alleviating the issue. Note that not all peaks are detected as a chunk segmentation point (e.g., a double peak at around 800 s or a smaller peak at around 2,400 s), which is due to the minimum chunk length that is necessary to ensure a reliable computation of the spatial filter in the first place. Also, note that this is a single case example only meant to illustrate the algorithm

range of alpha oscillations (8–13 Hz) and also above the third subharmonic of 50 Hz, which was present in some MEG data sets.

- *maxfreq* (default = 99): Maximum frequency to be considered as noise when searching for noise freqs automatically. We chose this default as it is below the second harmonics of the 50 Hz line noise. If the line noise cannot be removed successfully in the original frequency, trying to remove the harmonics can potentially lead to overcleaning.
- *adaptiveNremove* (default = true): Boolean if the automatic detection of number of removed components (Section 2.4) should be used. If set to false, a fixed number of components will be removed in all chunks. As this is a core feature of the algorithm it is switched on by default.
- *fixedNremove* (default = 1): Fixed number of removed components per chunk. If adaptiveNremove is set to true, this will be the minimum. Will be automatically adapted if "adaptiveSigma" is set to true. We chose this default to remove at least one component at all times, no matter whether or not a noise oscillation was detected per chunk, as the detector can fail to find an oscillation that should be removed, and removing a single component does not lead to a large effect if no oscillation was present in the chunk.
- *detectionWinsize* (default = 6): Window size in Hz for the detection of noise peaks. As the detector uses the lower and upper third of the window to determine the center power (Section 2.3) this leaves a noise bandwidth of 2 Hz. In our tests, some data sets indeed had such a large bandwidth of line noise, which can occur if the noise varies in time.
- *coarseFreqDetectPowerDiff* (default = 4): Threshold in $10\log^{10}$ scale above the center power of the spectrum to detect a peak as noise frequency. If this is too high, weaker noise can go undetected and thus uncleaned. If it is too low, spurious peak oscillations can be wrongfully classified as noise artifacts. This default corresponds to a 2.5-fold increase of the noise amplitude over the center power in the detection window which worked well in our tests.
- *coarseFreqDetectLowerPowerDiff* (default = 1.76): Threshold in $10\log^{10}$ scale above the center power of the spectrum to detect the end of a noise artifact peak. This is necessary for the noise frequency detector to stop. This default corresponds to a 1.5× increase of the noise amplitude over the center power in the detection window which worked well in our tests.
- *searchIndividualNoise* (default = true): Boolean whether or not individual noise peaks should be applied on the individual chunks instead of the noise frequency specified or found on the complete data (Section 2.3). As this is a core feature of the algorithm it is switched on by default.
- *freqDetectMultFine* (default = 2): Multiplier for the 5% quantile deviation detector of the fine noise frequency detection for adaption of *SD* thresholds for too strong/weak cleaning (Section 2.3). If this value is lowered, the adaptive changes of Section 2.5 are stricter, if it is increased, these adaptations happen more rarely.
- *detailedFreqBoundsUpper* (default = [−0.05 0.05]): Frequency boundaries for the fine threshold of too weak cleaning. This is also used for the search of individual chunk noise peaks as well as the computation of analytics values of removed power and the ratio of noise power to surroundings. Low values mean a more direct adaptation to the peak, but too low values might mean that the actual noise peaks are missed.
- *detailedFreqBoundsLower* (default = [−0.4 0.1]): Frequency boundaries for the fine threshold of too strong cleaning. Too strong cleaning usually makes a notch into the spectrum slightly below the noise frequency, which is why these boundaries are not centered around the noise peak.
- *maxProportionAboveUpper* (default = 0.005): Proportion of frequency samples that may be above the upper threshold before cleaning is adapted. We chose this value since it allows a few potential outliers before adapting the cleaning.
- *maxProportionBelowLower* (default = 0.005): Proportion of frequency samples that may be below the lower threshold before

cleaning is adapted. We chose this value since it allows a few potential outliers before adapting the cleaning.

- *noiseCompDetectSigma* (default = 3): Initial *SD* threshold for iterative outlier detection of noise components to be removed (Section 2.4). Will be automatically adapted if "adaptiveSigma" is set to 1. This value led to the fewest adaptations in our tests.
- *adaptiveSigma* (default = 1): Boolean if automatic adaptation of noiseCompDetectSigma should be used. Also adapts fixedNremove when cleaning becomes stricter (Section 2.5). As this is a core feature of the algorithm it is switched on by default.
- *minsigma* (default = 2.5): Minimum when adapting noiseCompDetectSigma. We found that a lower *SD* than 2.5 usually resulted in removing too many components and a distortion of the data.
- *maxsigma* (default = 4): Maximum when adapting noiseCompDetectSigma. We found that a *SD* higher than 4 usually did not relax the cleaning meaningfully anymore.
- *chunkLength* (default = 0): Length of chunks to be cleaned in seconds. If set to 0, automatic, adaptive chunking based on the data covariance matrix will be used.
- *minChunkLength* (default = 30): Minimum length of the chunks when adaptive chunking is used. We chose a minimum chunk length of 30 s because shorter chunks resulted in both, a sometimes suboptimal decomposition within Zapline and a lower frequency resolution for the chunk noise peak detector. Smaller chunks result in better adaptation to nonstationary noise, but also potentially worse decomposition within Zapline. The necessary minimum chunk length for ideal performance may also depend on the sampling rate.
- *winSizeCompleteSpectrum* (default = 0): Window size in samples of the pwelch function to compute the spectrum of the complete data set for detecting the noise frequencies. If 0, a window length of sampling rate × chunkLength is used. This parameter mainly adjusts the resolution of the computed spectrum. We chose relatively long windows to ensure a high resolution for the noise frequency detector.
- *nkeep* (default = 0): Principal Component Analysis dimension reduction of the data within Zapline. If 0, no reduction will be applied. This option can be useful for extremely high number of channels in which there is a risk of overfitting, but in our tests, even on high-density EEG and MEG data, it did not lead to better results.
- *plotResults* (default = 1): Boolean if the plot should be created.

After completing the cleaning, Zapline-plus passes out the complete configuration struct including all adaptations that were applied during the cleaning. This allows a perfect replication of the cleaning when applying the configuration to the same raw data again and facilitates reporting the procedure. Additionally, the generated analytics values that can be found in the plot are also passed out as a struct: raw and final cleaned log spectra of all channels, *SD* used for detection, proportion of removed power of the complete spectra, the noise frequencies, and below noise frequencies, ratio of noise powers to surroundings before and after cleaning per noise frequency, proportion of spectrum samples above/below the threshold for each frequency, matrices of number of removed components per noise frequency and chunk, of artifact component scores per noise frequency and chunk, of individual noise peaks found per noise frequency and chunk, and whether or not the noise peak exceeded the threshold, per noise frequency and chunk. These values allow an easy check of the complete Zapline-plus cleaning both for each subject and on the group level.

2.8 | A note on the sampling rate of the data

Modern M/EEG setups typically record data at high sampling rates of at least 500 Hz (1,200 Hz is common for MEG), which allows for high temporal resolution and investigation of very high frequencies. However, brain activity is typically not quantified beyond 100 Hz, and lower sampling rates such as 250 Hz are typically deemed sufficient for ERP studies investigating the onset of neural responses. Importantly, the presence of high frequencies in the data poses a major challenge for line noise removal with Zapline, because Zapline also needs to handle the (sub)-harmonics (integer divisions and multiples of the line noise frequency) that emerge with frequency-specific noise. For example, at a sample rate of 1,200 Hz, Zapline will remove line noise at 50 Hz also at multiples of 50 Hz all the way up to 600 Hz (Nyquist frequency), yielding as many as 12 harmonics. In addition, noise removal at 25 Hz (beta range) can also often be observed. We noticed that Zapline performed worse with data at higher sampling rates, due to the increased complexity of the data. Thus, to make Zapline's task easier, it is advisable to downsample the data prior to running Zapline-plus. For the MEG data analyzed here, we downsampled it to 350 Hz, for the EEG data to 250 Hz, such that only 50 and 100 Hz, and 150 Hz for the MEG data, are considered for noise removal. Indeed, we found that Zapline-plus performed much better at lower sampling rates.

3 | EXAMPLE APPLICATIONS

3.1 | Data sets

In order to test the efficacy of the Zapline-plus algorithm, we ran it on four different openly available data sets, two EEG data sets containing both stationary and mobile conditions, and two stationary MEG data sets. Notably, line noise is usually extremely strong in MEG, despite extensive shielding of the equipment that is commonly applied.

3.2 | EEG study I

This is an open data set available at https://openneuro.org/datasets/ds003620/versions/1.0.2 (Liebherr et al., 2021). Data of 41 participants (aged 18–39 years, M = 23.1 years, 26 female and 15 male) is

available, of which we only used 24 sets for technical reasons. The experiment consisted of an auditory oddball task which was administered either in a laboratory environment, or on a grass field, or on the campus of the University of South Australia. Continuous EEG data was recorded with a 500 Hz sampling rate using 32 active Ag/AgCl electrodes and the BrainVision LiveAmp (Brain Products GmbH, Gilching, Germany). Electrode impedances were kept below 20 k Ohm and channels were referenced to the FCz electrode. See Liebherr et al. (2021) for details.

3.3 | EEG study II

This is an open data set available at https://doi.org/10.14279/depositonce-10493 (Gramann, Hohlefeld, Gehrke, & Klug, 2021). Data of 19 participants (aged 20–46 years, mean 30.3 years, 10 female and 9 male) are available, which we all used. The experiment consisted of a rotation on the spot, which either happened in a virtual reality environment with physical rotation or in the same environment on a two-dimensional monitor using a joystick to rotate the view. EEG data for each condition was recorded with a 1,000 Hz sampling rate using 157 active Ag/AgCl electrodes (129 on the scalp in a custom equidistant layout, 28 around the neck in a custom neckband) and the BrainAmp Move System (Brain Products GmbH, Gilching, Germany). Electrode impedances were kept below 10kΩ for scalp electrodes and below 50k Ohm for neck electrodes, and channels were referenced to the FCz electrode. See Gramann et al. (2021) for details.

3.4 | MEG study I

This open data set is available at https://data.donders.ru.nl/collections/di/dccn/DSC_3011020.09_236?0 (Schoffelen et al., 2019). We randomly selected 12 of the 204 subjects to test Zapline-plus. Subjects performed a language task, during which they had to process linguistic utterances that either consisted of normal or scrambled sentences. Four of the analyzed subjects were reading the stimuli (subject IDs V1001, V1012, V1024, V1036), the other eight listened to the stimuli (subject IDs A2027, A2035, A2051, A2064, A2072, A2088, A2101, A2110). Magnetoencephalographic data were collected with a 275-channel axial gradiometer system (CTF). The MEG recording for each subject lasted about 45 min. The signals were digitized at a sampling frequency of 1,200 Hz (cutoff frequency of the analog anti-aliasing low pass filter was 300 Hz). See Schoffelen et al. (2019) for details.

3.5 | MEG study II

This data set comprises open MEG data from the Cam-CAN set of the Cambridge Centre for Aging and Neuroscience, available at http://www.mrc-cbu.cam.ac.uk/datasets/camcan (Shafto et al., 2014; Taylor et al., 2017). We randomly selected 23 of the 647 participants. Participants performed a sensory-motor task on audio-visual stimuli (bilateral sine gratings and concurrent audio tone). Participants were asked to respond each time a stimulus was presented. The task lasted for 8 min and 40 s. Magnetoencephalographic data were collected with a 306-channel Elekta Neuromag Vectorview (102 magnetometers and 204 planar gradiometers) at a sampling rate of 1,000 Hz (bandpass 0.03–330 Hz). Only planar gradiometers were used in the analysis. See Shafto et al. (2014) and Taylor et al. (2017) for details.

3.6 | Processing

The following preprocessing steps were applied: removal of excess channels, resampling to 250/350 Hz (for the EEG and MEG sets, respectively), and merging of all conditions per study (EEG study II only). First, to test the different elements of the algorithm, we ran eight different sets of settings on EEG study II (which contained complex artifacts that differed between the two conditions):

1. Using a fixed removal of three components and no chunks, corresponding to standard Zapline use.
2. Using a fixed removal, but chunking the data into 150 s segments.
3. Using the automatic detector of noise components, but no chunks.
4. Combining 150 s chunks and automatic noise component detector.
5. Using 150 s chunks with individual peak detection and automatic noise component detector.
6. Using 150 s chunks without peak detection and automatic noise component detection with adaptive changes for over- or undercleaning.
7. Using 150 s chunks with individual peak detection, as well as automatic detection with adaptive changes
8. Using all features (default): adaptive chunk length with individual peak detection, as well as automatic detection with adaptive changes.

All conditions used the automatic detector of noise frequencies. With this approach, we tried to mimic the creation of the algorithm with successive improvements.

Subsequently, we ran Zapline-plus additionally on EEG study I and on the MEG studies. For EEG study I, we used only default values, for the MEG studies we set "noisefreqs" to 50 as we expected only line noise and wanted to prevent false-positive noise frequency detection due to very strong (sub-)harmonics of the line frequency.

4 | RESULTS

Overall, the cleaned spectra show that Zapline-plus successfully removed the strong line noise peaks while introducing only minimal notches. The results of the cleaning of all example studies are depicted in Figure 4, and Table 1 lists the results for analytics for the cleaning using successively enabled features for EEG study II (the

TABLE 1 Algorithm steps applied to an example data set

	1. Original Zapline	2. Fixed chunks	3. Auto comp. Detection	4. Auto + fixed chunks
Final # of removed components	3 (0)	3 (0)	4.55 (1.93)	5.76 (1.61)
Clean ratio nois/surroundings	1.22 (0.21)	1.09 (0.12)	1.08 (0.04)	0.97 (0.07)
% removed power below noise	1.73 (0.57)	1.87 (0.46)	3.05 (2.10)	4.30 (1.36)
% below lower threshold	0 (0)	3.36 (6.67)	0.20 (0.63)	20.08 (16.78)
% above upper threshold	23.13 (11.37)	16.79 (13.67)	14.74 (9.94)	6.82 (7.06)
	5. Auto + fixed chunks + peaks	**6. Auto + adaptive + fixed chunks**	**7. Auto + adaptive + fixed chunks + peaks**	**8. Auto + adaptive + variable chunks + peaks**
Final # of removed components	4.08 (1.09)	3.63 (0.88)	3.31 (0.92)	3.20 (0.78)
Clean ratio noise/surroundings	1.00 (0.05)	1.01 (0.05)	1.03 (0.04)	1.03 (0.08)
% removed power below noise	3.01 (0.95)	2.41 (0.72)	2.20 (0.69)	2.12 (1.92)
% below lower threshold	9.31 (12.84)	5.26 (11.01)	2.94 (6.51)	0.36 (1.44)
% above upper threshold	7.25 (7.90)	8.33 (6.85)	8.37 (7.47)	5.64 (10.81)

Note: Analytics (mean and *SD*) when using varying features enabled during cleaning of EEG study II. The removed power below noise refers to −11 Hz to −1 Hz below the detected noise frequency, the percentage below/above thresholds refer to the proportion of samples in the spectrum exceeding the thresholds for fine-grained adaptation. Although they were not always used, they are always available for analysis. The values are first averaged over all detected noise frequencies per subject. "1. original Zapline" refers to the basic fixed version of Zapline, "2. fixed chunks" refers to applying the basic Zapline on regular 150 s chunks, "3. auto comp. detection" refers to using automatic detection of components to remove, "4. auto + fixed chunks" refers to using automatic noise component detection on regular 150 s chunks, "5. auto + fixed chunks + peaks" refers to using automatic noise component detection on regular 150 s chunks with individual chunk noise peak detection, "6. auto + adaptive + fixed chunks" refers to using automatic noise detection on regular 150 s chunks with adaptive detection strength, "7. auto + adaptive + fixed chunks + peaks" refers to using automatic noise component detection on regular 150 s chunks with individual peak detection and adaptive detection strength, and "8. auto + adaptive + variable chunks + peaks" refers to using automatic noise component detection on automatically detected variable chunks with individual peak detection and adaptive detection strength (see also Section 3.6). $N = 19$.

TABLE 2 Analytics results of the cleaning of four openly available data sets (mean and *SD*)

	EEG study I ($N = 24$)	EEG study II ($N = 19$)	MEG study I ($N = 12$)	MEG study II ($N = 23$)
Final *SD* of detector	2.63 (0.18)	3.10 (0.40)	3.42 (0.73)	3.38 (0.61)
Final # of removed components	2.83 (0.95)	3.20 (0.78)	17.18 (8.62)	8.21 (3.66)
Raw ratio noise/surroundings	6.99 (6.26)	2.40 (1.91)	962.6 (1799.6)	232.6 (369.4)
Clean ratio noise/surroundings	1.22 (0.17)	1.03 (0.08)	1.32 (0.61)	1.00 (0.05)
Percentage of removed power below noise	6.20 (2.60)	2.12 (1.92)	−31.34 (75.71)	3.52 (1.38)

Note: The removed power below noise refers to −11 Hz to −1 Hz below the detected noise frequency. For EEG study II the values are first averaged over all detected noise frequencies per subject, the other studies had only 50 Hz line noise removed.

number of removed components per cleaning step, the ratio of noise/surroundings after cleaning, the proportion of removed power below noise, and the proportion of frequency samples below and above the adaptation threshold). Only EEG study II had noise frequencies different from line, which is why we specifically show the raw and clean 50 Hz / surroundings power ratios. Table 2 shows the results for the four example data sets (the final *SD* value for detection, the number of removed components per cleaning step, the ratio of noise/surroundings before and after cleaning, and the proportion of removed power below noise).

4.1 | Suboptimal case results

Viewing only the average results of the final cleaning, however, yields only a limited understanding of the detailed processes. Some data sets had less-than-ideal results, for example, they showed a distortion of the spectrum below noise such that the power was actually increased. This could be seen mostly in data sets with particularly strong noise contamination, especially in MEG study I where four data sets had more than 800 times stronger power at noise frequency than surroundings, up to almost 7,000 times for the noisiest data set (Figure 4, MEG study I, right panel). All these four data sets, but only them, exhibited a negative removal of power below noise, that is, an increase of power in the cleaned data, and they drive the average that can be seen in Table 2 and Figure 4, MEG study I, left panel (green line above black). Also, while all data sets showed a reduction in power of the noise, some of them had comparably strong residual noise peaks (ratios of noise/surroundings above 1.2, these usually also had very high ratios before cleaning), indicating that Zapline-plus could not fully clean these data sets.

4.2 | Automatic segmentation

In order to understand and evaluate the automatic segmentation, we investigated the algorithm's tolerance against violations of the stability assumption of the spatial filter. In Figure 5, we show a single case example of the internal workings of the automatic segmentation, visualizing the distances of the covariance matrices of the narrowband-filtered signal as well as the resulting segmentation points (Section 2.2). Large peaks in this signal correspond to violations of the stability assumption, and the effectiveness of a spatial filter applied to ranges in which several such large peaks occur is limited. Evidently, such violations of the assumption can exist, underlining the need for segmentation of the data. The minimal chunk length prevents some of the peaks to be taken as segmentation points, however, all relevant major peaks are correctly detected and used as the indication of a new segment. Note that this is only an example and cannot be assumed to work identically for all data sets, it is only meant to give an impression of the algorithm.

5 | DISCUSSION

In this work, we extended Zapline to allow fully automatic removal of line noise and other spectral peaks, while giving the user a maximum of flexibility and information, as well as allowing complete replicability of the processing. We evaluated the algorithm on two EEG and two MEG data sets. First, we checked whether the different parts of the algorithm improved the cleaning on one EEG study, then we applied the final default values to the three other data sets. Taken together, the results show that the new features allow for fully automatic noise removal and make the algorithm applicable for different kinds of electrophysiological data, resulting in a substantial decrease of frequency-specific noise with a minimal negative impact on true neural activity.

5.1 | Efficacy of the algorithm

Examination of the algorithm components on EEG study II showed that they do improve the results. However, the improvement is not a simple linear relationship. Both, using fixed 150 s chunks, and using automatic detection of to-be-removed components improved the clean ratio of noise/surrounding power similarly to using the standard fixed approach. In doing so, using auto-detection affected the power below noise frequencies (−11 to −1 Hz) more than chunks did, but chunks had a larger proportion of samples below the threshold directly at the noise frequency, meaning chunks introduced a slight notch into the spectrum, whereas auto cleaning without chunks distorted the spectrum more generally. Interestingly, combining these two approaches led to the lowest ratio of noise/surroundings power while also introducing substantial amounts of overcleaning, both in terms of general distortion (% removed below noise) and a notch (% below lower threshold). This combination also had the fewest samples above the adaptation threshold, corresponding to the low noise/surroundings ratio.

The strong overcleaning effect can be explained by the fact that not all noise oscillations were present in all chunks. Although the automatic detection of components to remove should be able to select fewer samples with less noise, it requires some sort of "knee-point" or "corner" in the artifact scores. In chunks with no oscillation in the given noise frequency, the scores exhibit an almost linear decrease, which can lead to erroneously removing large numbers of components. This negative interaction effect can be fixed by either adapting the *SD* level the detector uses, or by simply not using auto-detection when no noise is present. Using either improvement alone led to similar levels of cleaning in terms of noise/surroundings power as well as % of samples above threshold, while the adaptive cleaning had a slightly reduced impact on the spectrum below noise and a reduced notch. Combining all options, chunks with individual peak detection, as well as automatic detection with adaptation, led to even better overall results.

Finally, adding the adaptive variable chunk length based on the spatial stability of the noise (using the full feature set of the algorithm) improved the specificity of the cleaning even further. This combination had a lower % of samples below and above the adaptation threshold and a lower impact on the spectrum below noise. Overall, the combination of all features of the algorithm successfully cleans the data, while keeping the distortions to the spectrum as low as possible.

Applying this final combination to all example data sets led to substantial improvements in the spectra. In EEG study I, there was 50 Hz line noise present in the data, and an unknown oscillation at around 7 Hz, plus harmonics. The former was detected and successfully cleaned by Zapline-plus, whereas the latter was too small to be detected. EEG study II is a particularly heavily contaminated study, as can be seen by the various peaks in the spectrum. However, Zapline-plus was able to successfully clean these data, not only at line noise, but also all other strong peaks. This example emphasizes the importance of the automatic noise frequency detector, as these oscillations are difficult to anticipate.

Applying Zapline-plus on the MEG studies shows that even extremely noisy data is successfully cleaned. It can be seen in MEG study I, however, that Zapline-plus may have an impact on the overall spectrum by increasing the broadband power. This effect is driven by four of the 12 data sets, which show extreme levels of noise before cleaning, the other eight do not show such an increase. In these cases, the actual impact of the cleaning on final measures must be closely examined in order to decide whether the trade-off of reduced noise versus spectrum distortions is worth it in this particular analysis or if the cleaning must be adapted.

5.2 | Other notes

In EEG study II, it was clearly visible that some noise frequencies were only present in the first or second part of the data (body vs. joystick rotation, see Figure 3 for an example of a noise frequency only present in the second half). This underlines the importance of the

chunking and individual frequency detection, as this allows checking whether the oscillation is actually present in that chunk and prevents overcleaning. We would also like to point out the importance of fine-tuned noise frequency detection for some frequencies, especially the one has seen in Figure 3. The separation of 20.9 Hz and, subsequently, 21.1 Hz noise is important as the two frequencies cannot be cleaned together. This would be impossible to see without a high resolution of the frequency spectrum, and simply cleaning with a fixed 21 Hz setting does lead to subpar results. Also, as shown in Figure 2, the peak frequency of the line noise is not always stationary and Zapline-plus is able to detect these variations. On a final note, although Zapline attempts to preserve the power and phase angle of physiological signals that remain at the cleaned frequencies, analyzing and interpreting these should always be done with care.

5.3 | Limitations

As we showed, the cleaning is not always perfect. Especially with data that is heavily contaminated with noise, it is possible to (a) change the spectrum below the noise frequency such that the power is actually increased, (b) leave residual noise in the data, or (c) after cleaning, leave a small notch in the spectrum. Although the default values of the algorithm are chosen to fit most of the data sets, in some cases it might be better to adjust them according to the results obtained from the automatic cleaning and then re-run Zapline-plus. The user is strongly advised to always check the resulting analytics plots after applying Zapline-plus.

5.4 | Future directions

It might also be that no matter the parameter adjustment, the cleaning will remain suboptimal. In these cases, it could be useful to combine Zapline-plus with CleanLine, since these two methods rely on distinct, complementary algorithms to isolate and remove line noise. Zapline, on the one hand, applies a fixed spatial filter over the entire data segment, allowing it to account for variations in noise amplitude in the temporal domain, but strictly not changes in noise topography. CleanLine, in contrast, removes a fixed oscillatory noise signal in the time-domain data in each channel separately, allowing full flexibility in the spatial, but not the temporal domain. Indeed, a recent paper shows that combining the two methods can improve the cleaning of heavily contaminated data (Miyakoshi et al., 2021). Examining the possibility of an automatic extra CleanLine step if Zapline-plus alone yielded sub-optimal results would be an option for future investigations.

Another interesting possibility is to visualize the topographies of the removed artifacts. As Zapline internally uses spatial filters, these can be visualized like any other spatial filter and be added to the analytics information feedback for the user. However, this is not straightforward as Zapline-plus specifically uses different spatial decompositions and a different number of removed components for each chunk. Still, if the filters vary only slightly, visualizing the average of the removed topographies could be valuable feedback.

Finally, it could be explored whether Zapline-plus can also be used for other applications. For example, some of our tests suggest that one could remove very regular mechanical walking artifacts in mobile EEG studies, or the steps could be extracted to create events for subsequent analysis. Another option would be to extract alpha oscillations (8–13 Hz) that exceed the $1/f$ background activity. This topic has already been mentioned in the original Zapline paper (de Cheveigné, 2020), but with a focus on removing alpha for other analyses. Extracting only the oscillatory alpha time series by switching the "clean" with the "noise" data could result in more specific alpha signals than using a standard band-pass filter. In sum, Zapline-plus is essentially a tool created for noise removal, but it can also be used to extract all kinds of oscillatory activity to be used in other analyses, which makes it a versatile tool in any analysis pipeline.

5.5 | Implications for the field

Removing line noise is an undeniably important part of electrophysiological data processing, and having the option to do so without risking the analysis of potentially important frequencies while retaining full data rank is a valuable tool. The newly added features of fully automatic and documented processing including the detection of noise oscillations are especially important considering the current trend towards complete automatic processing pipelines (Bigdely-Shamlo, Mullen, Kothe, Su, & Robbins, 2015; da Cruz, Chicherov, Herzog, & Figueiredo, 2018; Gabard-Durnam, Leal, Wilkinson, & Levin, 2018; Pedroni, Bahreini, & Langer, 2019) and the need for more rigorous methods in neurophysiological analysis (Cohen, 2017) due to the replication crisis (Open Science Collaboration, 2015). Also, although the impact of preprocessing has been investigated in parts (Robbins, Touryan, Mullen, Kothe, & Bigdely-Shamlo, 2020), and some pipelines create comprehensive documentation of their processes, a documentation of the line noise removal as detailed as provided by Zapline-plus is lacking thus far. Zapline-plus contributes to the field by making the removal of line noise and other oscillation artifacts in large data sets automatic, easy, transparent, and reproducible, while limiting its potential negative impact on downstream analysis. It can easily be integrated into any automatic processing pipeline.

ACKNOWLEDGMENTS

We are thankful to the researchers who made the data sets freely available and to Alain de Cheveigné who kindly allowed the adaptation and re-hosting of parts of the original Zapline code. We would also like to thank the members of the Berlin Mobile Brain/Body Imaging Lab of Prof. Klaus Gramann for valuable discussions throughout the development of the algorithm. We acknowledge the support of this work by the DFG (GR2627/8-1) and USAF (ONR 10024807). Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST

The authors have declared no conflicts of interest for this article.

DATA AVAILABILITY STATEMENT

The data used in this study is available for download as laid out in the Data sets section. The MATLAB source code of the software is available for download at https://github.com/MariusKlug/zapline-plus.

ORCID

Marius Klug https://orcid.org/0000-0001-8667-3457
Niels A. Kloosterman https://orcid.org/0000-0002-1134-7996

REFERENCES

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*, 16. https://doi.org/10.3389/fninf.2015.00016

Cohen, M. (2021). *Linear algebra: Theory, intuition, code*. Amsterdam: Sincxpress BV.

Cohen, M. X. (2017). Rigor and replication in time-frequency analyses of cognitive electrophysiology data. *International Journal of Psychophysiology*, *111*, 80–87. https://doi.org/10.1016/j.ijpsycho.2016.02.001

da Cruz, J. R., Chicherov, V., Herzog, M. H., & Figueiredo, P. (2018). An automatic pre-processing pipeline for EEG analysis (APP) based on robust statistics. *Clinical Neurophysiology*, *129*, 1427–1437. https://doi.org/10.1016/j.clinph.2018.04.600

de Cheveigné, A. (2020). ZapLine: A simple and effective method to remove power line artifacts. *NeuroImage*, *207*, 116356. https://doi.org/10.1016/j.neuroimage.2019.116356

de Cheveigné, A., & Parra, L. C. (2014). Joint decorrelation, a versatile tool for multichannel data analysis. *NeuroImage*, *98*, 487–505. https://doi.org/10.1016/j.neuroimage.2014.05.068

de Cheveigné, A., & Simon, J. Z. (2007). Denoising based on time-shift PCA. *Journal of Neuroscience Methods*, *165*, 297–305. https://doi.org/10.1016/j.jneumeth.2007.06.003

Dehais, F., Duprès, A., Blum, S., Drougard, N., Scannella, S., Roy, R. N., & Lotte, F. (2019). Monitoring Pilot's mental workload using ERPs and spectral power with a six-dry-electrode EEG system in real flight conditions. *Sensors*, *19*, 1324. https://doi.org/10.3390/s19061324

Dehais, F., Lafont, A., Roy, R., & Fairclough, S. (2020). A Neuroergonomics approach to mental workload, engagement and human performance. *Frontiers in Neuroscience*, *14*, 268. https://doi.org/10.3389/fnins.2020.00268

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dimigen, O. (2020). Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. *NeuroImage*, *207*, 116117. https://doi.org/10.1016/j.neuroimage.2019.116117

Gabard-Durnam, L. J., Leal, A. S. M., Wilkinson, C. L., & Levin, A. R. (2018). The Harvard automated processing pipeline for electroencephalography (HAPPE): Standardized processing software for developmental and high-artifact data. *Frontiers in Neuroscience*, *12*, 97. https://doi.org/10.3389/fnins.2018.00097

Gehrke, L., & Gramann, K. (2021). Single-trial regression of spatial exploration behavior indicates posterior EEG alpha modulation to reflect egocentric coding. *The European Journal of Neuroscience*, *54*, 8318–8335. https://doi.org/10.1111/ejn.15152

Gramann, K., Ferris, D. P., Gwin, J., & Makeig, S. (2014). Imaging natural cognition in action. *International Journal of Psychophysiology*, *91*, 22–29. https://doi.org/10.1016/j.ijpsycho.2013.09.003

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T.-P., Lin, C.-T., … Makeig, S. (2011). Cognition in action: Imaging brain/body dynamics in mobile humans. *Reviews in the Neuroscience*, *22*, 593–608. https://doi.org/10.1515/rns.2011.047

Gramann, K., Hohlefeld, F. U., Gehrke, L., & Klug, M. (2021). Human cortical dynamics during full-body heading changes. *Scientific Reports*, *11*, 18186. https://doi.org/10.1038/s41598-021-97749-8

Hämäläinen, M., Hari, R., Ilmoniemi, R. J., Knuutila, J., & Lounasmaa, O. V. (1993). Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, *65*, 413–497. https://doi.org/10.1103/revmodphys.65.413

Hoogenboom, N., Schoffelen, J.-M., Oostenveld, R., Parkes, L. M., & Fries, P. (2006). Localizing human visual gamma-band activity in frequency, time and space. *NeuroImage*, *29*(764), 773. https://doi.org/10.1016/j.neuroimage.2005.08.043

Hyvarinen A. (1997). A family of fixed-point algorithms for independent component analysis. In *Proccedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 21–24 April 2017, Munich, Germany, 3917–3920. doi:https://doi.org/10.1109/icassp.1997.604766

Jungnickel, E., Gehrke, L., Klug, M., & Gramann, K. (2019). MoBI—Mobile brain/body imaging. *Neuroergonomics*, *2019*, 59–63. https://doi.org/10.1016/b978-0-12-811926-6.00010-5

Kloosterman, N. A., de Gee, J. W., Werkle-Bergner, M., Lindenberger, U., Garrett, D. D., & Fahrenfort, J. J. (2019). Humans strategically shift decision bias by flexibly adjusting sensory evidence accumulation. *eLife*, *8*, e37321. https://doi.org/10.7554/elife.37321

Klug, M., & Gramann, K. (2020). Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *The European Journal of Neuroscience*, *54*(12), 8406–8420. https://doi.org/10.1111/ejn.14992

Leske, S., & Dalal, S. S. (2019). Reducing power line noise in EEG and MEG data via spectrum interpolation. *NeuroImage*, *189*, 763–776. https://doi.org/10.1016/j.neuroimage.2019.01.026

Liebherr, M., Corcoran, A. W., Alday, P. M., Coussens, S., Bellan, V., Howlett, C. A., … Bornkessel-Schlesewsky, I. (2021). EEG and behavioral correlates of attentional processing while walking and navigating naturalistic environments. *Scientific Reports*, *11*, 22325. https://doi.org/10.1038/s41598-021-01772-8

Makeig, S., Gramann, K., Jung, T.-P., Sejnowski, T. J., & Poizner, H. (2009). Linking brain, mind and behavior. *International Journal of Psychophysiology*, *73*, 95–100. https://doi.org/10.1016/j.ijpsycho.2008.11.008

Miyakoshi, M., Schmitt, L. M., Erickson, C. A., Sweeney, J. A., & Pedapati, E. V. (2021). Can we push the "quasi-perfect artifact rejection" even closer to perfection? *Frontiers in Neuroinformatics*, *14*, 597079. https://doi.org/10.3389/fninf.2020.597079

Open Science Collaboration. (2015). Psychology. Estimating the reproducibility of psychological science. *Science (New York, N.Y.)*, *349*, eaac4716. https://doi.org/10.1126/science.aac4716

Pedroni, A., Bahreini, A., & Langer, N. (2019). Automagic: Standardized preprocessing of big EEG data. *NeuroImage*, *200*, 460–473. https://doi.org/10.1016/j.neuroimage.2019.06.046

Raja, P., & Matthew, R. (2009). *Neuroergonomics: The brain at work*. New York, NY: Oxford University Press.

Robbins, K. A., Touryan, J., Mullen, T., Kothe, C., & Bigdely-Shamlo, N. (2020). How sensitive are EEG results to preprocessing methods: A benchmarking study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*, 1081–1090. https://doi.org/10.1109/tnsre.2020.2980223

Schoffelen, J.-M., Oostenveld, R., Lam, N. H. L., Uddén, J., Hultén, A., & Hagoort, P. (2019). A 204-subject multimodal neuroimaging dataset to study language processing. *Scientific Data*, *6*, 17. https://doi.org/10.1038/s41597-019-0020-y

Shafto, M. A., Tyler, L. K., Dixon, M., Taylor, J. R., Rowe, J. B., Cusack, R., … Cam-CAN. (2014). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: A cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurology*, *14*, 204. https://doi.org/10.1186/s12883-014-0204-1

Taylor, J. R., Williams, N., Cusack, R., Auer, T., Shafto, M. A., Dixon, M., … Henson, R. N. (2017). The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. *NeuroImage*, *144*, 262–269. https://doi.org/10.1016/j.neuroimage.2015.09.018

Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, *15*, 70–73. https://doi.org/10.1109/tau.1967.1161901

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data—A practical approach. *Journal of Neuroscience Methods*, *250*(34), 46. https://doi.org/10.1016/j.jneumeth.2014.08.002

Wunderlich, A., & Gramann, K.. (2018). Electrocortical evidence for long-term incidental spatial learning through modified navigation instructions. In *Proceedings of the 11th International Conference on Spatial Cognition 2018*, September 5–8, 2018, Tübingen, Germany. doi:https://doi.org/10.1101/280842

How to cite this article: Klug, M., & Kloosterman, N. A. (2022). Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG. *Human Brain Mapping*, *43*(9), 2743–2758. https://doi.org/10.1002/hbm.25832

2 Obtaining an optimal ICA

Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments

Klug, M., & Gramann, K. (2021). Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *The European Journal of Neuroscience*, 54(12), 8406–8420. https://doi.org/10.1111/ejn.14992.

Received: 29 May 2020 | Revised: 28 August 2020 | Accepted: 23 September 2020

DOI: 10.1111/ejn.14992

SPECIAL ISSUE ARTICLE

EJN European Journal of Neuroscience FENS WILEY

Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments

Marius Klug[1] | **Klaus Gramann**[1,2,3]

[1]Biopsychology and Neuroergonomics, Institute of Psychology and Ergonomics, TU Berlin, Berlin, Germany

[2]Center for Advanced Neurological Engineering, University of California San Diego, La Jolla, CA, USA

[3]School of Computer Science, University of Technology Sydney, Ultimo, NSW, Australia

Correspondence
Marius Klug, Biopsychology and Neuroergonomics, TU Berlin, Berlin 10623, Germany.
Email: marius.klug@tu-berlin.de

Funding information
This work was supported by the DFG (GR2627/8-1) and USAF (ONR 10024807).

Abstract

Recent developments in EEG hardware and analyses approaches allow for recordings in both stationary and mobile settings. Irrespective of the experimental setting, EEG recordings are contaminated with noise that has to be removed before the data can be functionally interpreted. Independent component analysis (ICA) is a commonly used tool to remove artifacts such as eye movement, muscle activity, and external noise from the data and to analyze activity on the level of EEG effective brain sources. The effectiveness of filtering the data is one key preprocessing step to improve the decomposition that has been investigated previously. However, no study thus far compared the different requirements of mobile and stationary experiments regarding the preprocessing for ICA decomposition. We thus evaluated how movement in EEG experiments, the number of channels, and the high-pass filter cutoff during preprocessing influence the ICA decomposition. We found that for commonly used settings (stationary experiment, 64 channels, 0.5 Hz filter), the ICA results are acceptable. However, high-pass filters of up to 2 Hz cut-off frequency should be used in mobile experiments, and more channels require a higher filter to reach an optimal decomposition. Fewer brain ICs were found in mobile experiments, but cleaning the data with ICA has been proved to be important and functional even with low-density channel setups. Based on the results, we provide guidelines for different experimental settings that improve the ICA decomposition.

KEYWORDS

artifact removal, electroencephalogram, independent component analysis, mobile brain/body imaging, preprocessing

Abbreviations: ASR, artifact subspace reconstruction; EEG, electroencephalography; EOG, electrooculography; ERP, event-related-response; EXG, electrooculography and electrocardiography; IC, independent component; ICA, independent component analysis; MEG, magnetoencephalography; MoBI, Mobile Brain/Body Imaging; RV, residual variance; SNR, signal-to-noise ratio; VR, virtual reality.

Edited by Martin Seeber

1 | INTRODUCTION

Over the last decade, the development of lightweight portable electroencephalography (EEG) amplifiers and new data-driven analyses approaches led to the investigation of the neural basis of ecologically valid cognitive processes in actively behaving human participants outside established laboratory environments. Experiments now allow active behavior of participants both in the lab (De Sanctis et al., 2014; Djebbara et al., 2019; Ehinger et al., 2014; Gehrke et al., 2018; Gramann et al., 2010; Nenna et al., 2020, this issue) and in the real world, which increases our understanding of human brain dynamics accompanying embodied cognitive processes as well as the impact of real world environments (Debener et al., 2012; Ladouce et al., 2017; Protzak & Gramann, 2018; Wascher et al., 2014; Wunderlich & Gramann, 2018). While these experimental protocols provide new insights into the neural activity subserving cognition in more realistic and natural settings, they present new challenges. Mobile EEG or Mobile Brain/Body Imaging (MoBI; Gramann et al., 2011, 2014; Jungnickel et al., 2018; Makeig et al., 2009) recordings are impacted by movement-related electrical activity stemming from facial muscles, neck muscles and eye movements that naturally accompany active behaviors. While these physiological contributions are usually considered to be artifacts, they may still provide additional insights if analyzed separately. Other artifactual contributions to the recording are even less welcome. For example movement in mobile protocols might lead to mechanical artifacts like cable sway or micro movement of electrodes that contribute artifactual activity into the recording. Finally, environmental sources and the equipment necessary for the experiment itself like head mounted virtual reality (VR) systems or treadmills can be another unavoidable source of electrical artifacts in mobile recordings. All these signals mix at the sensor level and render it difficult to dissociate brain from non-brain activity to investigate the neural basis of the cognitive processes of interest. While movement-related non-brain sources are specifically problematic for experiments including actively behaving participants, contributions from sources like eye movements, facial muscles, and neck muscle activity can also be found in EEG data recorded in established desktop settings. These forms of biological signals are traditionally considered artifacts and are one of the main reasons why established EEG research minimizes any kind of participant movement, including eye movements or blinks. Thus, the ability to interpret EEG data from both classic stationary as well as MoBI experiments depends greatly on the ability to dissociate signals of interest originating in the brain from those of other sources.

Mechanical and electrical artifacts do not correlate highly with physiological recordings and thus are typically easier to detect and to remove than physiological contributions (Chang et al., 2020). The dissociation of potentially correlating physiological sources (brain, eyes, and muscles) is more difficult. It can be achieved by applying spatial filter methods to the data, exploiting the fact that electrical activity is recorded with multiple electrodes on the scalp. Among different spatial filter approaches, blind source separation techniques (Bell & Sejnowski, 1995; Hyvärinen et al., 2001; Makeig et al., 1997) proved to be very successful and specifically independent component analysis (ICA) applied to EEG and magnetoencephalography (MEG) data demonstrated increasing popularity among researchers. With the number of ICA applications to EEG data constantly growing over the last 25 years from 16 publications in 1995 to 5,450 publications in 2019 alone (search term "EEG" + "Independent component analysis", Google Scholar, accessed on 2020-05-18), the variations of preprocessing the data to eventually applying ICA also increased. In most cases a channel density of 64 and upwards is being used for ICA since spatial filtering typically improves with more degrees of freedom, but less consensus is reached considering the applied filter. Often a high-pass filter of 1 Hz is used, but lower frequencies like 0.5 Hz or higher ones like 2 Hz or even 3 Hz can be found in the literature as well. Sometimes, additional low-pass filters are applied while many times none is used. While some of the preprocessing steps have been evaluated regarding their impact on the subsequent ICA decomposition, not all factors have been systematically investigated. Quantitative validation of different ICA algorithms and their efficacy in separating brain from other data was often done with simulated data, since a ground truth for signal and noise is available in that case. However, simulated data are cleaner than natural data and cannot reflect the true complexity of artifacts and the intricate variations in physiological activity occurring in real experiments. Researchers working with natural EEG data need to understand the effects of different preprocessing steps on this data and the subsequent ICA decomposition. Consequently, the purpose of this study is to shed light on the relevant contributions of different factors on ICA for both stationary and mobile experiments using natural data, and to provide a "best practices" guideline to improve the ICA decomposition.

In this paper, we will first introduce the EEG mixing model and discuss prior research on the effect of different data preprocessing settings on ICA. Formulating our hypotheses, we will then present our approach in investigating the impact of the three most common factors influencing ICA decompositions: high-pass filter settings, channel density, and movement, by evaluating ICA decompositions with respect to the number of components categorized as brain or non-brain origin, independent component (IC) dipolarity, and the signal-to-noise ratio (SNR) of event-related potentials (ERPs). Finally, the results will be discussed and recommendations will be given.

1.1 | The EEG linear mixing model

Analyzing EEG data with ICA is based on a general assumption that the data matrix $X \in \mathbb{R}^{N \times M}$ recorded by the EEG electrodes is a linear mixture of different sources $S \in \mathbb{R}^{N \times M}$ with a mixing matrix $A \in \mathbb{R}^{N \times N}$ such that $X = AS$ (N being both the number of sources and EEG channels, and M being the number of samples in the dataset; Hyvärinen et al., 2001; Hyvärinen & Oja, 2013). Sources are assumed to be statistically independent and stationary. These assumptions can now be leveraged to compute an inverse un-mixing matrix $W = A^{-1}$ ($W \in \mathbb{R}^{N \times N}$), such that $S = WX$. Finding W is an ill-posed problem without an analytical solution. Different ICA algorithms use different heuristcs and thus compute slightly different un-mixing matrices (Hyvärinen et al., 2001), and even the same algorithm does not always converge on the same solution for the same data (Artoni et al., 2014; Groppe et al., 2009).

1.2 | Achieving an optimal decomposition

Since ICA is becoming increasingly popular for EEG research, efforts have been made to identify the best algorithms and prerequisites to obtain a good decomposition of the data. Comparing different algorithms, Delorme et al. (2012) and Leutheuser et al. (2013) found that AMICA (Palmer et al., 2011) performed best among different algorithms. This was confirmed in part by (Zakeri et al., 2014), but it was concluded that the choice of preprocessing was more relevant to the decomposition quality than the algorithm itself.

Already early work on ICA has found the choice of preprocessing to be relevant, as "[t]he success of ICA for a given data set may depend crucially on performing some application-dependent preprocessing steps" (Hyvärinen et al., 2001, p. 263). One often used method to improve the decomposition quality is that of high-pass filtering. High-pass filtering is essentially another linear transformation of the data and thus does not violate the ICA assumptions, as it can be expressed by multiplying the first equation with a component-wise filtering matrix F from the right such that $X_{filtered} = XF = ASF = AS_{filtered}$. It is thus possible to compute the mixing matrix A on filtered data and apply it to the unfiltered data without modification (see also Hyvärinen et al., 2001; Winkler et al., 2015), which is common practice in ICA analysis. While low-pass as well as high-pass filtering may remove noise from the data, filtering also bears the risk of removing relevant information. For low-pass filtering this is the case especially for high-frequency activity stemming from muscle contractions while for high-pass filtering this concerns slow cortical potentials in the data. Noise in form of slow drifts in the data affecting multiple channels (e.g. from cable sway or strong sweating) often occurs in all or many EEG channels and is thus hard to separate from brain signals (Winkler et al., 2015). Removing slow drifts can thus benefit the decomposition. While being used in practice almost universally as a preprocessing step before ICA, the exact filter specifications, especially the cut-off frequency of the high-pass filter, are not always agreed upon.

Several studies have investigated the effect of high-pass filtering on the ICA decomposition. Groppe et al. (2009) have found that removing the mean of epoched data (which acts as a leaky high-pass filter) resulted in a more reliable decomposition. Following up on this result, Zakeri et al. (2014) compared the effects of filtering, epoching, de-meaning, and including electrooculography and electrocardiography channels (EXG) on the ICA decomposition. By assessing IC dipolarity and mutual information, the authors found that the best approach was to compute the ICA on filtered continuous data including EXG. However, the applied filter was a band-pass filter of 0.16–40 Hz, which is a low high-pass filter compared to previous studies that used filters of 0.5 Hz or higher (Delorme et al., 2012; Leutheuser et al., 2013). Additionally, the application of a band-pass filter does not allow any conclusions for high-pass filtering specifically. This was addressed in detail by Winkler et al. (2015), who compared the effect of high-pass filtering in frequencies of 0 (no filter) to 40 Hz on the number of dipolar ICs and both SNR of ERPs and classification accuracy when artifactual ICs were removed. It was found that filters of <0.5 Hz were indeed suboptimal, and the best results were achieved with filters of 1–2 Hz. In a recent study, Frølich and Dowding (2018) found that filtering data that had already been band-pass filtered from 2–40 Hz with another 14 Hz high-pass filter increased the number of dipolar ICs and event-related desynchronization measures in a scenario with high muscular contributions to the data. Considering especially the impact on data with high amounts of ocular movements, Dimigen (2020) found that a high-pass filter cut-off of 1–1.5 Hz produced best results when comparing filters of 0 (no filter) to 30 Hz by assessing the residual eye artifacts, the size of the saccadic spike potential, and the distortion of artifact-free intervals. In addition, the study specifically investigated low-pass filtering and found that low-pass filtering with 40 Hz was detrimental compared to 100 Hz. Lastly, in a study using a phantom head to simulate EEG recordings during walking, Richer et al. (2019) found that adding EMG channels to the recording before computing ICA improved the recovery of simulated brain signals.

Taken together, previous studies suggest that a high-pass filter between 1 and 2 Hz and no low-pass filter seems to be the best choice to improve ICA decompositions. However, the results are inconclusive, and two factors have not yet been addressed that can be observed in several EEG studies. Firstly, no study yet compared the different requirements of standard stationary experiments and active MoBI experiments. While (Winkler et al., 2015) used a stationary experimental protocol

where comparatively low amounts of artifacts were to be expected, other studies only investigated muscle artifacts (Frølich & Dowding, 2018; Richer et al., 2019; Zakeri et al., 2014), with one study exploring the removal of ocular artifacts in great detail (Dimigen, 2020). The second not yet examined factor is the number of channels which were used for the decomposition, as none of the above-mentioned studies compared scalp channel montages, and the reported studies used channel densities ranging from 45 to 71 channels. Yet, as the number of cortical and artifactual sources in a given recording stays the same, no matter how many channels are used, the distinction between signal and noise could become more evident with an increasing number of channels, as the sources might be more clearly separated. This is especially important in mobile studies as more and stronger contributions from non-brain physiological (eye and muscle activity) and other sources (mechanical and electrical noise) are expected in these types of experiments. Here, the available degrees of freedom for the decomposition might play a crucial role.

1.3 | Current study

We thus specifically asked how movement in EEG experiments would influence the quality of the decomposition. We were further interested whether and how the number of channels would impact the decomposition results. Lastly, we investigated how the filter settings during preprocessing influence the outcome of the decomposition, especially considering the differences between stationary and mobile experiments with different spatial densities of the montages. The quality of the decomposition was assessed using the dipolarity of IC topographies, the noise in the event-related potential data after backprojection to the sensor level using only brain ICs, and the number of brain components automatically classified from all resultant ICs. We assumed that higher-density channel montages would be beneficial in general, and more so for data recorded from mobile participants. Especially for a mobile setting, we expected that adding EMG data from sensors placed on the neck would improve the decomposition. We further expected the best decomposition results for preprocessing with a high-pass filter cut-off in the range between 0.5 and 2 Hz. Finally, we hypothesized that mobile experiments include more slow drifting signals due to mechanical and movement-related artifacts and thus require a higher cut-off frequency than stationary experiments to achieve the best decomposition.

2 | METHODS

We analyzed data from a spatial orienting MoBI experiment, which is particularly useful for this study as it included a stationary as well as a mobile condition in which participants solved the identical task with comparable visual input. It allows for a comparison of stationary and mobile EEG setups and thus the impact of movement on decomposition quality. Details of the experiment can be found in (Gramann et al., 2018).

2.1 | Experiment and dataset

2.1.1 | Participants

20 healthy adults participated in the study (11 females, aged 20–46 years, $M = 30.25$ years) and were compensated with either 10/h or course credits. One participant aborted the experiment due to motion sickness, the remaining 19 datasets were used for analysis. The experiment was approved by the local ethics committee (Technische Universität Berlin, Germany) and all participants gave written informed consent in accordance with the Declaration of Helsinki.

2.1.2 | Experimental paradigm

Participants were situated in a virtual environment that displayed only floor texture. They were instructed to follow a sphere that rotated around them and stopped unpredictably on a trial at different eccentricities. The task of participants was to rotate back to indicate their initial heading direction. The task was self-paced and participants initiated a trial with a button press with their index finger. Each trial started with the appearance of a red pole indicating the starting position participants had to face. After signaling alignment with a second button press, the pole disappeared and a red sphere appeared, circling around the participant in a distance of 30m. Participants rotated on the spot to keep the sphere in the center of their view. The sphere stopped and turned blue to mark the end of the outward rotation. Participants then rotated back and indicated their estimated initial heading by a button press. Participants rotated both clockwise and counter-clockwise, in varying velocities and eccentricities (30° to 150°), in a randomized order, summing up to 140 trials. The task was completed twice, once using a traditional 2D monitor setup where movement was controlled through a joystick (stationary condition), and once with a virtual reality setup where movement was controlled through physical body movement (mobile condition). The order was balanced across participants. An overview of the paradigm can be seen in Figure 1.

In the stationary condition, participants stood in front of a TV monitor (Samsung UE42F5000AW, 1.5m distance, 40” diagonal size, HD resolution, 60 Hz refresh rate) and were instructed to move as little as possible. In the mobile

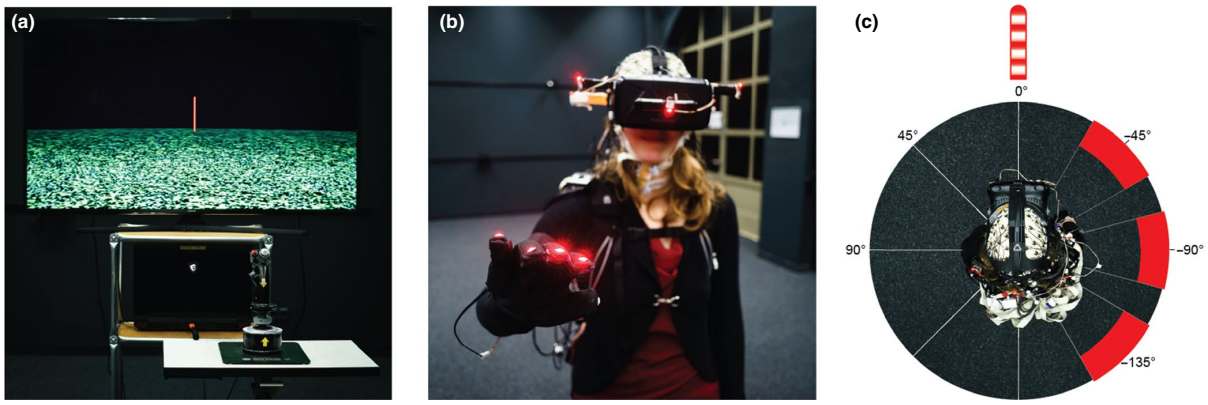


FIGURE 1 Experimental setup and paradigm. (a) Setup of the stationary condition with joystick rotation (visual flow only), displaying a sparse virtual environment with a local landmark providing the initial heading direction (pole). The joystick was placed on a table in front of the standing participant. (b) Mobile Brain/Body Imaging setup with a participant wearing a head-mounted virtual reality (VR) display, high-density EEG including an EMG neckband, and motion capture devices (red LEDs on gloves and VR). (c) Top-down view of a participant in the mobile condition, displaying the rotation eccentricities (varying ±15° around 45°, 90°, and 135°, respectively).

condition, they were wearing a head-mounted virtual reality display (HTC Vive, 110° field of view, 2x1080x1200px resolution, 90 Hz refresh rate) and a backpack PC so no cables constrained their movement, and completed the task by physically rotating on the spot. Each condition was preceded by a baseline of three minutes during which participants were asked to stand still, keep their eyes open, and to look straight ahead. Completing each condition took around 30 min, with the mobile condition being slightly shorter than the stationary condition due to faster physical rotations during the response movement.

2.1.3 | Data recording

In both conditions, EEG was recorded from 157 active electrodes on both the scalp (129 electrodes) and neck of the participant (28 electrodes). The latter were used to specifically record neck muscle activity for a potential benefit in data cleaning. Electrodes on the scalp were placed using an elastic cap with an equidistant design. The electrodes on the neck were placed with a custom design neckband (EASYCAP, Herrsching, Germany). All channels were referenced to an electrode close to the standard FCz position and data were recorded with a sampling rate of 1,000 Hz. The data were band-pass filtered from 0.016–500 Hz (BrainAmp Move System, Brain Products, Gilching, Germany) and impedances were kept below 10k Ω for electrodes on the scalp, and below 50k Ω for neck electrodes. Individual electrode locations were recorded using an optical tracking system (Polaris Vicra, NDI, Waterloo, ON, Canada).

In addition to the EEG, motion capture data were recorded using either the camera location in the virtual environment, or, in the mobile condition, the VR lighthouse tracking system (HTC Corporation, Taoyuan, Taiwan) of the head-mounted display, and active LEDs on the feet, around the hip, and on the shoulders with the Impulse X2 System (PhaseSpace Inc., San Leandro, CA, USA), all with a sampling rate of 90 Hz. Motion capture data were not used for the analyses presented here. Data and event marker streams of different sources were time-stamped and recorded using Lab Streaming Layer.[1]

2.2 | Processing pipeline

The data were analyzed in MATLAB (R2016b version 9.1; The MathWorks Inc., Natick, Massachusetts, USA), using custom scripts based on the EEGLAB toolbox (Delorme & Makeig, 2004, version 14.1.0). We investigated the effects of different factors on the quality of the resulting ICA decomposition. To this end, we systematically assessed the impact of the experimental protocol (stationary versus. mobile condition), the channel density (five different montages subsampled from the original 157-channel motage), and the high-pass filter cut-off frequency (from no filter up to 4 Hz cutoff). A schematic overview of the data processing pipeline can be seen in Figure 2.

2.2.1 | Preprocessing

Data from both conditions was first appended and individual channel locations were loaded. Raw EEG data

[1]https://github.com/sccn/labstreaminglayer

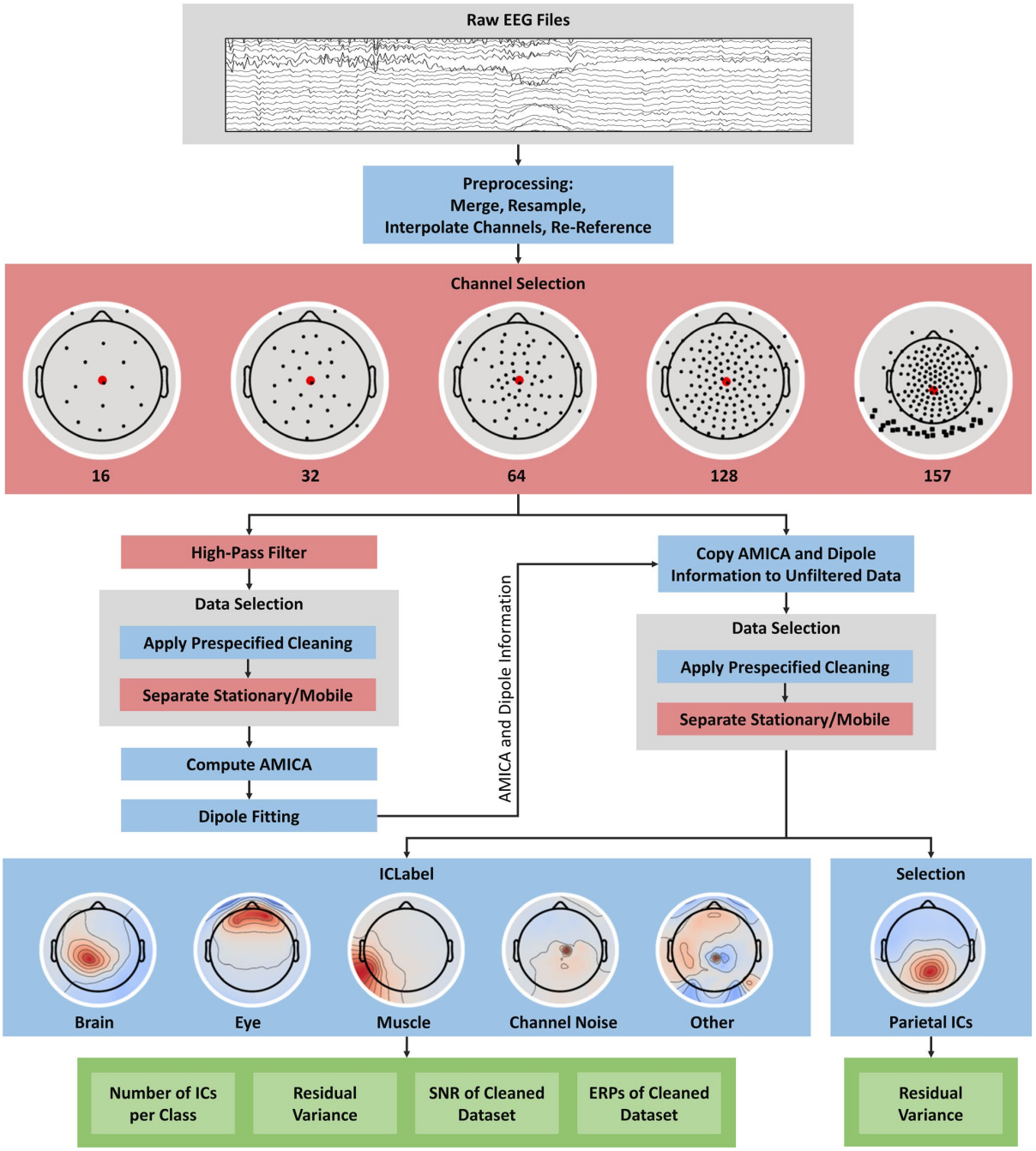


FIGURE 2 Schematic overview of the processing pipeline. Blue boxes mark processing steps which were executed identically for all datasets, red boxes mark a selection of conditions, green boxes mark final quality measures. Steps are described in section *Processing pipeline*

were then low-pass filtered to the new Nyqvist-frequency to prevent aliasing (zero-phase Kaiser-windowed sinc Finite Impulse Response (FIR) filter, Kaiser beta = 5, cutoff = 112.5 Hz, stopband = 125 Hz, transition bandwidth = 25 Hz, default when using the pop_resample function of EEGLAB) before being resampled to 250 Hz. Subsequently, bad channels were detected manually to remove strong outliers which were then interpolated (e.g. channels heavily contaminated by line noise, transient artifacts from electrode shifts, or strong drifts, 17.6 channels on average, $SD = 9.5$). Lastly, channels were re-referenced to the average reference.

2.2.2 | Channel selection

In the next step, we selected channels of the dataset to be included in the analyses. These included either all channels, using the full equidistant setup with 129 scalp and 28 neck electrodes or a subset of only the scalp electrodes, resulting in a 128, 64, 32, and a 16 channel scalp setup. The subsampled channel layouts of 64 and less channels were chosen such that the whole head was covered while the mean of the channel locations remained within 1 cm of the mean of the 128 channel layout. Since the data contained free eye movements, the two electrooculogram (EOG) electrodes below the eyes were kept for all setups. Earlier testings pointed toward different results when using a more dorsal channel layout in the 16 channel recordings, which is why an additional channel layout was tested. To not inflate the results section, the effects of this layout can be found in the supplementary material. The channel subset selection was identical for all participants, see Figure 2 for an exemplary visualization of the channel layouts. These data constituted the basic datasets ("datasets A").

2.2.3 | High-pass filtering

In order to compare the impact of different high-pass filter frequencies on ICA decompositions, the five datasets A were filtered with a zero-phase Hamming window FIR-filter (EEGLAB firfilt plugin, version 1.6.2) with varying cut-off frequencies. In many cases, it is advisable to specify the filter order in detail to achieve maximal control of the process (see Widmann et al., 2015) for a practical guide to filtering EEG data). The filter passband-edge defines where signal attenuation begins, the cut-off frequency is the frequency where the signal is attenuated by 6 db and can be regarded as the frequency where the filter starts to have a noticeable effect. The transition bandwidth is double the difference between passband-edge and cut-off frequency and is specified by the filter order. The stopband-edge is the passband-edge minus the transition bandwidth and can be regarded as the frequency where the signal attenuation reaches its full effect. At this point, it should be noted that in EEGLAB filters are specified by passband edge and follow a heuristic to find a suitable filter order (and thus transition bandwidth) depending on the frequency. For example, a default 1 Hz filter as used by EEGLAB routines has a transition bandwidth of 1 Hz and a cut-off of 0.5 Hz, whereas a 3Hz filter has a transition bandwidth of 2 Hz and a cut-off frequency of 2 Hz. For the present study, we used a constant filter order of 1,650 to ensure comparability, resulting in a transition bandwidth of 0.5 Hz independently of the passband-edge.[2] This means that a filter with a specified passband edge of 1 Hz and a transition bandwidth of 0.5 Hz leads to a cut-off frequency of 0.75 Hz and a stopband frequency of 0.5 Hz. In the further course of this paper, we use the cut-off frequency to specify the filter. As the literature suggests, we focused our analysis on lower frequencies. Since the transition bandwidth was 0.5 Hz, the lowest cut-off frequency that could be applied was 0.25 Hz. We then increased the frequency in steps of 0.25 Hz for lower frequencies up to 1.5 Hz, then in steps of 0.5 Hz up to 3 Hz, and added a 4 Hz filter as the highest frequency. Additionally, we added an analysis without any additional filtering ("0 Hz") for comparison. This resulted in 11 different filter settings for all of the datasets A.

[2]This can be reproduced in MATLAB/EEGLAB with [EEG, com, ~] = pop_eegfiltnew(EEG, highpassPassbandEdge, 0, 1,650, 0, [], 1), note that in EEGLAB the specified value is the passband edge, not the cutoff frequency, which in this case is desiredCutoff + 0.25.

2.2.4 | Data selection

After filtering, segments in the data which were not part of the experiment were rejected and subsequently a manual cleaning followed where the data were scored for strong artifacts (on average 11.1% of the experiment data was removed, $SD = 5.6\%$). The marked timepoints were saved and rejected from all filtered datasets. The separation of the stationary and mobile experimental conditions was made based on the event markers present in the data. Importantly, to ensure comparability, both the stationary and mobile conditions had to be of the same length. As a consequence, the longer dataset was cut to the length of the shorter dataset (on average, datasets were 27 min long, $SD = 5.8$ min). Overall, this resulted in 110 datasets per subject composed of 2 (movement conditions) × 5 (channel montages) × 11 (filter cutoff) that entered an ICA decomposition.

2.2.5 | Independent Component Analysis

All final 2090 datasets (110 datasets × 19 participants) were then decomposed using the AMICA algorithm (Palmer et al., 2011). AMICA was chosen since it is considered the best ICA algorithm (see section *Achieving an optimal decomposition*) and is widely used by different research groups. Although the impact of filtering has been evaluated for algorithms other than AMICA, AMICA itself was not often subject to these investigations. We used one model and ran AMICA for 2000 iterations on all datasets. Since we interpolated channels previously and used an average reference for our datasets, we also let the algorithm perform a principal component analysis rank reduction to the number of channels minus 1 (average reference) minus the number of interpolated channels. All computations were performed using four threads on machines with identical hardware, an AMD Ryzen 1,700 CPU and 32GB of DDR4 RAM. Overall, computation time amounted to 4,340 hr for all participants and datasets.

2.2.6 | Dipole fitting

Subsequently, for every resulting IC, an equivalent dipole model was computed as implemented by the DIPFIT plugin for EEGLAB. For this purpose, the individually measured electrode locations of every participant were warped (rotated and rescaled) to fit a boundary element head model based on the MNI brain (Montreal Neurological Institute, MNI, Montreal, QC, Canada). The dipole model includes an estimate of IC topography variance which is not explained by the model (residual variance, RV).

2.2.7 | Transfer of AMICA and equivalent dipole model structures

Since the final quality measures of the resultant AMICA decomposition needed to be computed on comparable unfiltered data to allow for a direct comparison of ICs, we copied the resulting weight matrices of the AMICA and the equivalent dipole model back to dataset A. This also allowed an automatic IC classification based on ICLabel (Pion-Tonachini et al., 2019) to be performed on data containing the complete spectrum which increases classification accuracy. Subsequently, the data were cleaned and separated into the two movement conditions (identical to section *Data selection*).

2.2.8 | Automatic component classification

The next part of the processing was the automatic classification of ICs using the ICLabel algorithm (Pion-Tonachini et al., 2019). ICLabel is a classifier trained on a large database of expert labelings of ICs, which classifies whether or not ICs are of brain or non-brain origin, including eye, muscle, and heart sources as well as channel and line noise artifacts and a category of other, unclear, sources. The class probability is provided allowing both for a more fine-grained analysis of probabilities and a classic popularity vote classification. Classifying based on a class probability threshold per class can be beneficial when the focus of interest lies mainly on one class, but it can also lead to ICs which have zero or more than one class labels assigned. Since we were interested in comparing the different classes, we used the popularity vote for our analysis. As a result, ICs received the class with the maximal probability as their class label. Two versions of the ICLabel algorithm exist: i) the *default* version which uses the IC activity spectrum (1-100Hz), IC topography, and IC activity autocorrelation as features for classification, and ii) the *lite* version which does not take autocorrelation into account. The latter is faster to compute and uses less RAM, especially for larger datasets, and although the classification of brain ICs can be slightly better in the default version, classification of other sources like eyes and muscles can be better using the lite version. Hence, we ran ICLabel twice using both versions but focus our analysis on the lite version.[3] See Figure 2 for example patterns of the most important classes.

2.2.9 | Automatic selection of parietal components

In addition to the ICLabel classification, we automatically selected one parietal IC for each decomposition, based on a topographic weight map. To allow automatic selection of this parietal component, we took the first 10+ number of ICs/3 ICs with a RV of <10% into account. The analysis of a specific brain IC allowed for an additional investigation of the impact of the preprocessing independent of ICLabel. Additionally, this allowed for investigating the effect of channel density, filtering, and movement on specific scalp topographies as opposed to a general decomposition quality. This can be important when using ICA to examine the data on the source level, for example in a parietal region of interest. See Figure 2 for an example parietal pattern. The low-density layouts with 16 and 32 channels were excluded from this analysis because parietal patterns could not be detected reliably. Additionally, two subjects had to be excluded even in the high-density layouts because the algorithm failed to reliably detect a parietal pattern.

2.3 | Quality measures

In order to compare the decomposition quality, we extracted several features addressing both general and practical considerations.

First, we considered the ICLabel classifications. The focus of most EEG research lies on brain signal analysis and the removal of other sources that are considered artifactual contributions. In MoBI research, in contrast, analyzing muscle and eye activity as signals can be very important to make sense of the data and potentially to be used as a source of insight into cognitive processes. Hence, we were interested in the amount of brain, eye, and muscle ICs as signal sources, and the amount of other ICs as a proxy of general decomposition quality.

Additionally, we were interested in the residual variance of ICs after fitting an equivalent dipole model. The RV, especially of brain components, is an important measure to

[3] A comparison of the two algorithms' effect on the number of ICs per class can be found in the supplementary material. For further inquiry refer to (Pion-Tonachini et al., 2019).

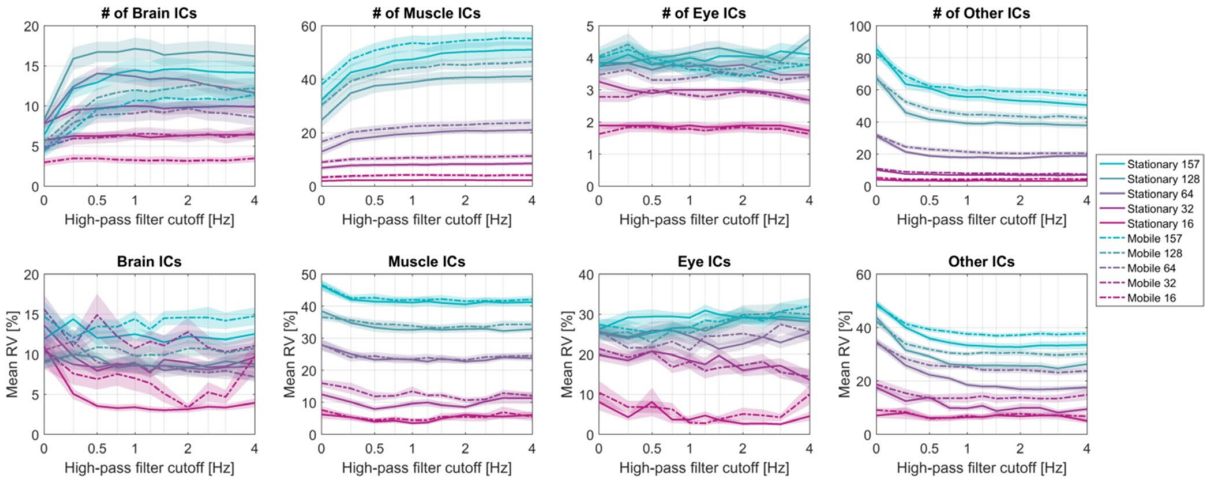


FIGURE 3 Results for the ICLabel classifications ($n = 19$). Shaded areas depict the standard error of the mean. 0 Hz refers to no additional filter being applied before computing ICA. Note the logarithmic scaling of the abscissa with grid lines for each available filter frequency. Top row: amount of ICs per class, bottom row: mean residual variance (RV) per class.

estimate the quality of a component (Delorme et al., 2012). A low RV means that the respective independent component is largely dipolar in nature, which in turn indicates more physiologically plausible sources that are more likely to be of brain origin, since the standard head models only include dipoles in the cortex. Often, this measure is used to separate brain ICs from other ICs where ICs with an RV <15% are treated as more likely originating in the brain. We were interested in the mean intra-class RV for the ICLabel classes as well as the mean RV of the parietal ICs.

Finally, as a practical measure for researchers, event-related-responses (ERPs) were computed for all datasets and further examined on their the signal-to-noise ratio (SNR). To this end, the data were pruned with ICA by removing all ICs that were classified as non-brain classes and only brain ICs were backprojected to the sensor level. In case no IC was classified as brain, the one with the highest brain probability was used (over all conditions, this occurred 16 times). Since the data were previously scored for strong artifacts in the time domain, only trials not containing these artifacts entered the ERP. The ERPs were computed at an electrode in equidistant layout which was positioned closest to the POz electrode in the standard layout (POz'). Importantly, to not distort the results we used no frequency filter on the data (as the ICA results were copied back to the unfiltered data), but only the spatial filter of the ICA. We then extracted epochs (−600 ms to +1,200 ms) around the trial-onset event (onset of the moving sphere) for which we expected a parietal late positive complex to occur, and removed the pre-stimulus baseline activity. The two mobility conditions (stationary, mobile) did not contain the same amount of events, as the stationary condition had to be cut short to fit in length to the mobile condition in which participants rotated back faster and thus were able to answer more trials within the same time. To ensure comparability between the conditions, we determined the minimal number of available events for both movement conditions per subject and used this number of events in both conditions to compute the ERPs. On average, 77.8 ($SD = 18.2$) epochs were used per subject and condition, and the final measures for signal and noise were computed. To this end, the mean amplitudes from 250 ms to 450 ms served as the signal which was divided by the standard deviation in the 500 ms pre-stimulus interval to compute the SNR (Debener et al., 2012).

3 | RESULTS

As the effects are either clearly visible in the figures or a reflection of the arbitrarily chosen filter steps, statistical testing was not performed. Overall, the ICA decompositions were sensitive to the different preprocessing parameters. Figure 3 shows the results for the number of ICs in the *Brain*, *Muscle*, *Eye*, and *Other*, classes, as well as their mean RV. The results of the RV values of the parietal ICs can be seen in Figure 4. Finally, the practical quality measures of ERPs and SNR values can be found in Figure 5.

Clear differences could be observed between the stationary and mobile data. The stationary data contained more brain ICs and less muscle ICs than the mobile data (see Figure 3), and additionally, the mobile data contained more ICs classified as "other". Interestingly, the number of eye ICs did not differ between the mobility conditions. A larger RV of brain ICs could be observed in the mobile condition, however, this difference was not very large. Considering the parietal ICs specifically (see Figure 4), the mobile condition consistently

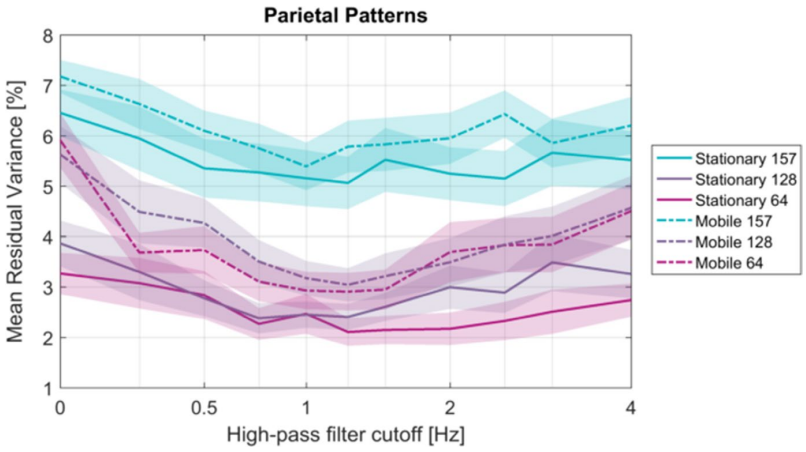


FIGURE 4 Residual variance (RV) of the parietal patterns ($n = 17$). Shaded areas depict the standard error of the mean. Only channel montages of 64 and more channels were considered. 0 Hz refers to no additional filter being applied before computing ICA. Note the logarithmic scaling of the abscissa with grid lines for each available filter frequency.

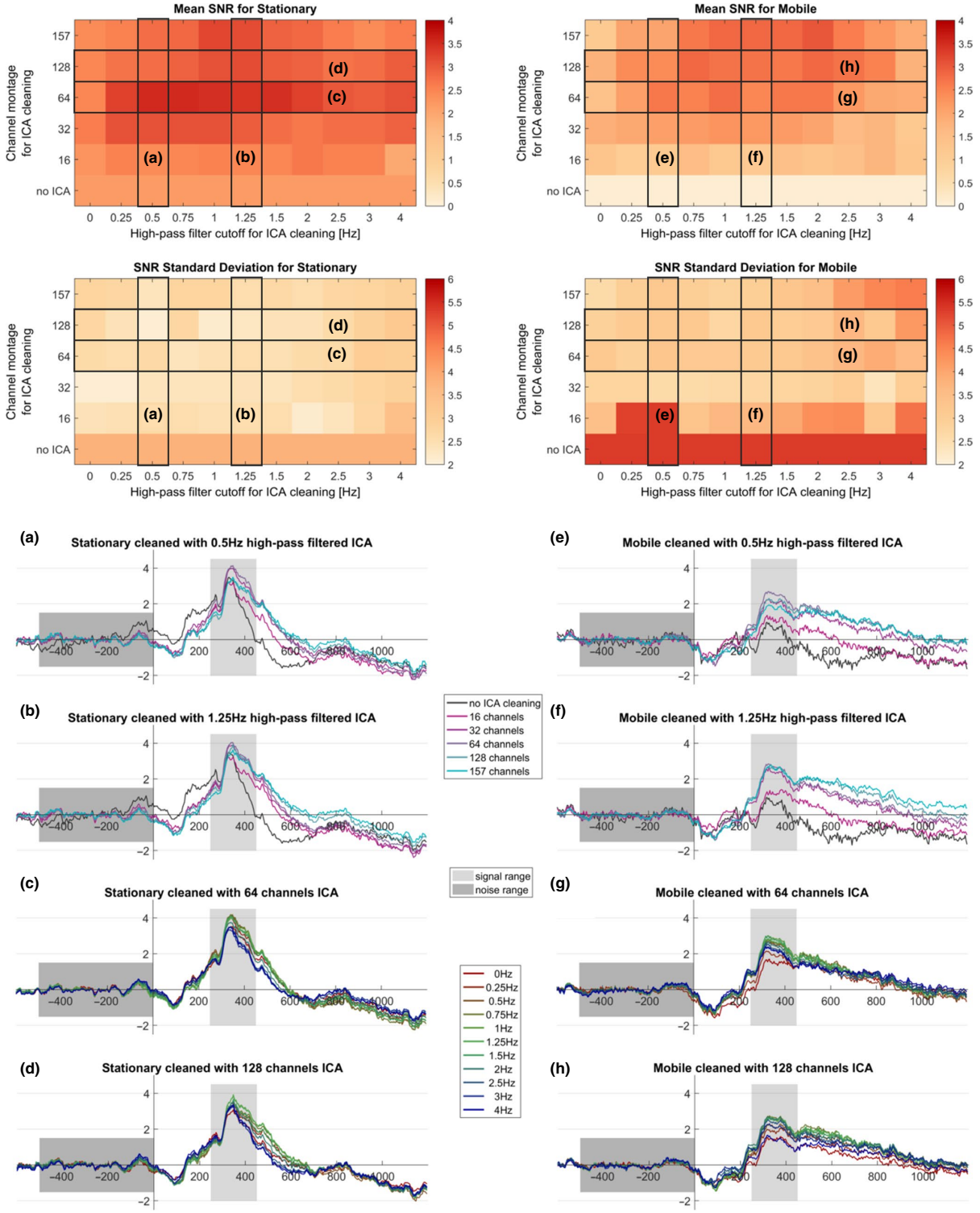
exhibited a slightly higher RV than the counterparts in the stationary condition. The SNR of the ERPs was considerably greater in the stationary condition than in the mobile condition (see Figure 5), and the shape of the ERP in the two mobility conditions was different, with a larger late positive peak including a steeper offset in the stationary condition.

The ICA decomposition was also clearly influenced by the channel montage, with generally more ICs being present in each class in higher channel densities. This was evident also in the number of brain ICs, but even with 16 channels there were still ICs classified as brain. Importantly, the difference in brain ICs between montages with different channel densities appeared less pronounced than the difference in muscle and other ICs, indicating a possibly more pronounced stability of the brain ICs. Note that the maximum of brain ICs was not reached with the montage containing the neck band (157 channels), but with 128 scalp channels. This was true for both mobility conditions, but the detrimental effect of the neck band was less pronounced in the mobile condition. The number of eye ICs was stable across channel montages, with densities of 64 channels and upward containing four eye ICs, the 32 channel montage containing three eye ICs, and only the 16 channel montage containing only two ICs classified as reflecting eye movement activity. The RV of channel montages with fewer channels was lower in general. The 16 channel montages reached RV values of <10% in most cases, including not only physiological, but also other ICs. The difference between the channel montages were more pronounced for muscle and other ICs than for brain ICs, which means that the difference in RV values between brain and non-brain ICs was larger when using more channels. In the parietal ICs, the 157 channel montage showed a general increase of RV (around 2 percentage points) above the 128 and 64 channel montages, whereas only a slight increase could be observed in the 128 channel montage over the 64 channel montage. Interestingly, when looking at the SNR and the ERP waveforms, the 64 channel montage already led to very good or the best results and cleaning the data with ICA based on 32 channels already led to a substantial improvement in SNR. The improved SNR values did not come at the cost of increased standard deviations which would indicate more outliers. On the contrary, the standard deviation of the SNRs was generally lower when the data were cleaned with ICA. Visual inspection of the ERPs showed improvements already for 16 channels when cleaned with ICA, but the ERP waveform, especially in the mobile condition, was more similar to the uncleaned ERP than to the ERP of the 32 channels condition. However, employing a channel layout more focused on dorsal electrodes led to an improvement of the SNR and ERP waveforms (see supplementary material).

The high-pass filter applied before computing ICA had a considerable influence as well. For brain and muscle ICs an increase in the number of ICs in these classes with increasing high-pass frequencies could be observed, especially when compared to the data without additional filter ("0 Hz"). The number of eye ICs appeared to be insensitive

FIGURE 5 Practical quality measures ($n = 19$). Top: Signal-to-noise ratios (SNRs) of the event-related-responses (ERPs) computed on uncleaned data and data that were cleaned with ICA by removing all non-brain ICs as classified by ICLabel. ERPs were computed on the POz' electrode for the trial-onset event. Note that as the ICA results were copied back to the unfiltered datasets prior to ERP computation, the datasets themselves only differed in the ICA decompositions that were used for cleaning, no frequency filter was applied before computing the ERPs. SNR was defined as the mean amplitude in the 250 ms–450 ms interval divided by the standard deviation in the 500 ms pre-stimulus interval. 0 Hz refers to no additional filter being applied before computing ICA. Bottom: Corresponding ERPs, plotted either for different channel montages and a fixed filter cutoff frequency used before computing ICA (columns of SNR plots, a, b, e, f), or different cutoff frequencies and a fixed channel montage (rows of SNR plots, c, d, g, h).

to high-pass filtering, whereas the number of other ICs dropped with increasing filter frequency. The effect of filtering also exhibited a ceiling (brain/muscle ICs) or floor (other ICs) effect, where a filter higher than 1–2 Hz did not affect the results any further. The number of muscle ICs increased a little slower, approaching an optimum from 2Hz onward and continued to increase even up to 4 Hz cut-off (maximum filter applied). The RV of brain ICs appeared



relatively stable across different filter frequencies, implying that an increased number of brain ICs did not coincide with classifying less dipolar ICs as originating in the brain. The mean RV of brain ICs ranged from 3% (stationary condition with 16 channels) to 15% (mobile condition with 157 channels). RV values of muscle and other ICs droppped with increasing filter frequency up to 1 Hz, which was more noticeable with higher channel densities. In the parietal ICs, a slightly different pattern was observed, where the RV did not approach a floor asymptote, but increased again after reaching the minimum around 1 Hz. This inverted U-shape with increasing filter frequencies could also be observed for the SNR measures of the ERPs and the ERP waveforms themselves, where mid-range filters showed a larger late positive signal, not only in the range used for the SNR computation (250 ms to 450 ms post-stimulus) but continuing until around 600 ms post-stimulus. In the parietal ICs, the already small difference between the 64 and the 128 channel montage disappeared at the optimal filter frequency. The SNR in higher channel densities generally required a higher filter to reach its maximum which could also be seen in the ERPs themselves.

Finally, the effects of filter and mobility conditions appeared to interact as well. For the number of brain ICs, the necessary filter to reach the maximum showed a marked difference between stationary and mobile data: The maximum number of 17.2 brain ICs in the stationary condition was reached with 128 channels and a 1 Hz filter. In contrast, the maximum of 12.7 brain ICs in the mobile condition was reached also with 128 channels but requiring a filter of 2.25 Hz. The SNR of the ERPs also showed that in the mobile condition a higher filter led to the best results. As expected from the SNR values, the ERP waveforms with the highest late positive values in the mobile condition were the ones cleaned with ICA computed on higher filtered data, especially noticeable with 128 channels where the maximal late positive peak occurred with the 2.25 Hz filter in the mobile condition as opposed to 1.25 Hz in the stationary condition.

4 | DISCUSSION

EEG is a widely adopted tool in neuroscientific research and in the recent years new trends toward more active and mobile experiments emerged, allowing for the investigation of more natural cognitive processes "in the wild". These experiments, however, come with the drawback of additional and stronger artifactual contributions in the data which can mask electrical activity originating from the brain. Separating the different sources is thus a key step in modern EEG research and does not only allow for an analysis of clean data but also an estimation of the source activity and their cortical origins. Although blind source separation techniques like ICA are widely adopted as a tool to achieve this goal, the influence of different factors on this decomposition is not always clear. In this study, we investigated the impact of movement of participants, channel density, and high-pass filter cut-off frequency during preprocessing on the decomposition of EEG data with ICA. We evaluated the outcome of ICA based on differently preprocessed data using the number of brain, muscle, eye, and other ICs as classified by ICLabel, their dipolarity, and the SNR of ERPs on the cleaned data.

The results show that, as expected, participant movement has a detrimental effect on the decomposition and generally leads to fewer and less dipolar brain ICs but more muscle ICs in the data. This is not surprising as more artifacts and especially more muscle activity is present in MoBI data which take up degrees of freedom for the ICA decomposition. Importantly, ICA is still a powerful tool for cleaning EEG data even in light of increasingly noisy recordings from MoBI and mobile EEG experiments, as the effect of cleaning on SNR and ERPs shows. In fact, in mobile EEG studies researchers risk to obtain a very low SNR of an ERP without cleaning the data, but ERP quality might be significantly improved by removing non-brain ICs. When inspecting the ERPs of the uncleaned data some residual signal was observed even though the average varied around the baseline activity, possibly indicating an underestimation of the SNR in this case. This notwithstanding, the ERP waveform is clearly different from the cleaned ones. Analyzing MoBI data with ICA or a comparably powerful cleaning method thus appears to be vital. It is to note that the strong difference of the ERP waveform between stationary and mobile data is not necessarily an effect of artifacts and increased noise alone. Since the brain needs to fulfill a variety of additional tasks when moving the body by preparing and executing motor commands with constant sensory feedback, attention to a specific stimulus may be limited (Ladouce et al., 2019) and decreasing sensory mismatch might change oscillatory processes (Gramann et al., 2018).

Considering the effect of the number of EEG channels, it can be stated that a higher scalp electrode density generally leads to a better ICA decomposition. However, there seems to be a ceiling effect when cleaning the sensor data with ICA which is reached already when using 64 channels, as shown in the ERP SNR analysis of both mobility conditions. Although the ICA decomposition of the 16 channel subsampled montage did not reach the same level of data cleaning as observed for the high-density montages, it seems to be powerful enough to reconstruct event-related activity in stationary experiments to a useful degree and is still an improvement over uncleaned data analysis for mobile protocols. Recent advancements using ear-EEG have already shown promising results in detecting EEG artifacts with ICA in low-density recordings (Bleichner & Debener, 2019), and channel layouts focusing on more dorsal electrode sites appear to result in better cleaning capabilities (see supplementary material).

Selecting a suitable channel layout might thus be especially important in low density recordings. Additionally, since the ICLabel classifier was not trained on low-density recordings, it might be possible to improve the ERP reconstruction by selecting specific ICs manually. Nonetheless, using more channels resulted in more brain ICs, which in turn led to a more precise source-level analysis, making EEG a powerful tool to truly image the brain in action. A second surprise was the observed detrimental effect of an EMG neckband on the number of brain ICs and their RV. This might have two reasons: First, it could be that the neckband is not an ideal candidate for measuring EMG activity. As 28 electrodes were placed around the neck, the width of the band might have been too large in some participants, leading to movement of the neck band and the incorporated electrodes and thus artifacts due to changes in the electrode-skin contact. Additionally, since the neckband was fixed, turning the head might have led to electrodes shifting over the skin, leading to artifacts and EMG measurements that were potentially spatially unstable, introducing non-stationarity into the ICA decomposition and thus violating assumptions of the ICA model. In sum, the EMG neckband that was used might have introduced more artifacts to the data than adding useful information and degrees of freedom for the spatial filter. Another explanation could be that the images of the IC topography used by ICLabel as a feature for classification did not incorporate topographies based on additional neck channels (see Figure 2). This might have led to less accurate classifications and thus more incorrectly classified brain ICs. When looking at the SNRs and ERPs, although not being particularly helpful, the neckband seemed to be unproblematic, and especially in the mobile case the highest SNR was reached with the 157 channel montage.

In light of these considerations and the beneficial effect of EMG found in simulation studies (Richer et al., 2019), it needs to be further examined whether using sticky electrodes for recording neck muscle activity instead of a neckband improves the results. One other observation we made is that the RV of ICs decreases with fewer channels, seemingly suggesting a better, more dipolar, decomposition. This effect could be caused by an actually better decomposition due to more samples available relative to the number of channels (as the dataset size was kept identical to ensure that the same information entered the ICA). On the other hand, it might be caused by less measurement points (channels) available to compute the RV in these recordings. Extrapolating to the case of a single-channel recording, no RV would be measurable any more. Exploring this factor by adjusting the dataset length to the number of channels is an important option for future investigations. Independently of the underlying cause, however, it is important to note that in experiments of typical lengths of 30 to 60 min, RV may only be useful to dissociate brain and non-brain ICs when recording with higher-density montages of 64 channels and more.

Lastly, we were able to confirm our hypothesis that high-pass filtering before computing ICA does improve the decomposition when the data are filtered with a cutoff between 0.5–2 Hz. However, there is not one optimal filter as the filtering frequency should be adjusted depending on other factors of the experiment. In standard stationary experiments with 64 channels a high-pass filter cutoff of 0.5 Hz is acceptable, but with increasing number of channels a higher filter cut-off of up to 1.25 Hz should be employed to achieve the best decomposition. This effect was even more pronounced in the mobile condition where the decomposition improved further with cut-off frequencies of up to 2 Hz, corresponding to results of Winkler et al. (2015) and Dimigen (2020). Interestingly, even though we came to similar conclusions as Winkler et al. (2015) regarding the filter cut-off, we did observe clear changes in the ERP waveforms which the authors did not report. We believe this could be due to differences in the experimental paradigm, with the present study requiring participants to stand upright even in the stationary condition controlling the visual flow with a joystick. This likely introduced more artifacts than would be observed in a classic auditory oddball paradigm with seated participants and the effect of cleaning the data with ICA thus became more noticeable. Comparing our results to those of Frølich and Dowding (2018), we could not confirm that a very high cut-off frequency led to better results, as we saw detrimental effects after reaching the optimal filter cut-off. These conflicting results may be due to the fact, that Frølich and Dowding (2018) employed a 45 Hz low-pass filter even though they specifically investigated muscle activity, which is more prevalent in higher frequencies.

It should be noted that cleaning data with classic ICA alone is not the only option to remove undesired artifacts. Dimigen (2020) was able to improve the ICA cleaning capabilities by leveraging eye tracking data to overweight saccadic potentials before computing ICA. Artifact Subspace Reconstruction (ASR; Kothe & Jung, 2015) is another cleaning method that gained increased attention in the last years, especially since it also works in an online fashion. Recently, Chang et al. (2020) evaluated ASR in terms of its efficacy when using different cleaning sensitivities and also showed that an ICA decomposition could be improved by first cleaning the data with ASR. ASR is particularly helpful in removing transient burst artifacts, but when setting the sensitivity to a degree which removes physiological artifacts from eyes and muscles reliably from the data, it bears the risk of removing too much brain activity as well. Using a cautious ASR cleaning in combination with the classic ICA appears to be a promising approach and needs further evaluation. Additional modifications like Riemannian ASR (Blum et al., 2019) could also be of interest here. Another potentially promising online-capable unified source imaging and artifact cleaning

approach was proposed by Ojeda et al. (2019), but further comparisons and evaluations are needed. Taken together, the field of EEG research is clearly moving toward more sophisticated artifact removal techniques which advance our abilities to investigate the human brain in everyday life. Extending the present study to include and compare these recent data cleaning methods is a promising step for future investigations.

We conclude that obtaining an optimal ICA decomposition when analyzing EEG data is highly relevant, not only for source-level analysis but also for cleaning sensor data, and it is especially effective and necessary when expecting increased artifactual contributions to the recording. We would like to finalize this paper by providing some recommendations as a set of "best practices" when performing ICA on EEG data.

First of all, when computing ICA to remove eye and muscle artifacts it is important to do this on data which was high-pass filtered but not low-pass filtered, and it is unproblematic to apply the obtained decomposition to unfiltered data for further analysis. Second, higher-density recordings of 64 and more channels should be used when aiming for an optimal recovery of the brain signals and especially when doing source-level analysis, as low-density recordings cannot separate neural sources adequately. Third, an increasing channel density is required with increasing movement range and velocity in the experimental protocol. Fourth, when no high-density recording is possible, ICA can still be used to clean the sensor data from eye and muscle activity artifacts. Last, but not least, we recommend using higher high-pass filter cut-offs than traditionally used. We want to emphasize again that when discussing filters in this paper we used the cut-off frequency, not the passband-edge as the defining parameter, and when using EEGLAB it is recommended to specify the correct filter (see section *High-pass filtering*). While 0.5 Hz might be acceptable for 64 channels in stationary experiments, using a 1 Hz filter is not detrimental and ensures a good decomposition also for higher-density recordings with more noise being present in the data. For MoBI experiments with significant noise even higher filters of 1.5 or even 2 Hz should be employed before computing ICA, depending on the channel montage.

ACKNOWLEDGMENTS

This work was supported by the DFG (GR2627/8-1) and USAF (ONR 10024807). We thank Jonna Jürs and Yiru Chen who assisted in collecting the data, and Emma Auerbach Brode for her initial contributions to the project. We also sincerely thank Olaf Dimigen for important notes on filter specifications. Open access funding enabled and organized by ProjektDEAL.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

M.K. and K.G. designed the research; M.K. participated in the original data collection; M.K. performed the data analysis and wrote the first draft of the paper; K.G. and M.K. edited the paper.

DATA AVAILABILITY STATEMENT

Data relating to these experiments are available for download at http://dx.doi.org/10.14279/depositonce-10493. Source code for running the pipeline and plotting the figures is available for download at https://github.com/MariusKlug/KeyFactorsForImprovingICAinEEG and the v1.0 release can be cited as https://doi.org/10.5281/zenodo.4003882.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.14992.

ORCID

Marius Klug https://orcid.org/0000-0001-8667-3457
Klaus Gramann https://orcid.org/0000-0003-2673-1832

REFERENCES

Artoni, F., Menicucci, D., Delorme, A., Makeig, S., & Micera, S. (2014). RELICA: A method for estimating the reliability of independent components. *NeuroImage*, *103*, 391–400. https://doi.org/10.1016/j.neuroimage.2014.09.010

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*, 1129–1159.

Bleichner, M. G., & Debener, S. (2019). Independent component decomposition of around ear EEG data to detect artifacts. *Conf. Proc. - IEEE Int. Conf. Syst. Man Cybern*, 3631–3634.

Blum, S., Jacobsen, N. S. J., Bleichner, M. G., & Debener, S. (2019). A riemannian modification of artifact subspace reconstruction for EEG artifact handling. *Frontiers in Human Neuroscience*, *13*, https://doi.org/10.3389/fnhum.2019.00141

Chang, C. Y., Hsu, S. H., Pion-Tonachini, L., & Jung, T. P. (2020). Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel EEG recordings. *IEEE Transactions on Biomedical Engineering*, *67*, 1114–1121.

De Sanctis, P., Butler, J. S., Malcolm, B. R., & Foxe, J. J. (2014). Recalibration of inhibitory control systems during walking-related dual-task interference: A Mobile Brain-Body Imaging (MOBI) Study. *NeuroImage*, *94*, 55–64. https://doi.org/10.1016/j.neuroimage.2014.03.016

Debener, S., Minow, F., Emkes, R., Gandras, K., & de Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology*, *49*, 1617–1621. https://doi.org/10.1111/j.1469-8986.2012.01471.x

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*, 9–21.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS One*, *7*, e30135. https://doi.org/10.1371/journal.pone.0030135

Dimigen, O. (2020). Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. *NeuroImage*, *207*, 116117. https://doi.org/10.1016/j.neuroimage.2019.116117

Djebbara, Z., Fich, L. B., Petrini, L., & Gramann, K. (2019). Sensorymotor brain dynamics reflect architectural affordances. *Proceedings of the National Academy of Sciences*, 1–31.

Ehinger, B. V., Fischer, P., Gert, A. L., Kaufhold, L., Weber, F., Pipa, G., & König, P. (2014). Kinesthetic and vestibular information modulate alpha activity during spatial navigation: A mobile EEG study. *Frontiers in Human Neuroscience*, *8*, 71. https://doi.org/10.3389/fnhum.2014.00071

Frølich, L., & Dowding, I. (2018). Removal of muscular artifacts in EEG signals: A comparison of linear decomposition methods. *Brain Informatics*, *5*, 13–22. https://doi.org/10.1007/s40708-017-0074-6

Gehrke, L., Iversen, J. R., Makeig, S., & Gramann, K. (2018). The invisible maze task (IMT): Interactive exploration of sparse virtual environments to investigate action-driven formation of spatial representations. In S. Creem-Regehr, J. Schöning, & A. Klippel (Eds.), *Spatial Cognition XI* (pp. 293–310). Springer International Publishing.

Gramann, K., Ferris, D. P., Gwin, J., & Makeig, S. (2014). Imaging natural cognition in action. *International Journal of Psychophysiology*, *91*, 22–29.

Gramann, K., Gwin, J. T., Bigdely-Shamlo, N., Ferris, D. P., & Makeig, S. (2010). Visual evoked responses during standing and walking. *Frontiers in Human Neuroscience*, *4*, 202.

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T. P., Lin, C. T., …, Makeig, S. (2011). Cognition in action: Imaging brain/body dynamics in mobile humans. *Reviews in the Neurosciences*, *22*, 593–608.

Gramann, K., Hohlefeld, F. U., Gehrke, L., & Klug, M. (2018). Heading Computation in the Human Retrosplenial Complex during full-body Rotation, *bioRxiv*, 417972.

Groppe, D. M., Makeig, S., & Kutas, M. (2009). Identifying reliable independent components via split-half comparisons. *NeuroImage*, *45*, 1199–1211. https://doi.org/10.1016/j.neuroimage.2008.12.038

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent Component Analysis*. John Wiley & Sons.

Hyvärinen, A., & Oja, E. (2013). Independent Component Analysis: Algorithms and Applications. *Neural Networks*, *56*, 963–976.

Jungnickel, E., Gehrke, L., Klug, M., & Gramann, K. (2018). MoBI-Mobile Brain/Body Imaging. In H. Ayaz, & F. Dehais (Eds.), *Neuroergonomics: The Brain at Work and in Everyday Life*, 1st ed. (pp. 59–63). Elsevier.

Kothe, C. A. E., & Jung, T.-P. (2015). Artifact removal techniques with signal reconstruction. U.S. Patent, 047462 A9.

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., & Ietswaart, M. (2017). Understanding Minds in Real-World Environments: Toward a Mobile Cognition Approach. *Frontiers in Human Neuroscience*, *10*, 1–14.

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., & Ietswaart, M. (2019). Mobile EEG identifies the re-allocation of attention during real-world activity. *Scientific Reports*, *9*, 1–10.

Leutheuser, H., Gabsteiger, F., Hebenstreit, F., Reis, P., Lochmann, M., & Eskofier, B. (2013). Comparison of the AMICA and the InfoMax algorithm for the reduction of electromyogenic artifacts in EEG data. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., EMBS*, 6804–6807.

Makeig, S., Gramann, K., Jung, T. P., Sejnowski, T. J., & Poizner, H. (2009). Linking brain, mind and behavior. *International Journal of Psychophysiology*, *73*, 95–100.

Makeig, S., Jung, T. P., Bell, A. J., Ghahremani, D., & Sejnowski, T. J. (1997). Blind separation of auditory event-related brain responses into independent components. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 10979–10984.

Nenna, F., Do, C.-T., Protzak, J., & Gramann, K. (2020). Alteration of brain dynamics during natural dual-task walking. *European Journal of Neuroscience*, 1–27.

Ojeda, A., Klug, M., Kreutz-delgado, K., Gramann, K., Mishra, J. (2019). A Bayesian framework for unifying data cleaning, source separation and imaging of electroencephalographic signals. *bioRxiv*.

Palmer, J. A., Kreutz-delgado, K., & Makeig, S. (2011). AMICA: An adaptive mixture of independent component analyzers with shared components, 1–15.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026

Protzak, J., & Gramann, K. (2018). Investigating established EEG parameter during real-world driving. *Frontiers in Psychology*, *9*, 1–11.

Richer, N., Downey, R. J., Nordin, A. D., Hairston, W. D., & Ferris, D. P. (2019). Adding neck muscle activity to a head phantom device to validate mobile EEG muscle and motion artifact removal. *2019 9th Int. IEEE/EMBS Conf. Neural Eng.*, 275–278.

Wascher, E., Heppner, H., & Hoffmann, S. (2014). Towards the measurement of event-related EEG activity in real-life working environments. *International Journal of Psychophysiology*, *91*, 3–9.

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods*, *250*, 34–46.

Winkler, I., Debener, S., Muller, K. R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, 4101–4105.

Wunderlich, A., & Gramann, K. (2018). Electrocortical evidence for long-term incidental spatial learning through modified navigation instructions. In S. Creem-Regehr, J. Schöning, & A. Klippel (Eds.), *Spatial Cognition XI* (pp. 261–278). Springer International Publishing.

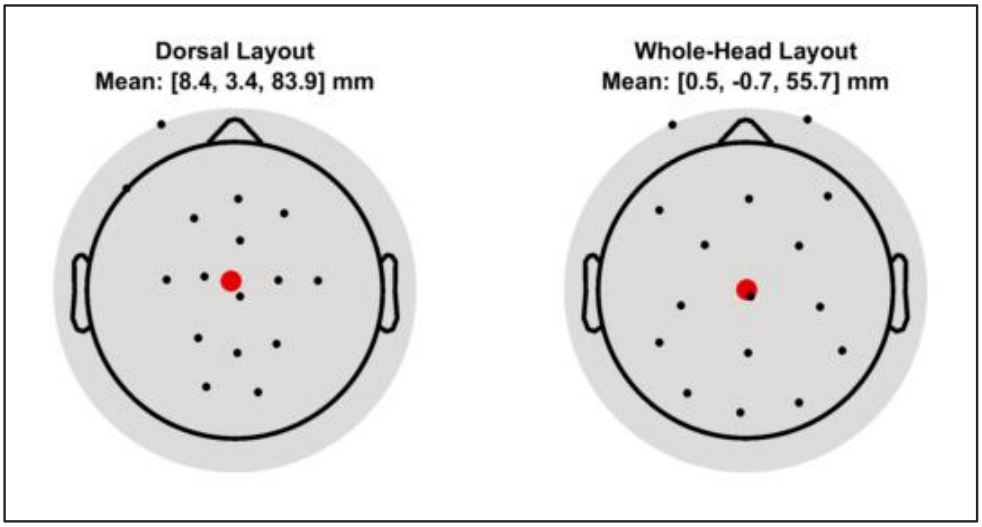
Zakeri, Z., Assecondi, S., Bagshaw, A. P., & Arvanitis, T. N. (2014). Influence of Signal Preprocessing on ICA-Based EEG Decomposition. *IFMBE Proceedings*, *41*, 563–566.

SUPPORTING INFORMATION

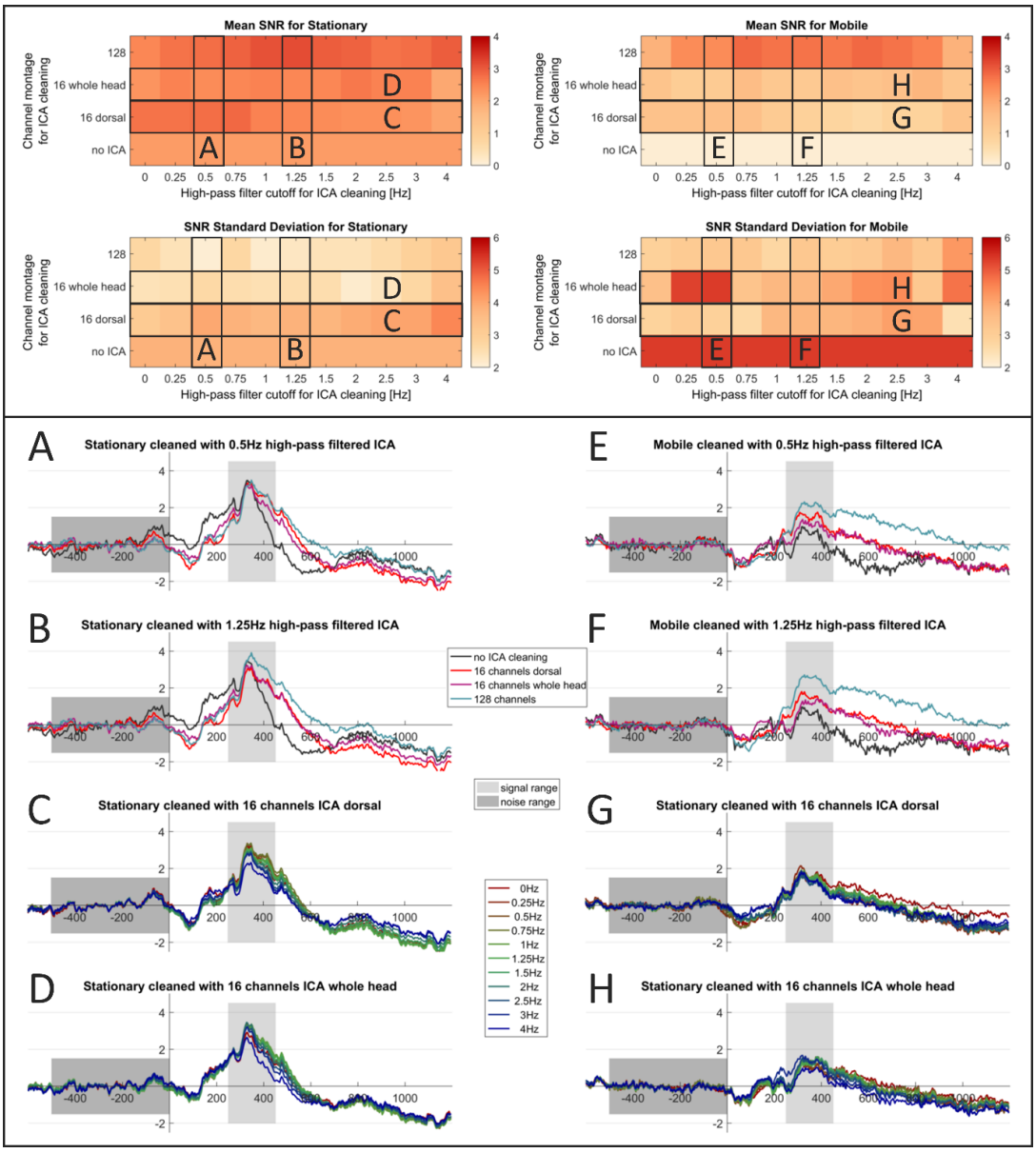
Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Klug M, Gramann K. Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *Eur J Neurosci*. 2021;54:8406–8420. https://doi.org/10.1111/ejn.14992

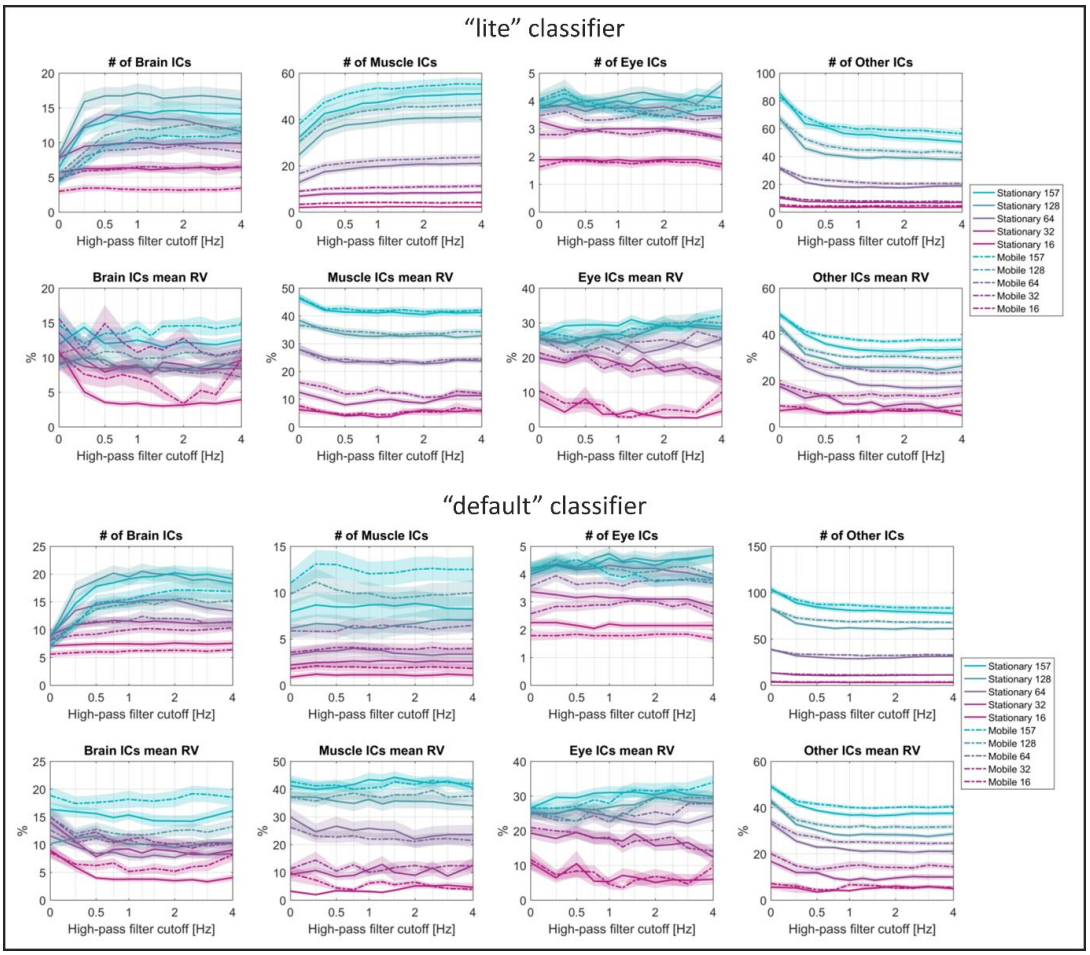
Supplementary material



1: Comparison of the 16 channel montages. Left: Focusing on dorsal electrodes, right: whole-head subsampling of the original 128 scalp channels which kept the mean stable (<1cm change).



2: Comparison of SNR and ERP results for 16 channel montages on the whole head and dorsal focus (n=19). Top: Signal-to-noise ratios (SNRs) of the event-related-responses (ERPs) computed on uncleaned data and data that was cleaned with ICA by removing all non-brain ICs as classified by ICLabel. ERPs were computed on the POz' electrode for the trial-onset event. Note that as the ICA results were copied back to the unfiltered datasets prior to ERP computation, the datasets themselves only differed in the ICA decompositions that were used for cleaning, no frequency filter was applied before computing the ERPs. SNR was defined as the mean amplitude in the 250 ms - 450 ms interval divided by the standard deviation in the 500 ms pre-stimulus interval. 0 Hz refers to no additional filter being applied before computing ICA. Bottom: Corresponding ERPs, plotted either for different channel montages and a fixed filter cutoff frequency used before computing ICA (columns of SNR plots, A, B, E, F), or different cutoff frequencies and a fixed channel montage (rows of SNR plots, C, D, G, H).



3: Results for the ICLabel classifications using the lite and the default classifier (n=19). Shaded areas depict the standard error of the mean. 0 Hz refers to no additional filter being applied before computing ICA. Note the logarithmic scaling of the abscissa with grid lines for each available. Also note the difference in classifications of muscle ICs.

No need for extensive artifact rejection for ICA - A multi-study evaluation on stationary and mobile EEG datasets

*Klug, M., *Berg, T., & Gramann, K. (submitted). No need for extensive artifact rejection for ICA - A multi-study evaluation on stationary and mobile EEG datasets.

** authors contributed equally*

No need for extensive artifact rejection for ICA - A multi-study evaluation on stationary and mobile EEG datasets

*Klug, M.[1], *Berg, T.[1], & Gramann, K.[1]

1) Biopsychology and Neuroergonomics, Institute of Psychology and Ergonomics, TU Berlin, Berlin, Germany

* *authors contributed equally*

Keywords: *electroencephalography, mobile brain/body imaging, signal processing, artifact removal, preprocessing, independent component analysis*

Abstract

Objective. Electroencephalography (EEG) studies increasingly make use of more ecologically valid experimental protocols involving mobile participants that actively engage with their environment leading to increased artifacts in the recorded data (MoBI; Gramann et al., 2011). When analyzing EEG data, especially in the mobile context, removing samples regarded as artifactual is a common approach before computing independent component analysis (ICA). Automatic tools for this exist, such as the automatic sample rejection of the AMICA algorithm (Palmer et al., 2011), but the impact of both movement intensity and the automatic sample rejection has not been systematically evaluated yet.
Approach. We computed AMICA decompositions on eight datasets from six open-access studies with varying degrees of movement intensities using increasingly conservative sample rejection criteria. We evaluated the subsequent decomposition quality in terms of the component mutual information, the amount of brain, muscle, and “other” components, the residual variance of the brain components, and an exemplary signal-to-noise ratio.
Main results. We found that increasing movements of participants led to decreasing decomposition quality for individual datasets but not as a general trend across all movement intensities. The cleaning strength had less impact on decomposition results than anticipated, and moderate cleaning of the data resulted in the best decompositions.
Significance. Our results indicate that the AMICA algorithm is very robust even with limited data cleaning. Moderate amounts of cleaning such as 5 to 10 iterations of the AMICA sample rejection with 3 standard deviations as the threshold will likely improve the decomposition of most datasets, irrespective of the movement intensity.

1 Introduction

Removing artifacts from electrophysiological data in the time-domain can be a task as time-consuming as it is important. Rejecting periods of “bad” data that should not be taken into account for further downstream analysis has become a staple in electroencephalography (EEG) analysis from the outset of the method. This includes the

rejection of bad epochs when computing event-related measures, but also the rejection of bad samples before running an independent component analysis (ICA; Bell & Sejnowski, 1995; Hyvärinen et al., 2001). ICA decomposes the acquired sensor data into components that can subsequently be interpreted regarding the underlying physiological processes (e.g. brain, eyes, muscle, other), and as a common preprocessing step before running ICA, bad samples are removed from the data. Similar to applying a high-pass filter before ICA and copying the decomposition results back to unfiltered data (Klug & Gramann, 2021; Winkler et al., 2015), the ICA results computed on a dataset that had bad time points removed can be applied to the complete uncleaned data in the end (e.g. Gramann et al., 2021; Jacobsen et al., 2021) to retain as much data as possible for downstream analyses. Although time-domain cleaning is regularly used, to our knowledge no study has yet investigated the effect of time-domain cleaning on ICA decomposition in depth. The present study addresses this issue and systematically investigates the effect of time-domain cleaning on the resultant ICA decomposition while taking different experimental protocols into account that increase in mobility from stationary to Mobile Brain/Body Imaging (MoBI; Gramann et al., 2011, 2014; Makeig et al., 2009) setups.

1.1 Cleaning of data from mobile experiments is a complex problem

While data cleaning in the time-domain is important for stationary, seated, experiments, it is even more relevant for experiments collecting data from mobile participants. These mobile EEG (Debener et al., 2012) and MoBI (Jungnickel et al., 2019) studies are gaining popularity. Especially with analytical options to remove non-brain activity from high-density EEG data, this approach allows for imaging the human brain in its natural habitat – in participants moving in and actively engaging with their environment. This increased mobility naturally comes with increased non-brain activity, traditionally considered artifacts, contributing to the recorded signal. On the one hand, more physiological activity stemming from the eyes and muscles will be present, on the other hand, also electrical and mechanical artifacts stemming from additional devices, cable sway, or electrode shifts on the scalp will be more prevalent in mobile EEG data (Gramann et al., 2011; Gwin et al., 2010; Jungnickel & Gramann, 2016). In traditional seated experimental protocols, all these contributions to the recorded signal would be candidates for removal. However, as eye and muscle activity can be found throughout the recordings in stationary as well as mobile EEG data, and they typically can be removed with ICA, it is not always clear or easy to decide which time points to remove during time-domain cleaning. Considering mechanical artifacts, large transient spikes from electrode shifts can be detected comparably easily but cable sway, for example, although potentially high in amplitude, is not a clear case for removal since it might be present in the entire experiment and/or especially in times that are interesting in the experimental paradigm. Taken together, mobile EEG protocols complicate the time-domain cleaning of electrophysiological experiments and traditional heuristics for data cleaning can not always be applied.

1.2 Different automatic cleaning options - different challenges

Removing samples from data manually is a sub-optimal approach, as it is both time-consuming and subjective. With the varying experience of the persons cleaning the data, the resulting cleaning strategy will also vary. Even within the same person, different mental states or varying noise levels in different datasets may alter the cleaning procedure.

To ensure a reliable, repeatable, and transparent cleaning, it is thus preferred to make use of automatic cleaning algorithms. Several such options exist, ranging from methods based on simple amplitude criteria to the identification of artifactual time periods based on the spectral characteristics to more complex approaches that identify artifactual time periods based on artifact subspaces such as the EEGLAB (Delorme & Makeig, 2004) *clean_rawdata* function, which uses Artifact Subspace Reconstruction (ASR; Kothe & Jung, 2015).

These methods do have their challenges, though. Identifying bad time points by their amplitude alone will either be a very lax measure or it will also remove all periods containing eye blinks since these are high-amplitude signals. Removing eye-blinks before ICA decomposition, however, is not desired, as ICA can typically remove these more reliably and preserve the respective time points for downstream analyses. Spectral measures will be prone to removing periods with muscle activity since these are usually detected by having increased high-frequency broad-band power (Onton & Makeig, 2006). But especially in mobile experiments, removing time periods with muscle activity would result in excessive cleaning, and the computed ICA decomposition would not be readily applicable to the entire dataset since it was not informed by time points containing muscle activity. And while the cleaning threshold of ASR can be adjusted to remove mainly large transient spikes, ASR is very sensitive to this threshold (Chang et al., 2020), and especially for mobile data, it does not always find a suitable baseline by itself. ASR thus requires a specifically recorded baseline and sometimes different cleaning thresholds for different movement modalities and even different datasets within the same modality, which renders it unsuitable for automatic data cleaning as targeted in this study.

1.3 AMICA sample rejection

The Adaptive Mixture ICA (AMICA; Palmer et al., 2011), currently one of the most powerful ICA algorithms (Delorme et al., 2012), includes an inbuilt function to reject bad samples that might not be well-utilized by researchers working with the algorithm: AMICA can reject bad samples based on their log-likelihood while computing the decomposition model. The log-likelihood is an objective criterion corresponding to the algorithm's estimate of the model fit, effectively leading to the rejection of samples AMICA cannot easily account for. Hence, unlike other cleaning methods, this option will only remove those kinds of artifacts that negatively affect the decomposition and retain those that can be decomposed and removed with ICA. This is done in an iterative fashion: First, several steps of the model estimation are performed, then samples are rejected based on the difference of their log-likelihood from the mean in standard deviations (SDs), then the model is estimated for several steps again before the next rejection, and so on. The start of the rejection, the number of rejection iterations, the SD threshold as well as the number of model computation steps between each rejection can be set in the AMICA computation parameters. This artifact rejection approach is model-driven and allows users to automatically remove time-domain data to improve the decomposition.

When applied from the EEGLAB user interface (AMICA plugin v1.6.1), this is disabled by default, but when opening the rejection sub-interface, it is enabled with 5 iterations with 3 SDs, starting after the first AMICA step, with one step between each iteration. When the *runamica15* function is applied directly from the command line, the cleaning is disabled by default, but when enabled, it rejects three times with 3SDs, starting after the second AMICA

step, with 3 steps between each iteration. As a consequence, different settings will impact whether and how AMICA uses time-domain cleaning during the decomposition. However, while previous evaluations have shown that AMICA is one of the currently best algorithms for EEG decomposition (Delorme et al., 2012), the impact of the integrated time-domain cleaning procedure has not been evaluated yet.

1.4 Current study

In this study, we thus investigate the impact of automatic sample rejection on the quality of the AMICA decomposition of data from six experiments with different levels of mobility. To this end, we varied the cleaning intensity in terms of the number of cleaning iterations as well as the rejection threshold. As measures of decomposition quality, we used the number of brain, muscle, eye, and unspecified components, the residual variance of brain components, and the component mutual information. Additionally, we examined the signal-to-noise ratio in one standing and one mobile condition of the same experiment. We hypothesized that 1) increasing mobility affects the decomposition negatively, 2) cleaning affects the decomposition positively, and 3) an interaction exists, where experiments with more movement require more cleaning to improve their decomposition quality. We had no hypothesis as to what the optimal amount of cleaning should be. Based on the empirical evidence from this study, we formulate recommendations for using sample rejection with AMICA.

2 Methods

2.1 Datasets

For a reliable estimate of the effect of time-domain cleaning on the quality of the ICA decomposition for datasets from stationary as well as mobile EEG protocols, we included open access EEG datasets with a wide range of movement conditions in this study. This resulted in eight datasets from six studies containing standard seated protocols but also a gait protocol, arm reaching, and irregular movement protocols. We categorized the datasets into the four groups of low, low-to-medium, medium-to-high, and high movement intensity as laid out in the individual dataset descriptions. We used datasets that used at least 60 EEG channels (not including EOG), had a sampling rate of at least 250 Hz, and either contained channel locations or standard 10-20 system electrode layouts. Representation of different EEG setups was ensured by including no more than two datasets from one lab. We manually subsampled channels where necessary as described in the section *Preparation*. The following datasets were used:

Video Game:
This dataset is available at https://doi.org/10.18112/openneuro.ds003517.v1.1.0 (Cavanagh & Castellanos, 2021) and contains data of 17 participants (6 female and 11 male, mean age = 20.94 years, SD = 5.02 years). Participants were sitting while playing a video game using a gamepad. Data was recorded with a 500 Hz sampling rate using 64 electrodes (Brain Products GmbH, Gilching, Germany), and filtered with a high-pass filter of .01 Hz and a low-pass filter of 100 Hz. Channels were manually downsampled by selecting 58 channels of only scalp electrodes. This dataset was categorized as having low movement intensity.

Face Processing:
This dataset is available at https://doi.org/10.18112/openneuro.ds002718.v1.0.5 (Wakeman & Henson, 2021) and contains data of 19 participants (8 female and 11 male, age range 23-37 years). Participants were seated in front of a screen and exposed to images of faces. Data was recorded with a 1100 Hz sampling rate using 70 electrodes and a 350 Hz low-pass filter was applied. Channels were manually downsampled by selecting 65 channels that were closest to corresponding electrodes from a 10-20 layout of scalp electrodes. This dataset was categorized as having low movement intensity.

Spot Rotation (stationary/mobile):
This dataset is available at https://doi.org/10.14279/depositonce-10493 (Gramann et al., 2021) and contains data of 19 participants (10 female and 9 male, aged 20–46 years, mean age = 30.3 years). The experiment consisted of a rotation on the spot, which either happened in a virtual reality environment with physical rotation or in the same environment on a two-dimensional monitor using a joystick to rotate the view. Participants were standing in front of the computer screen in the stationary condition. The data was split into the two conditions of joystick rotation (stationary) and physical rotation (mobile) for the purpose of this study. Data was recorded with a 1000 Hz sampling rate using 157 electrodes (129 on the scalp in a custom equidistant layout, 28 around the neck in a custom neckband) with the BrainAmp Move System (Brain Products GmbH, Gilching, Germany). Channels were manually downsampled by selecting 60 channels that were closest to corresponding electrodes from a 10-20 layout of scalp electrodes. This dataset was categorized as having low-to-medium (stationary) and medium-to-high (mobile) movement intensity.

Beamwalking (stationary/mobile):
This dataset is available at https://doi.org/10.18112/openneuro.ds003739.v1.0.2 (Peterson & Ferris, 2021) and contains data of 29 participants (15 female and 14 male, mean age = 22.5 years, SD = 4.8 years). Participants either stood or walked on a balance beam and were exposed to sensorimotor perturbations. Perturbations were either virtual-reality-induced visual field rotations or side-to-side waist pulls. Because of the different degrees of movement, the data was split according to the two conditions: stationary (standing) and mobile (walking). Data was recorded at a 512 Hz sampling rate using 136 electrodes (BioSemi Active II, BioSemi, Amsterdam, The Netherlands). Channels were manually downsampled by selecting 61 channels that were closest to corresponding electrodes from a 10-20 layout of scalp electrodes. This dataset was categorized as having low-to-medium (stationary) and high (mobile) movement intensity.

Prediction Error:
This dataset is available at https://doi.org/10.18112/openneuro.ds003846.v1.0.1 (Gehrke et al., 2021) and contains data of 20 participants (12 female, mean age = 26.7 years, SD = 3.6 years) of which one was removed by the authors due to data recording error. Participants were seated at a table and equipped with an HMD. The task consisted in reaching for virtual cubes that appeared in front of participants on the table. Participants moved their arm and upper torso to reach the virtual goal on the table. Data was recorded with a 1000 Hz sampling rate using 64 electrodes (Brain Products GmbH, Gilching, Germany). Channels were manually downsampled by selecting 58 channels of only scalp electrodes. This dataset was categorized as having medium-to-high movement intensity.

Auditory Gait:
This dataset is available at https://doi.org/10.1038/s41597-019-0223-2 (Wagner et al., 2019) and contains data of 20 participants (9 females and 11 males, aged 22–35 years, mean age = 29.1 years, SD = 2.7 years). Participants had to walk on a treadmill and synchronize their steps to a regular auditory pacing stream that included infrequent, sudden shifts in tempo. Data was recorded with a 512 Hz sampling rate using 108 electrodes with seven 16-channel amplifiers (g.tec GmbH, Graz, Austria), high pass filtered >0.1 Hz, low pass filtered <256 Hz and a notch filter was applied at 50 Hz to remove power line noise. Channels were manually downsampled by selecting 61 channels that were closest to corresponding electrodes from a 10-20 layout of scalp electrodes. This dataset was categorized as having high movement intensity.

2.2 Data processing

All data was processed in an automated fashion with identical preprocessing steps as displayed in figure 1. The main processing steps can be summarized under pre-processing, AMICA with sample rejection, and ICA post-processing, followed by the computation of quality measures to evaluate the decomposition.

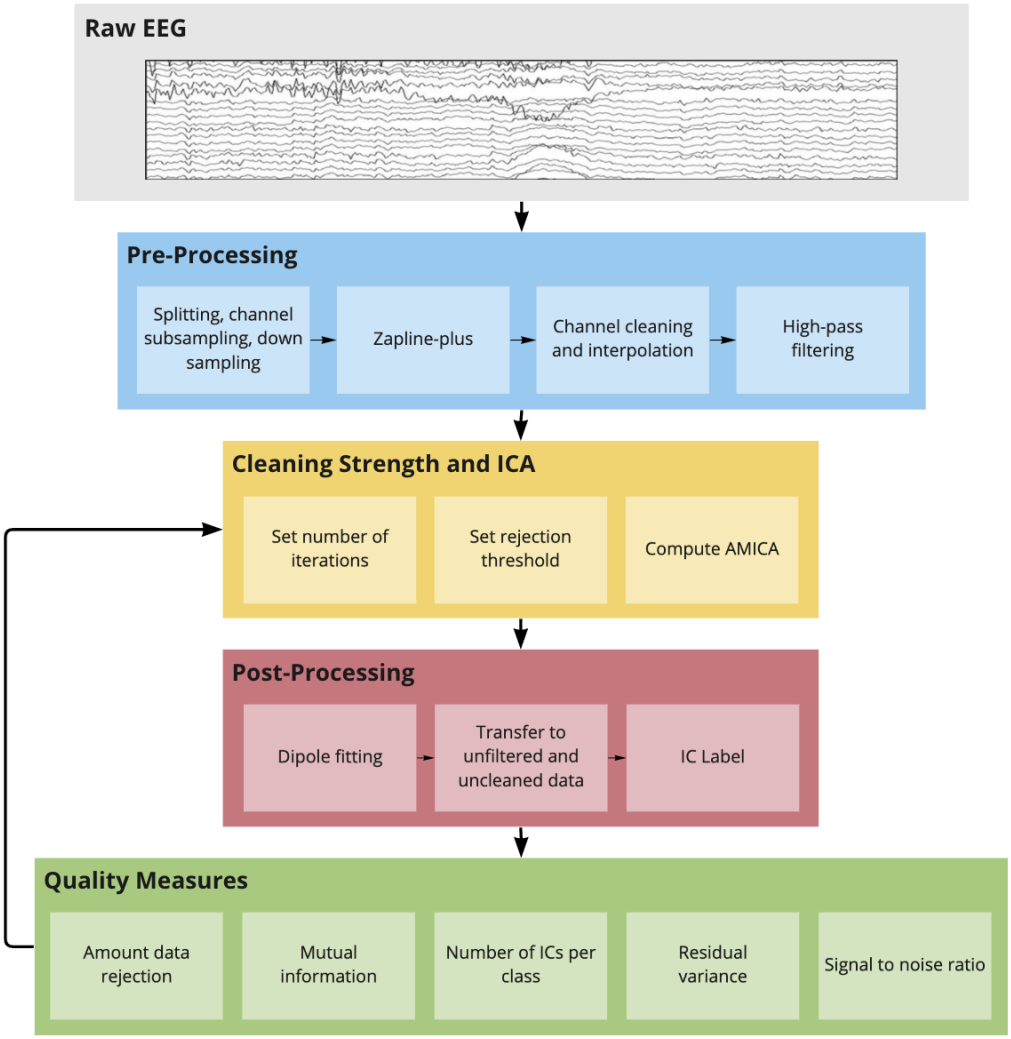


Figure 1: Data processing pipeline. The number of cleaning iterations and the rejection threshold were varied several times to compute AMICA and post-processing as well as quality measures repeatedly. Details can be found in section *Data processing*.

2.2.1 Preparation

All datasets were first loaded into EEGLAB, and if a study contained data of two different conditions these were split and subsequently treated separately. We then manually selected channels to reduce the number of channels to a range of 58 to 65, excluding EOG or neck channel locations, and matched the remaining channels to the 10-20 layout as close as possible if the original layout was equidistant (see section *Datasets*). We did not subsample all studies to exactly 58 channels to allow an evenly distributed whole head coverage in all datasets. Full comparability of the channel layout between studies was not given, nor was it intended since the data were recorded in different labs with different devices. In addition, a previous investigation revealed a ceiling effect in obtaining brain ICs with an increasing number of electrodes used for ICA decomposition (Klug & Gramann, 2021). A doubling of channels from 64 to 128 resulted in only 3 to 4 more brain ICs while differences in the number of brain ICs due to different movement intensities were more substantial with around

5 to 6 brain ICs. As a consequence, we expected our differences in channel count to have minimal impact on the computed quality measures. After channel reduction, all datasets were downsampled to 250 Hz and reduced to a length of 10 minutes (150,000 samples), ensuring there was the same amount of data available for all datasets. All data were subsequently processed using the *BeMoBIL pipeline* (Klug et al., 2022) with identical parameters except for varying time-domain cleaning types and strengths.

2.2.2 Zapline-plus

The *Auditory Gait* dataset had a notch filter applied before uploading. All other datasets were processed with Zapline-plus (de Cheveigné, 2020; Klug & Kloosterman, 2022) to remove line noise and other frequency-specific noise peaks. Zapline (de Cheveigné, 2020) removes noise by splitting the data into an originally clean (data A) and a noisy part (data B) by filtering the data once with a notch filter (A) and once with the inverse of this notch filter (B). It then uses a spatial filter to remove noise components in the noisy part (B) to get a cleaned version of that part (B'). Finally, the two clean parts (A and B') are added back together to result in a cleaned dataset with full rank and full spectrum except for the noise. Zapline-plus is a wrapper for Zapline that chunks the data to improve the cleaning and adapts the cleaning intensity automatically, thus maximizing the cleaning performance while ensuring minimal negative impact. We used default parameters in all datasets containing mobile data (*Spot Rotation, Prediction Error, Beamwalking*), but limited the noise frequency detector to line noise for the *Face Processing* and *Video Game* datasets to avoid the removal of minor spectral peaks in stationary datasets that were not expected to be influenced by additional electronic equipment or mechanical artifacts.

2.2.3 Channel cleaning and interpolation

We detected bad channels using the *clean_rawdata* plugin of EEGLAB in an iterative way. We did not use the *flatline_crit*, the *line_noise_crit*, and samples were not rejected, nor was ASR applied. The only used criteria to detect bad channels was the *chancorr_crit* criterion, which interpolates a channel based on a random sample consensus (RANSAC) of all channels and then computes the correlation of the channel with its own interpolation in windows of 5 s duration. If this correlation is below the threshold more than a specified fraction of the time (50% in our case), it is determined to be bad. Since the function has a random component, it does not necessarily result in a stable rejection choice, which is why this detection was repeated ten times, and only channels that were flagged as bad more than 50% of the time were finally rejected. Removed channels were then interpolated and the data was subsequently re-referenced to the average using the *full rank average reference* plugin of EEGLAB, which preserves the data rank while re-referencing.

2.2.4 High-pass filtering

High-pass filtering has a positive effect on the ICA decomposition quality and is especially important in mobile studies (Klug & Gramann, 2021). For a dataset with 64 channels, a filter of 0.5 to 1.5 Hz cutoff resulted in the best decomposition for both stationary and mobile conditions (Klug & Gramann, 2021), which is why we chose a cutoff of 1 Hz in this study. We specified the filter manually as recommended (Widmann et al., 2015) and used the same filter specifications as in Klug & Gramann (2021): a zero-phase Hamming window FIR-filter (EEGLAB firfilt plugin, v1.6.2) with an order of 1650 and a passband-edge of 1.25 Hz, resulting in a transition bandwidth of 0.5 Hz and a cutoff frequency of 1 Hz.

2.2.5 Independent component analysis with sample rejection

All final datasets were decomposed using AMICA with different numbers of sample rejection iterations and different rejection thresholds to compare the results of the decomposition for the eight datasets. We used one model and ran AMICA for 2000 iterations. Since we interpolated channels previously we also let the algorithm perform a principal component analysis rank reduction to the number of channels minus the number of interpolated channels. As we used the full rank average reference, we did not subtract an additional rank for this. All computations were performed using four threads on machines with identical hardware, an AMD Ryzen 1700 CPU with 32GB of DDR4 RAM.

In this step, we investigated the effect of different cleaning intensities using the AMICA sample rejection algorithm. All rejection was started after the *runamica15* default 2 iterations, with the default 3 iterations between rejections. We repeated the AMICA computation process either without sample rejection, or with 1, 3, 5, 7, or 10 iterations using 3 SDs as the threshold, and additionally with 10, 15, and 20 iterations using 2.8 SDs, and last with 20 iterations using 2.6 SDs.

2.2.6 Dipole fitting

An equivalent dipole model was computed for each resulting independent component (IC) using the DIPFIT plugin for EEGLAB with the 3-layer boundary element model of the MNI brain (Montreal Neurological Institute, Montreal, QC, Canada). The dipole model includes an estimate of IC topography variance which is not explained by the model (residual variance, RV). For datasets that had individually measured electrode locations of an equidistant channel layout, the locations were warped (rotated and rescaled) to fit the head model.

2.2.7 Transfer of ICA to unfiltered data

One of the goals of this study was to investigate the effect of time-domain cleaning on the component mutual information (MI) and the signal-to-noise ratio (SNR) when applied to the full dataset. To this end, the resulting AMICA decomposition and dipole models were copied back to the preprocessed dataset from section *Channel cleaning and interpolation* (line noise removed, channels interpolated, re-referenced to the average, but no high-pass filter and no time-domain cleaning).

2.2.8 Independent component classification using ICLabel

In order to categorize the ICs according to their likely functional origin, we applied the ICLabel algorithm (Pion-Tonachini et al., 2019). ICLabel is a classifier trained on a large database of expert labelings of ICs that classifies ICs into brain, eye, muscle, and heart sources as well as channel and line noise artifacts and a category of other, unclear, sources. As it was shown that the ‘lite’ classifier worked better than the ‘default’ one for muscle ICs (Klug & Gramann, 2021), we used the ‘lite’ classifier in this study. We used the majority vote to determine the final class, meaning the IC received the label with the highest probability.

2.2.9 Quality Measures

To measure the impact of the cleaning on the ICA decomposition, we used several measures that addressed both mathematical and practical considerations of the ICA decomposition: i) The mutual information (MI) of the components after applying the ICA solution to the complete dataset. The MI is essentially the mathematical description of how

well the ICA can decompose the data, as the ICA minimizes component MI. ii) The number of ICs categorized as stemming from *brain*, *muscle*, and *other* sources defined by ICLabel, as especially in MoBI research not only brain, but all physiological sources can be of interest to the experimental analysis. iii) The mean RV of brain ICs, as this can be considered a measure of the physiological plausibility of the IC (Delorme et al., 2012). iv) The combination of the above measures as the ratio of the number of brain ICs by the mean brain RV (higher indicates better decomposition), v) An exemplary computation of the signal-to-noise ratio (SNR) on the two *Spot Rotation* datasets (physical rotation in VR and 2D monitor rotation), using the same measures as in (Klug & Gramann, 2021). For this, we removed all non-brain ICs in the final dataset and computed event-related potentials (ERPs) of the trial onsets at the electrode closest to the Pz electrode in the 10-20 system. On average, 30.58 (SD = 7.31) epochs were used per subject and condition. This comparably low number of epochs is caused by the previous reduction in dataset length to 10 minutes, and the results of this approach should be interpreted with care. The signal was defined as the mean amplitude from 250 ms to 450 ms and the standard deviation in the 500 ms pre-stimulus interval was used as a measure of the noise (Debener et al., 2012).

3 Results

As the effects were either clearly absent or a reflection of the arbitrarily chosen steps in cleaning intensity, we did not perform a statistical analysis. We anticipated three effects regarding the time-domain cleaning and movement intensity on the decomposition quality of the ICA. First, we expected a higher movement intensity to decrease the decomposition quality. Second, higher cleaning intensity should remove more artifacts and therefore result in a better decomposition quality. Finally, we expected an interaction of movement intensity and cleaning intensity such that more movement would require more cleaning to reach a better decomposition.

3.1 The effect of movement intensity on the decomposition quality

The most fundamental effect of changing the hyperparameters of AMICA sample rejection is the amount of data that is rejected. If more movement results in more problematic artifacts (those that AMICA can not easily include in its model e.g. due to nonlinearities), more movement should also result in more rejected samples. As can be seen in figure 2a, the dataset with the lowest amount of rejection was indeed a set with low movement intensity, namely the *Face Processing* dataset. Yet, the dataset with the second lowest removal was a high movement intensity set (*Auditory Gait*). Furthermore, the most data was rejected in the *Spot Rotation (stationary)* set, which was a set with medium-to-low movement intensity. Overall, there was no discernible trend as to whether movement intensity affected the amount of data that was rejected by the AMICA.

Figure 2b shows the results for the MI, where the AMICA reached the lowest MI, and thereby its best decomposition from a mathematical point of view, in the *Auditory Gait* and *Spot Rotation (mobile)* datasets. The *Auditory Gait* and *Spot Rotation (mobile)* datasets were categorized as high and medium-to-high movement intensity, respectively. Even when comparing movement conditions of the same studies, no clear trend was identifiable. While the results for the *Beamwalking* study yielded a higher MI in its mobile condition, the *Spot*

Rotation study showed the opposite trend.

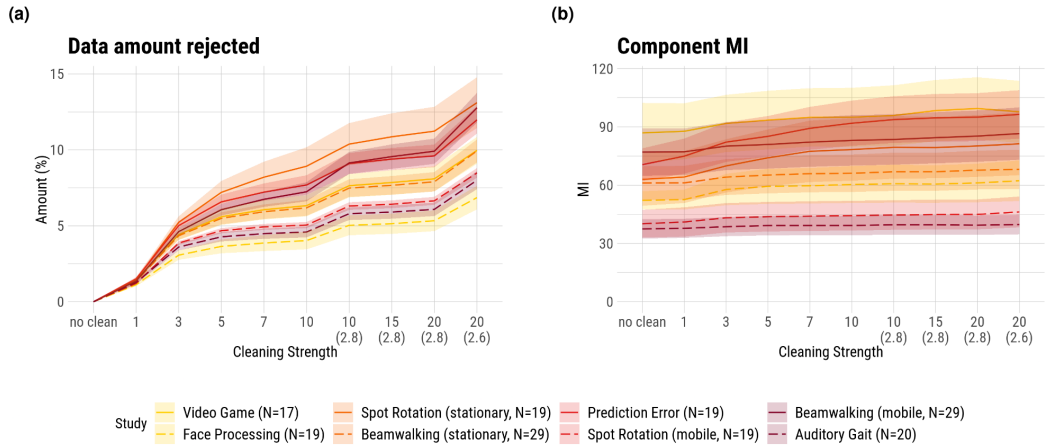


Figure 2: Results for the data rejection amount (a) and component mutual information (MI; b). Shaded areas depict the standard error of the mean (SE). The numbers on the abscissa refer to the number of iterations of the sample rejection, with a default of 3 SDs as the rejection threshold. The numbers in brackets on the abscissa refer to the rejection threshold in SDs when deviating from the default. “No clean” refers to no sample rejection being applied when computing AMICA. The colors denote the movement intensities: yellow - low, orange - low-to-medium, red - medium-to-high, violet - high.

The results of the number of resulting brain and muscle ICs can be seen in figure 3a. The highest amounts of brain ICs were found in the datasets *Face Processing*, *Beamwalking (stationary)*, and *Auditory Gait*, with the last dataset having high movement intensity. The group of datasets with the lowest number of brain ICs (with around 6 fewer than the highest) consisted of both medium-to-high movement intensities but also included the low movement intensity *Video Game* study. The number of muscle ICs also did not show a clear trend with movement intensity: one of the datasets with the least amount of muscle ICs was the high movement intensity set *Beamwalking (mobile)*. When looking at the within-study differences, the stationary conditions in both the *Spot Rotation* and the *Beamwalking* study showed around 2-3 brain ICs more than their respective mobile conditions. This was not the case for the number of muscle ICs, however, since both conditions of the *Beamwalking* study revealed almost identical amounts of muscle ICs and the *Spot Rotation* study even exhibited around 1-2 more muscle ICs in its stationary condition as compared to its mobile counterpart.

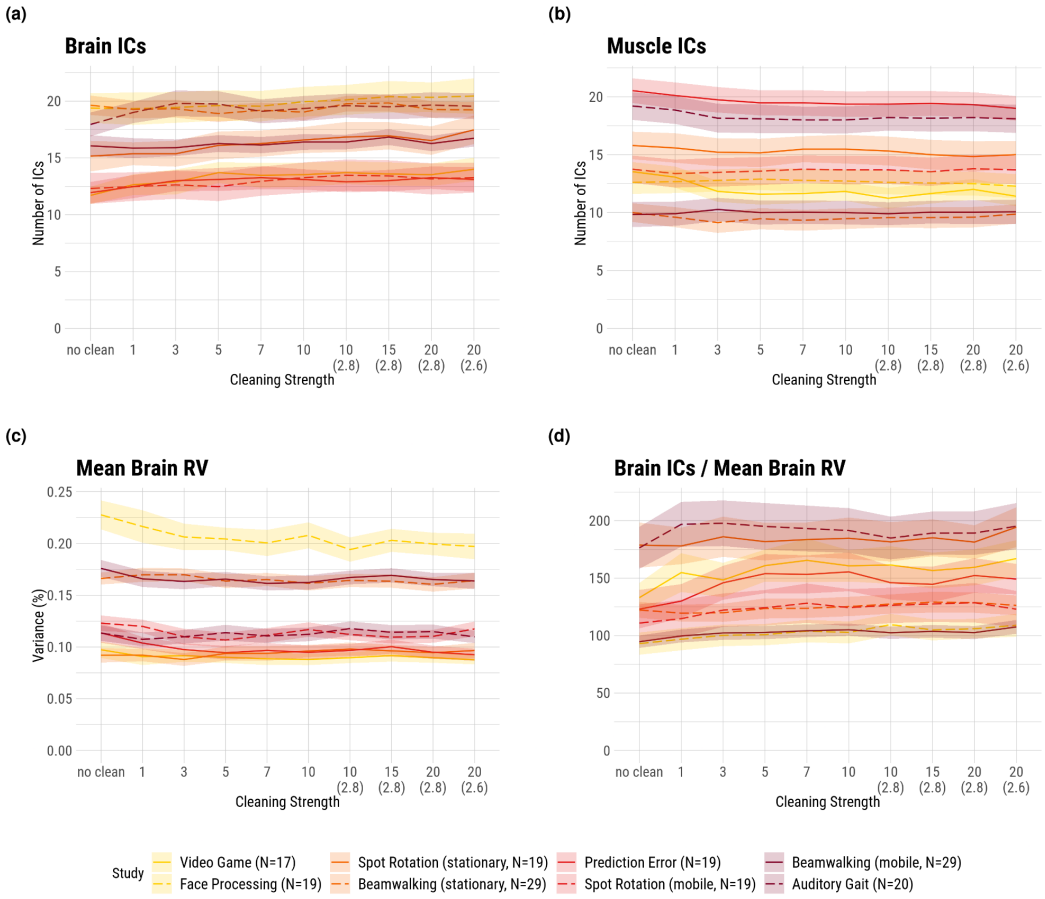


Figure 3: Results for the brain (a) and muscle (b) ICLabel classifications, residual variance (RV; c) and the ratio of the number of brain ICs to their mean RV (d). Shaded areas depict the standard error of the mean (SE). The numbers on the abscissa refer to the number of iterations of the sample rejection, with a default of 3 SDs as the rejection threshold. The numbers in brackets on the abscissa refer to the rejection threshold in SDs when deviating from the default. “No clean” refers to no sample rejection being applied when computing AMICA. The colors denote the movement intensities: yellow - low, orange - low-to-medium, red - medium-to-high, violet - high.

As can be seen in figure 3c, the highest RV values (indicating lowest physiological plausibility) for brain ICs (above 20%) were found in the *Face Processing* set, which was a low movement intensity study. The three datasets of *Prediction Error*, *Spot Rotation (stationary)*, and *Video Gaming*, coming from three different movement intensity groups, formed the cluster with the lowest RVs of around 10%. Taken together, the movement intensity of a study had no clear effect on the quality of the brain ICs. This is further supported by the fact that the best and worst ratio of the number of brain ICs to their mean RV was attained by two studies of the high movement intensity group. However, an effect within studies was present in the *Spot Rotation* study. Its stationary condition yielded RVs that were around 2-3% lower than its mobile condition. The *Beamwalking* study showed almost identical RVs in both conditions. This within-study effect was even more clear in the ratio of the number of brain ICs to their mean RV, where both stationary conditions showed a

noticeable increase over their mobile counterparts (see figure 3d).

Figure 4a shows the number of ICs labeled as “other”. The highest numbers of “other” ICs were attained by the datasets from the low-intensity group followed by the low-to-medium movement intensity group. The *Prediction Error* set, belonging to the medium-to-high movement intensity group, had 17-18 “other” ICs and therefore around 5 “other” ICs fewer than the cluster of the low-to-medium movement intensity group. The lowest number of “other” ICs had the high movement intensity study *Auditory Gait* reaching fewer than 15 “other” ICs. Not aligned with this trend were the *Spot Rotation (mobile)* and the *Beamwalking (mobile)* datasets stemming from the medium-to-high and high movement intensity group, respectively, but having similar amounts of “other” ICs as the medium-to-low cluster.

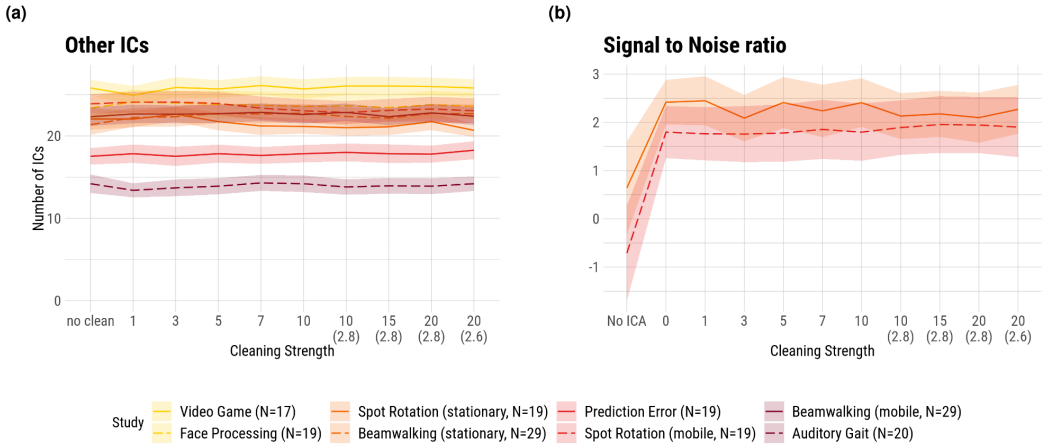


Figure 4: Results for other ICs (a) and exemplary signal-to-noise ratio (SNR; b). The SNR was computed on the *Spot Rotation* study only. Shaded areas depict the standard error of the mean (SE). The numbers on the abscissa refer to the number of iterations of the sample rejection, with a default of 3 SDs as the rejection threshold. The numbers in brackets on the abscissa refer to the rejection threshold in SDs when deviating from the default. “No clean” (left plot) / “0” (right plot) refers to no sample rejection being applied when computing AMICA. “No ICA” (right plot) refers to the SNR values being computed on the raw dataset without any ICA cleaning. The colors denote the movement intensities: yellow - low, orange - low-to-medium, red - medium-to-high, violet - high.

Lastly, in many cases, ICA is used as a means for data cleaning or extracting specific aspects of the signal. Hence, we used the SNR of an ERP in one exemplary study, *Spot Rotation*, to give a practical example of the effect of the use of ICA. As can be seen in figure 4b, this example revealed a lower SNR in the mobile condition.

3.2 The effect of cleaning intensity on the decomposition quality

Increasing the number of cleaning iterations and decreasing the rejection threshold resulted in more samples being rejected (figure 2a). The effect of cleaning iterations reached a ceiling after around 5 iterations and substantial increases in data rejection were only reached when in addition to more iterations the rejection threshold was lowered as well. As the MI is computed on the entire dataset but the ICA did not take the rejected samples into account, all studies showed a monotonous increase in MI with stronger cleaning (figure 2b).

Nevertheless, the magnitude of this effect was only moderate as in most datasets the MI remained almost constant and within the range of their SEs. Two sets (*Prediction Error*, *Spot Rotation (stationary)*) showed a stronger increase in the low iterations but also leveled off after around 7 iterations.

The number of brain ICs did not vary with increased cleaning intensity outside of the SE range for almost all studies (figure 3a). Only the *Audiocue* set showed an increase in low amounts of cleaning (3-5 iterations) of around 2 ICs but this effect reached a ceiling for stronger cleaning. Other sets exhibited small trends in the same direction, but no strong effect could be seen. All datasets except for *Beamwalking (stationary)* exhibited a slight decrease in mean RV with the first few iterations of the cleaning. This trend was within the range of the SE, however, and reached a floor after around 5 to 7 iterations. Combining this small trend with the small trend in the number of brain ICs, we did find a positive effect of time cleaning on the ratio of the number of brain ICs to their mean RV (figure 3d). Here, all studies exhibited an increase in this ratio with increased cleaning, up to a point of around 7 to 10 iterations. Especially the datasets *Video Game* and *Spot Rotation (stationary)* showed around a 20% increase in this ratio.

Considering the SNR, no effect of stronger time-domain cleaning was visible in our study (figure 4b). There was a pronounced effect when generally using the ICA as a means to select the signal of interest by removing all non-brain ICs, but the time cleaning itself did not noticeably affect the SNR as all variations were within the range of the SE.

3.3 The interaction effect of cleaning intensity and movement intensity on the decomposition quality

The changes in the amount of rejected data when cleaning was intensified were very similar for all studies (figure 2a). Only the *Beamwalking (mobile)* set responded stronger to reductions of the rejection threshold, yielding more rejected data compared to the other studies. More differences could be found when looking at the MI, where the *Prediction Error* set showed a stronger increase of MI in early cleaning iterations than the other studies, and the *Spot Rotation (stationary)* set showed a delayed level-off effect (figure 2b). Those datasets, however, contained different levels of movement intensity, and their respective movement intensity group members did not share these differences in trend. The number of brain ICs and their RV values also did not exhibit an interaction effect, as the changes with increased cleaning were either shared across datasets or did not vary systematically with movement intensity (figure 3). Lastly, the exemplary SNR shows a similar result (figure 4b). In the mobile condition, the SNR remains constant across different cleaning intensities. The stationary condition shows slightly more variation but none outside of the SE range and no clear trend either.

4 Discussion

In this study, we investigated the effect of time-domain cleaning and of different levels of mobility on the independent component analysis of EEG data. For this, we used eight datasets from six openly available studies and applied the AMICA algorithm using its own automatic sample rejection option with varying strengths. We evaluated the decomposition

quality on the basis of the component mutual information, the number of brain and muscle ICs as determined by the ICLabel classifier, as well as the brain IC residual variance. In addition, we measured the SNR in an exemplary application. We hypothesized that increased levels of mobility would lead to decreased decomposition quality, stronger cleaning would lead to an increase in decomposition quality, and higher mobility would require more cleaning to reach the best quality. While we found some indication that increased movement resulted in a worse decomposition and that moderate cleaning did improve the decomposition, these results are not conclusive and our hypotheses are not fully supported by the empirical data.

4.1 Ambiguous effect of movement intensity

The effect of movement intensity in the investigated datasets was modest at best. We found variations between studies in different metrics but these variations did not seem to depend on the movement intensity when comparing different studies. It could be assumed that higher movement intensities induce more muscle activity to be captured by the AMICA as muscle ICs. Since the number of ICs is limited by the number of channels, an increase in muscle ICs could come at the cost of a lower number of brain ICs. However, this does not seem to be the case in the present study. Neither did the physiological plausibility of the brain ICs vary systematically with movement intensity, as the best and worst scoring datasets were from the same movement intensity group. This lack of an effect of movement on the ICA decomposition quality could have several reasons: First, it may be the case that there simply is no adverse effect of mobility on data quality. However, another possible explanation could be that although we attempted to standardize the electrode layouts of the different studies, slight differences in the number and placement of electrodes remained, which could have affected the ICA. It has been shown that the number of electrodes does play a role in the number of resulting brain ICs and the quality of the ICA decomposition (Klug & Gramann, 2021), but we would not expect that effect to be so substantial that it results in large differences like the one observed in the present study. Another possible explanation could be that our classification of the datasets with respect to movement intensity primarily reflected the mobility of participants in the respective studies (sitting, standing, sitting and pointing, rotation on the sport, walking). Mobility, however, does not necessarily reflect the impact of movement on data quality or additional noise originating from biological and mechanical sources. Future studies should systematically employ alternative movement classification schemas that investigate different kinds of movements and their impact on the recorded EEG data quality. While slow walking might not impact mobile EEG recordings at all, upper torso and arm movements even in seated participants might be associated with head and cap movements that could lead to electrode micromovements associated with non-stationarity of the signal. Thus, even though participants' mobility is low while sitting and moving their arm, the movement itself might have a stronger impact than walking which can be considered the higher mobility condition. Finally, a major contributing factor to the decomposition quality might not be the mobility of the paradigm itself, but other aspects of the data recording such as the lab environment or the equipment used (Melnik et al., 2017).

Hence, it cannot be ruled out that if identical equipment and paradigms are used, the anticipated negative effect of movement on the decomposition quality might be found. We had the opportunity to test this in addition to our larger comparison across studies since we

included two studies that contained both a mobile and a stationary condition, allowing for a direct comparison of the results within the two studies. Both studies showed a decrease in the number of brain ICs in the mobile condition and an increase or no change in RV, resulting in a noticeable decrease in the ratio of the number of brain ICs to their mean RV. Although this did not hold true for other metrics such as the MI or the number of muscle ICs, this reduction in quality was also found in the SNR values of the *Spot Rotation* study.

Taken together, it is likely that in a lab environment where everything is kept constant except the mobility of the participant, an increase in movement intensity will have a negative impact on the data. This negative impact, however, is less pronounced than anticipated, as we did not find it when comparing different studies from different lab setups.

4.2 Moderate cleaning improved the decomposition

The impact of cleaning intensity on the quality of the decomposition was smaller than anticipated. While different datasets scored differently on various metrics, these scores stayed for the most part within the SE range and not all datasets exhibited a noticeable effect. We did observe a positive effect of moderate cleaning on the number of brain ICs and their RV values, but the magnitude was limited, and some datasets exhibited almost no change. Additionally, some datasets required relatively strong cleaning to reach their maxima, while others showed a negative impact of too strong cleaning.

This might be because the AMICA algorithm is more robust than anticipated and suitable for capturing or ignoring artifacts even without substantial cleaning. Especially considering physiological activity not stemming from the brain, removing single samples or small patches from the data does not remove the general activity of these sources. Thus, the AMICA algorithm will have to capture this activity regardless, and removing samples might not help much. Essentially, what researchers may consider artifacts in the data (such as eye movements, muscle activity, or recurrent cable sway from gait) is not necessarily an artifact for the underlying ICA model. If these signals are systematic and can be effectively modeled by the ICA, they will not be removed from the data and neither is it necessary to remove them beforehand. Artifact is a term from the user's perspective - the model is blind to such labels. Thus, only data that contains large, transient spikes or excessively strong other artifacts that can not easily be modeled by ICA but can be removed in the time-domain would benefit from cleaning. This could for example be time points where the participant was touching the EEG cap or other equipment, or moments where a virtual reality display is taken on or off. However, assuming that these artifacts are limited to periods of breaks or happen before or after the experiment, it might be suitable to just remove all non-experiment segments of the data and perform only minor additional time-domain cleaning before ICA.

If it is essential to capture as many brain components as possible because one is interested in deep or unusual regions of interest and intends to perform source-level analysis, it might be justified to clean the data more strongly. However, in these cases, one must keep in mind that the resulting decomposition will not be able to fully capture the artifacts as it was not computed with them included. This may result in no relevant change in the actual measure to investigate, such as ERPs, as could be seen in the absent effect of time-domain cleaning on the exemplary SNR of the *Spot Rotation* data when all non-brain ICs were removed.

4.3 More movement did not require more cleaning

As we expected an adverse effect of movement on the decomposition quality and a positive effect of time-domain cleaning, we also assumed that more cleaning would be necessary for data containing more movement. To our surprise, we found no trend indicating an interaction between the movement intensity and the required time-domain cleaning on the resulting decomposition quality. While we did find some indication for main effects, these trends were mostly shared across datasets in direction and magnitude. There were some exceptions but these were single datasets and their trend was usually not shared by the other dataset with the same class of movement intensity, and even if such an effect appeared, it was small. In accordance with our discussion of the expected main effects above, this again suggests that movement and thereby movement artifacts are less impactful on the ICA decomposition than previously assumed, and no substantially different time-domain cleaning is necessary for mobile EEG studies.

4.4 Limitations and possible improvements

As a first and major limitation, this study can only discuss the effects of the included datasets and does not necessarily generalize to other lab setups and experimental protocols. It was difficult to find an effect of variations in data processing without controlling for general data quality. A control for data quality, however, is not straightforward and would most likely only be possible when all investigated datasets share the same laboratory setup and recording equipment (Melnik et al., 2017), as well as experiment paradigm (such as an oddball task). Hence, although we tried to find a suitable amount of representative datasets with varying protocols, it would be favorable to have different studies repeat the same protocol in varying movement conditions. A taxonomy of different movement types such as gait, balancing, arm reaching, or tool use would be useful in this case, including a specification of the expected impact of these movement types on the EEG electrodes. Such a large dataset with consistent recording quality could help shed light on the smaller effects we found, especially since the results contradict our expectations.

A second limitation is the measure of decomposition quality. We used the number of brain ICs as classified by ICLabel, and their RV as a proxy for decomposition quality, but this approach has two limitations: i) The body of data that ICLabel used to train the classifier did not contain sufficient examples from mobile experiments, meaning that the classification results might not be fully reliable in our context. Extending the classifier to MoBI or mobile EEG studies would alleviate this issue. ii) RV values might also be problematic to interpret, especially those of non-brain sources, which is why we did not take those into account. However, especially in the MoBI context, having more physiologically plausible muscle and eye ICs would also be of value, and this is impossible to measure using the current version of dipole fitting in EEGLAB. In the future, this can be done using HArtMuT, a head model that contains sources for eyes and muscles and can thus lead to more reliable estimates of the IC source and its RV (Harmening et al., 2022), but was not yet available at the time of this study. Another option to take into account is to investigate the SNR after data cleaning in more depth. This, however, would also require the same study to be repeated in different movement conditions, akin to the *Spot Rotation* SNR evaluation we performed. This would shed light on more practical implications of the investigated effects.

A third limitation could be that we only used one method for time-domain cleaning. It is possible that other cleaning options could lead to different results. However, we believe that since the AMICA auto sample rejection uses its own objective metric, it is unlikely that the cleaning results will be substantially improved when using other algorithms. A separate investigation of the effect of a proposed time-domain cleaning algorithm in comparison with the AMICA auto sample rejection found no noticeable difference (Klug et al., 2022).

4.5 Conclusions

In our investigation of the effect of time-domain cleaning and movement intensity on the quality of the ICA decomposition, we did not find substantial evidence to support our hypotheses. While the expected adverse effect of movement on the data could be seen within studies, it is inconclusive between studies, pointing to the fact that lab setup, equipment, and possibly the paradigm itself might have a greater impact on the decomposition quality than the movement intensity. Additionally, while we did find some evidence that moderate cleaning prior to ICA computation improves the decomposition, this effect was far weaker than anticipated and it did not vary systematically with movement intensity in our study. This suggests that the AMICA algorithm is very robust and can handle artifacts even with limited data cleaning.

We thus recommend not to remove substantial parts of the data using time-domain cleaning before running AMICA. Moderate amounts of cleaning such as 5 to 10 iterations of the AMICA sample rejection starting after 2 iterations with the default 3 SDs as threshold and 3 iterations between rejections will likely improve the decomposition in most datasets, irrespective of the movement intensity. Only in special circumstances, strong cleaning will be relevant and more beneficial.

Acknowledgments

We are thankful to the researchers who made their datasets freely available. Without them, this investigation would not have been possible.

Data availability statement

The data used in this study is available for download as laid out in the *Datasets* section.

References

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.

Cavanagh, J. F., & Castellanos, J. (2021). *Continuous gameplay of an 8-bit style video game* [Data set]. Openneuro. https://doi.org/10.18112/OPENNEURO.DS003517.V1.1.0

Chang, C. Y., Hsu, S. H., Pion-Tonachini, L., & Jung, T. P. (2020). Evaluation of Artifact

Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings. *IEEE Transactions on Biomedical Engineering*, *67*(4), 1114–1121.

Debener, S., Minow, F., Emkes, R., Gandras, K., & de Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology*, *49*(11), 1617–1621.

de Cheveigné, A. (2020). ZapLine: A simple and effective method to remove power line artifacts. *NeuroImage*, *207*, 116356.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PloS One*, *7*(2), e30135.

Gehrke, L., Akman, S., Chen, A., Lopes, P., & Gramann, K. (2021). *Prediction Error* [Data set]. Openneuro. https://doi.org/10.18112/OPENNEURO.DS003846.V1.0.1

Gramann, K., Ferris, D. P., Gwin, J., & Makeig, S. (2014). Imaging natural cognition in action. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *91*(1), 22–29.

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T. P., Lin, C. T., Liao, L. D., & Makeig, S. (2011). Cognition in action: Imaging brain/body dynamics in mobile humans. *Reviews in the Neurosciences*, *22*(6), 593–608.

Gramann, K., Hohlefeld, F. U., Gehrke, L., & Klug, M. (2021). Human cortical dynamics during full-body heading changes. *Scientific Reports*, *11*(1), 18186.

Gwin, J. T., Gramann, K., Makeig, S., & Ferris, D. P. (2010). Removal of movement artifact from high-density EEG recorded during walking and running. *Journal of Neurophysiology*, *103*(6), 3526–3534.

Harmening, N., Klug, M., Gramann, K., & Miklody, D. (2022). HArtMuT - Modeling eye and muscle contributors in neuroelectric imaging. In *bioRxiv* (p. 2022.08.19.504507). https://doi.org/10.1101/2022.08.19.504507

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent Component Analysis*. John Wiley

& Sons.

Jacobsen, N. S. J., Blum, S., Witt, K., & Debener, S. (2021). A walk in the park? Characterizing gait-related artifacts in mobile EEG recordings. *The European Journal of Neuroscience*, *54*(12), 8421–8440.

Jungnickel, E., Gehrke, L., Klug, M., & Gramann, K. (2019). Chapter 10 - MoBI—Mobile Brain/Body Imaging. In H. Ayaz & F. Dehais (Eds.), *Neuroergonomics* (pp. 59–63). Academic Press.

Jungnickel, E., & Gramann, K. (2016). Mobile Brain/Body Imaging (MoBI) of Physical Interaction with Dynamically Moving Objects. *Frontiers in Human Neuroscience*, *10*(June), 306.

Klug, M., & Gramann, K. (2021). Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *The European Journal of Neuroscience*, *54*(12), 8406–8420.

Klug, M., Jeung, S., Wunderlich, A., Gehrke, L., Protzak, J., Djebbara, Z., Argubi-Wollesen, A., Wollesen, B., & Gramann, K. (2022). The BeMoBIL Pipeline for automated analyses of multimodal mobile brain and body imaging data. In *bioRxiv* (p. 2022.09.29.510051). https://doi.org/10.1101/2022.09.29.510051

Klug, M., & Kloosterman, N. A. (2022). Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG. *Human Brain Mapping*, *43*(9), 2743–2758.

Kothe, C. A. E., & Jung, T.-P. (2015). Artifact removal techniques with signal reconstruction. *U.S. Patent*, 047462 A9.

Makeig, S., Gramann, K., Jung, T. P., Sejnowski, T. J., & Poizner, H. (2009). Linking brain, mind and behavior. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *73*(2), 95–100.

Melnik, A., Legkov, P., Izdebski, K., Kärcher, S. M., Hairston, W. D., Ferris, D. P., & König, P. (2017). Systems, subjects, sessions: To what extent do these factors influence EEG data? *Frontiers in Human Neuroscience*, *11*(March), 1–20.

Onton, J., & Makeig, S. (2006). Information-based modeling of event-related brain dynamics. *Progress in Brain Research*, *159*, 99–120.

Palmer, J. A., Kreutz-delgado, K., & Makeig, S. (2011). *AMICA : An Adaptive Mixture of Independent Component Analyzers with Shared Components*. 1–15.

Peterson, S., & Ferris, D. (2021). *Perturbed beam-walking task* [Data set]. Openneuro. https://doi.org/10.18112/OPENNEURO.DS003739.V1.0.2

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*(May), 181–197.

Wagner, J., Martinez-Cancino, R., Delorme, A., Makeig, S., Solis-Escalante, T., Neuper, C., & Mueller-Putz, G. (2019). High-density EEG mobile brain/body imaging data recorded during a challenging auditory gait pacing task. *Scientific Data*, *6*(1), 211.

Wakeman, D. G., & Henson, R. N. (2021). *Face processing EEG dataset for EEGLAB* [Data set]. Openneuro. https://doi.org/10.18112/OPENNEURO.DS002718.V1.0.5

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods*, *250*, 34–46.

Winkler, I., Debener, S., Muller, K. R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2015-Noem*, 4101–4105.

3 The BeMoBIL Pipeline

The BeMoBIL Pipeline for automated analyses of multimodal mobile brain and body imaging data

Klug, M., Jeung, S., Wunderlich, A., Gehrke, L., Protzak, J., Djebbara, Z., Argubi-Wollesen, A., Wollesen, B., & Gramann, K. (submitted). The BeMoBIL Pipeline for automated analyses of multimodal mobile brain and body imaging data.

The BeMoBIL Pipeline for automated analyses of multimodal mobile brain and body imaging data

Klug, M.[1], Jeung, S.[1,2,3], Wunderlich, A.[1], Gehrke, L.[1], Protzak, J.[1], Djebbara, Z.[1,4], Argubi-Wollesen, A.[5], Wollesen, B.[6], & Gramann, K.[1]

1) Biopsychology and Neuroergonomics, Institute of Psychology and Ergonomics, TU Berlin, Berlin, Germany
2) Kavli Institute for Systems Neuroscience, Norwegian University of Science and Technology, Norway
3) Max-Planck Institute for Human Cognitive and Brain Sciences, Germany
4) Department of Architecture, Design, Media and Technology, Aalborg University, Aalborg, Denmark
5) exoIQ GmbH, Hamburg
6) Institute for Human Movement Science, Faculty of Psychology and Human Movement Science, University of Hamburg, Hamburg, Germany

Keywords: *electroencephalography, mobile brain/body imaging, signal processing, artifact removal, preprocessing, replicability, standardization*

Abstract

Advancements in hardware technology and analysis methods allow more and more mobility in electroencephalography (EEG) experiments. Mobile Brain/Body Imaging (MoBI) studies may record various types of data such as motion or eye tracking in addition to neural activity. Although there are options available to analyze EEG data in a standardized way, they do not fully cover complex multimodal data from mobile experiments. We thus propose the BeMoBIL Pipeline, an easy-to-use pipeline in MATLAB that supports the time-synchronized handling of multimodal data. It is based on EEGLAB and fieldtrip and consists of automated functions for EEG preprocessing and subsequent source separation. It also provides functions for motion data processing and extraction of event markers from different data modalities, including the extraction of eye-movement and gait-related events from EEG using independent component analysis. The pipeline introduces a new robust method for region-of-interest-based group-level clustering of independent EEG components. Finally, the BeMoBIL Pipeline provides analytical visualizations at various processing steps, keeping the analysis transparent and allowing for quality checks of the resulting outcomes. All parameters and steps are documented within the data structure and can be fully replicated using the same scripts. This pipeline makes the processing and analysis of (mobile) EEG and body data more reliable and independent of the prior experience of the individual researchers, thus facilitating the use of EEG in general and MoBI in particular. It is an open-source project available for download at https://github.com/BeMoBIL/bemobil-pipeline which allows for community-driven adaptations in the future.

Introduction

Electroencephalography (EEG) has rapidly evolved into a popular brain imaging method in various areas outside of traditional medical or psychological research. Recent developments in amplifier technology and improvements in data-driven analyses enable mobile EEG and Mobile Brain/Body Imaging studies (MoBI; Gramann et al., 2011, 2014), accelerating this progress. These methods do not only allow for participant movement (mobile EEG) but specifically address the interplay of active behavior and neural dynamics (MoBI) to better understand the neural foundation of embodied cognition (Jungnickel et al., 2019; Makeig et al., 2009). The electrical activity of the brain underlying cognitive processes is now investigated outside traditional laboratories in diverse scientific fields such as human factors (e.g. Protzak & Gramann, 2018), architecture (e.g. Djebbara et al., 2019), sports science (e.g. Büchel et al., 2021), clinical neuroscience (e.g. Short et al., 2020) and many more. As a result, researchers from various fields with manifold scientific and technical backgrounds handle data from multimodal brain-behavior assessments. This has led to a multitude of analytic approaches and analysis pipelines. In order to draw synergy from these exciting developments, we have identified the need for a standard pipeline that offers researchers the opportunity to focus on the actual research questions instead of overcoming methodological obstacles over and over again. To this end, we propose an easy-to-use and adaptive multimodal data processing pipeline, the BeMoBIL Pipeline, that enables documented, traceable, and objective data processing.

Despite some efforts to develop best practice guidelines (Chaumon et al., 2015) and to establish standard processing pipelines, especially for stationary recordings (Bigdely-Shamlo et al., 2015; da Cruz et al., 2018; Gabard-Durnam et al., 2018; Pedroni et al., 2019; Pernet et al., 2021; Rodrigues et al., 2021), a common basis in EEG data processing is still lacking (Robbins et al., 2020). In particular, pipelines targeting methodological obstacles such as data synchronization and subsequent processing of diverse data modalities (e.g. eye tracking data, motion capture) with different sampling rates are currently unavailable. In addition, more ecologically valid mobile EEG and MoBI protocols often come with increased noise levels in the recordings due to mechanical artifacts as well as biological activity stemming from the movement itself (Gramann et al., 2011, 2021; Richer et al., 2020). Handling increased noise levels in the recorded signal also requires new analysis approaches that often utilize information about the ongoing movement and thus rely on accurately synchronized multimodal recordings and specific analyses (Jungnickel & Gramann, 2016). While one toolbox for multimodal data analyses exists (MoBILAB; Ojeda et al., 2014), it is not supported anymore and lacks central analysis functions that are important for in-depth EEG processing and synchronized multimodal event extraction. Such data-driven event extraction, however, can be of central importance for highly realistic recordings. For example, eye blinks might be used for blink-based event-related analysis of EEG data when no external visual stimulation is available in natural outdoor experiments (Wascher et al., 2014; Wunderlich & Gramann, 2021).

This multitude of complex data types and the lack of common processing standards can lead to subjective, unjustified, or laboratory-specific parameter choices (e.g. filter design, artifact handling). As a consequence, peer-review processes can become complicated or, in the worst case, these factors can result in serious reproducibility issues (Cohen, 2017;

Kappenman & Keil, 2017; Larson & Moser, 2017; Open Science Collaboration, 2015). We believe that standards in multimodal neuroscientific computing are the inevitable prerequisite for researchers to converge on basic principles in the field, further our understanding of human brain function, and foster more efficient research. The goal of the present paper is thus to introduce a replicable, open-access, standardized, and transparent analysis approach to EEG data in general and to multimodal mobile EEG/MoBI data analyses specifically. Our pipeline is intended to serve as an analysis basis that can be adopted and continuously developed by the scientific community.

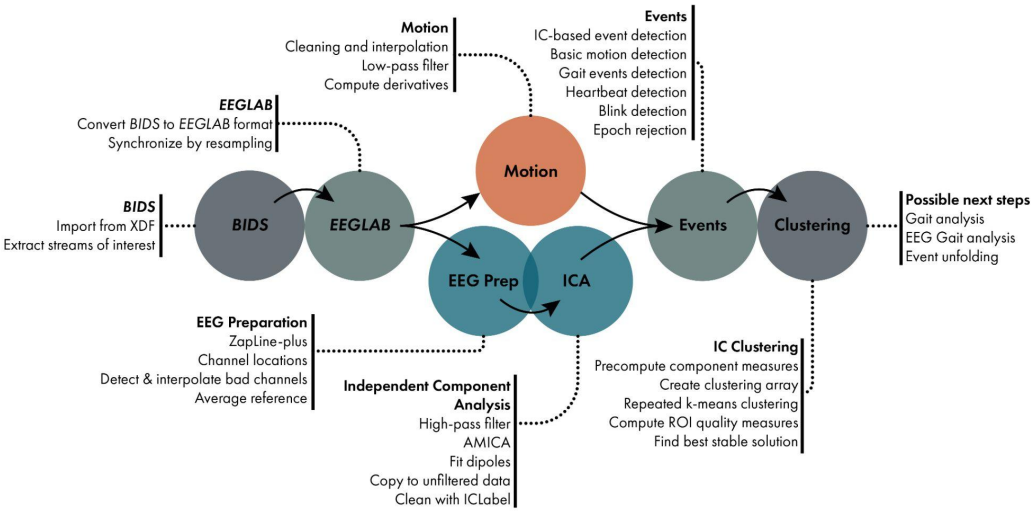


Figure 1: The BeMoBIL Pipeline workflow. Data is imported into BIDS format and then converted into synchronized EEGLAB files. EEG data is preprocessed, cleaned, and analyzed using ICA. Motion data is cleaned and prepared for analysis. Event markers can be extracted from different data modalities and repeated clustering can be used for robust source-level analysis. The individual steps are explained in detail in the respective sections.

The BeMoBIL Pipeline runs on MATLAB as we incorporated standard data processing routines from EEGLAB (Delorme et al., 2011; Delorme & Makeig, 2004) and Fieldtrip (Oostenveld et al., 2011), two of the currently leading EEG data processing toolboxes. Our pipeline can be used for the exclusive analyses of EEG data as well as for multimodal data processing including different data streams such as motion capture, eye tracking or force plate data. We provide raw data import and synchronization functions for the standard Brain Imaging Data Structure format (BIDS; Gorgolewski et al., 2016) that allows for easy data sharing. EEG data processing is available in the form of basic preprocessing routines (e.g. re-referencing, line noise removal, channel interpolation) as well as advanced artifact handling and source separation scripts (e.g. independent component analysis, ICA, or early-fusion approaches) and subsequent group-level source analysis methods. Motion data processing is integrated, including data preprocessing and the creation of derivatives. As a special feature for MoBI analyses, event extraction from motion data (e.g. heel strikes), eye tracking (e.g. blinks), electrocardiography (ECG, heartbeats), and EEG (using independent components representing eyes movements, heartbeats, or gait) are available. Various parameters can be adjusted for each processing step while the pipeline comes with informed recommendations for all steps. We encourage independent plausibility checks through automated data visualization at several processing milestones. All routines and parameter

selections are clearly described and documented in a wiki on the GitHub repository (https://github.com/BeMoBIL/bemobil-pipeline). The source code is freely available and extensions and improvements from the community are welcome. Figure 1 depicts the general structure of the pipeline. In the following sections, each pipeline element is described in detail.

Time-synchronized data collection and import

Handling simultaneously recorded multimodal data requires special considerations from the stage of data collection and import. A harmonized, time-synchronized representation of EEG and motion data is a prerequisite for joint analysis of the two modalities. Addressing time synchrony is especially non-trivial in MoBI studies because a dataset often consists of multiple modalities of different sampling rates and different levels of time precision. For instance, although EEG recording systems typically yield highly regular inter-sample intervals, motion data may have irregular temporal distances between samples and may even completely miss latency information for each sample. One solution to record such multimodal data and preserve all available information is Lab Streaming Layer (LSL, available at https://github.com/sccn/labstreaminglayer). LSL is an open-source data streaming protocol that allows the flexible definition and recording of data streams from various sources such as different EEG amplifiers, different motion capture hardware, eye gaze recordings, experiment event markers, and more. Using the LSL recorder, the multimodal dataset can be stored in the extensible data format (XDF), containing all selected data streams with definitions, samples, and timestamps for each sample.

Standardizing MoBI data

Although the XDF format is commonly used in the context of MoBI studies, it does not easily allow sharing the recorded data in a standardized way, as the recordings can contain self-defined data types with non-standard metadata and the XDF format is not based on common consensus. Additionally, the XDF files may not contain metadata about the experiment such as participant information that can be relevant for analysis. A key solution to easier data sharing is the BIDS standard. Initially covering standards for sharing fMRI datasets (Gorgolewski et al., 2016), the BIDS community has continued its effort to include more types of data in the framework by means of modality-specific extensions. The BeMoBIL Pipeline is designed to operate on datasets that adhere to the standards provided by the EEG-BIDS (Pernet et al., 2019) and MOTION-BIDS (in progress, see https://bids.neuroimaging.io/get_involved.html) extensions. Using BIDS data standards lowers the risk of data rot and makes data sharing easier by adhering to the FAIR principles for data management (Wilkinson et al. 2016), addressing the findability, interoperability, and reusability aspects. This in consequence can contribute to mitigating low reproducibility of scientific findings (Cohen, 2017; Kappenman & Keil, 2017; Open Science Collaboration, 2015).

To this end, the pipeline provides two features: One is the function *bemobil_xdf2bids* to convert data contained in XDF files into BIDS formatted data, and the other is the function *bemobil_bids2set* for converting BIDS data into EEGLAB-compatible data structures. The created BIDS dataset retains all necessary timing-relevant information from all streams,

especially all timestamps of the samples, while applying only minimal change to the data itself, for example converting motion data orientation values that are represented in quaternion values into Euler angles, or replacing samples with a defined value for missing samples (e.g. zeros or -999) with not-a-number (NaN) values. The function supports the flexible use of parameters to load XDF files and allows the addition of metadata such as information about the participants, the EEG recording, the motion tracking system, its manufacturer, or the spatial axes layout to be stored together with the dataset.

When importing this BIDS dataset into EEGLAB using *bemobil_bids2set*, the function aligns the EEG and other data streams by first resampling the EEG data to a given target sampling frequency and then resampling other data to match the latency of EEG samples (see figure 2). As it can be assumed that the EEG data is recorded with high precision and equidistant samples, it is resampled to the desired sampling rate using the EEGLAB function *pop_resample*, which uses the filter-based *resample* MATLAB function internally. Resampling the other data types to the same rate is done by default using the *pchip* interpolation option of the fieldtrip function *ft_resampledata*, including an anti-aliasing filter if the data is downsampled. This is chosen as it is not always guaranteed that the nominal sampling rate of the other data types is accurate (e.g. when recording motion from virtual reality environments, the sampling rate is dependent on the performance of the rendering and the refresh rate of the display), and the samples are not always evenly distributed. Even in cases of equidistant sampling from reliable measurement devices, the level of precision using filter-based resampling may not be high enough for very long data sets containing millions of samples, leading to a shift between EEG and other data of several hundreds of milliseconds towards the end of the recording. In contrast, using interpolation of other data streams to align with the EEG samples preserves the relative temporal structure between the different modalities within the precision of one sample at all times and is thus favored in our use case. As motion data usually varies mainly in much lower frequencies than EEG, imprecisions introduced during the interpolation should not be problematic for downstream analysis. However, the requirements unique to additional modalities other than EEG and motion (such as eye tracking) may be at a disadvantage with this approach due to possible distortion of the signal, especially in the high-frequency range. In these cases, the exact sampling rate can optionally be entered and used for filter-based resampling instead of the interpolation approach.

For every recording then, a plot is created that shows the first and last event and one channel of all modalities, both directly imported from the XDF file as well as after the EEGLAB import and data alignment process. These plots can be used to verify the integrity of the temporal structure of the multiple data streams. As a final step of the import, all data modalities are made the exact same length as the EEG data, even in cases where there was no recording of a given modality in some sessions. To this end, all missing samples are filled with not-a-number (NaN) values. This, in combination with the previous resampling step, leads to fully synchronized data structures in EEGLAB with the exact same amount of samples and identical event markers in all modalities. Such synchronized data allows for analyzing event-related activity or the creation of event markers from one modality that can be copied to others, such as extracting gait event markers from motion (see section *Extracting gait parameters from motion data*). A generic template and a specific example script to use this complete import processing are available in the repository.

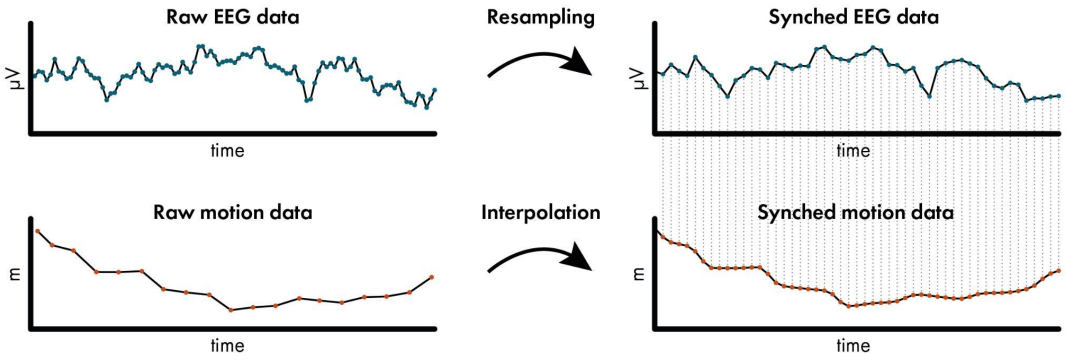


Figure 2: Example impression of the import synchronization. When importing BIDS data using *bemobil_bids2set*, EEG data is resampled using a filter-based method, while other data such as motion is interpolated to the same timestamps, leading to a fully synchronized dataset in EEGLAB. In this 200 ms example data, the EEG data was originally sampled at 500 Hz and the motion data at 90 Hz, and both are resampled/interpolated to 250 Hz. Plots of the raw and synchronized data that show the first channel of all modalities are created around the first and last event in each data file, allowing the inspection of the synchronization. This exemplary visualization is taken from subject 6 in the NeSitA example data that is available with the BeMoBIL Pipeline. For details, see section *Data import and time synchronization*.

Data cleaning and processing

The BeMoBIL Pipeline contains a variety of scripts and wrappers for cleaning and processing EEG, motion, and eye tracking data, all configured by one central file containing all relevant parameters and explanations thereof. EEG parameter default values are optimized for data from mobile experiments, however, they can also be easily adjusted for data from stationary experiments. Processing steps are documented within the data structures themselves, as well as by visualizing the outcome of every important step. As for the import, template files for the processing exist for all modalities and the configuration file. The pipeline wrapper scripts read the raw imported data and create several intermediate folders for preprocessing, other processing, and the final resulting data files. These folders and filenames can be adjusted in the configuration file. If the processing is stopped and restarted, it will by default load already created files instead of computing them again. This can be overridden if necessary. As it was shown that the data precision level has important effects on the processing (Bigdely-Shamlo et al., 2015), we ensure that double precision is used from the start and throughout the processing by selecting the appropriate EEGLAB option at several points within the pipeline.

EEG data processing

The BeMoBIL Pipeline allows for fully automatic processing of EEG data from raw files to the final clean datasets including ICA information. This raw data can be obtained either via our own import (see above) or from any other importer. As a very first step in EEG processing, it is recommended, but not mandatory, to remove the segments before and after the experiment, as well as breaks during the experiment. This can be done automatically based on experiment event markers, or manually if no such event markers are included in the data

structure. This step is important because these segments can contain very strong artifacts, e.g. excessive movement from stretching, or electrical or mechanical artifacts from touching cables, the cap, or putting off and on equipment such as a virtual reality head-mounted display. These artifacts can affect subsequent analysis steps negatively (such as channel rejection or ICA) and should thus be discarded. If removal based on event markers is impossible or manual removal is preferred for other reasons, it should be done once, and the removed indices should be stored for every participant. These indices can be obtained using the *eegh* command after removal, and can subsequently be applied automatically again. This is important on the one hand for reproducible processing, and, on the other hand, to be able to maintain synchronized datasets from different modalities. If segments are removed in the EEG data in this step, these segments must also be removed in motion and other physiological data for their respective processing. With only relevant experiment segments remaining in the EEG data, it is first preprocessed and then subjected to an ICA.

EEG preprocessing

Preprocessing of EEG data in the BeMoBIL Pipeline is done using the *bemobil_process_all_preprocessing* function. This is a wrapper that incorporates all necessary processing steps from the raw loaded EEG set (all blocks merged together and non-experiment parts removed) up to the preprocessed dataset which has line noise removed, channels interpolated, and the data re-referenced to the average. It stores intermediate files on disk in the location provided in the configuration file parameter *bemobil_config.EEG_preprocessing_data_folder* and plots several analytics plots which are saved alongside their respective files. Exemplary visualizations of the preprocessing can be seen in figure 3.

As a first step, some basic preparations are performed on the data using the *bemobil_process_EEG_basics* function. First, the EEG structure is filled with ur-data, specifically for event markers and channel locations. This is to ensure that in subsequent processing, the original event structure can always be recovered. Then, unused electrodes are removed. This can be set in the config file using the *bemobil_config.channels_to_remove* entry. This is to ensure that channels without channel location or channels in the montage that do not record data have no impact on downstream analysis. If the data is not already at the correct sampling rate, it is then resampled to the frequency declared in *bemobil_config.resample_freq*.

Frequency-specific noise is removed with Zapline-plus (Klug & Kloosterman, 2022). Zapline-plus is an EEGLAB plugin that removes spectral artifact peaks automatically. It includes a detector for artifactual peaks, chunks data to account for non-stationarity of the noise, runs Zapline (de Cheveigné, 2020) with the automatically detected number of components to remove, and creates a comprehensive analytics plot. Zapline removes noise by splitting the data into an originally clean (data A) and a noisy part (data B) by filtering the data once with a notch filter (A) and once with the inverse of said notch filter (B). It then uses a spatial filter to remove noise components in the noisy part (B) to get a cleaned version of that part (B'). Finally, the two clean parts (A and B') are added back together to result in a cleaned dataset with full rank and full spectrum except for the noise. Zapline is preferred over a notch filter because it preserves the spectrum, and it is preferred over a simple spatial filter because it preserves the full data rank. It was also shown to have more powerful

cleaning capabilities than the *cleanLine* plugin of EEGLAB (Miyakoshi et al., 2021). If *bemobil_config.zaplineConfig.noisefreqs* is declared empty, the function will use the complete automatic adaptation. This is the default and recommended. However, parameters can be adjusted if the cleaning does not work as intended. See (Klug & Kloosterman, 2022) for details about the processing and parameter tweaking. This step can be avoided by setting the whole *bemobil_config.zaplineConfig* field to empty.

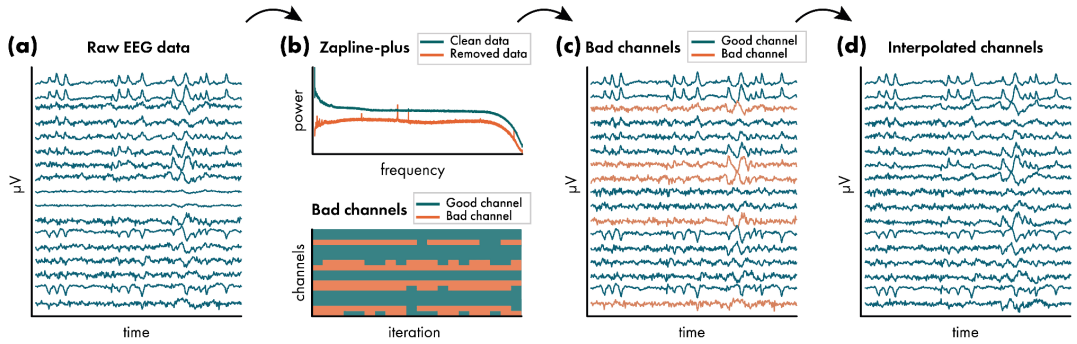


Figure 3: Exemplary impressions of the visualizations of the EEG preprocessing. During preprocessing of EEG data, several plots are created to allow an inspection of the workflow. (a) The raw data is plotted directly after import in six 10-second periods equally spaced throughout the entire data. This example shows the third such period in subject 76 of the visual discrimination datasets that are available with the BeMoBIL Pipeline. Only the first channels are shown in this plot for visualization purposes. Note that the scale of each channel is different and depends on the overall activity, and can thus not be interpreted. (b) Diagnostic plots of Zapline-plus (top) and the repeated bad channel detection (bottom) are available. This example shows a simplified version of plots from the same dataset as in (a). (c) In a plot similar to (a), the data that is cleaned with Zapline-plus is visualized with detected bad channels denoted in red color. (d) Lastly, the completely preprocessed data with interpolated bad channels is visualized again in a plot similar to (a). For details of the preprocessing, see section *EEG preprocessing*.

After this first step of data cleaning, channel locations are added. Here, channel names can be changed in case they were named incorrectly or contain an unnecessary prefix using the *bemobil_config.rename_channels* setting. This is to ensure that the lookup tables for channel locations can operate correctly, even if the channel names were incorrect. At this point, a reference channel can also be added with zero entries when declared in *bemobil_config.ref_channel*. This allows feeding back the data of the reference channel when the data is re-referenced to the average in a later step, similarly to the *Full Rank Average Reference* EEGLAB plugin. This is done before importing channel locations so the reference channel can also be located. Then, if channel locations were not already loaded during the import, they can be imported at this stage using the *bemobil_config.channel_locations_filename* entry. In this step, we either look up locations in the standard 10-20 system (when no filename is provided) or use a file provided by an electrode location digitizer. In the latter case, if a reference channel was declared before, the file must contain the location of the reference with the name specified above. Lastly, the channel types are declared to be either EEG (default), EOG (can be provided in *bemobil_config.eog_channels* and will be ignored in both bad channel detection and re-referencing, see below), or REF (if a reference channel was entered above). These steps form the foundational preparations for downstream cleaning and analysis, and data files will be saved with the name provided in *bemobil_config.basic_prepared_filename*.

As the next step, bad channels will be detected and interpolated in a repeated process using the *clean raw data* plugin of EEGLAB, which uses the algorithm proposed in the PREP pipeline (Bigdely-Shamlo et al., 2015). We do not use all options in this function and our processing slightly differs from the PREP pipeline, which is why we make use of the general concepts only. One key issue in this process is the order of the re-referencing and interpolation, in which the average reference should not contain strong artifacts anymore, but re-referencing is necessary for the detection of bad channels. The detection of bad channels also recommends a high-pass filter of 0.5Hz cutoff frequency, which is not part of our preprocessing in order to preserve as much data information as possible, and since spectral filters should be adapted exactly to the needs of the final analysis (Widmann et al., 2015). For these reasons, we decided to split the detection of bad channels from their interpolation.

Using the *bemobil_detect_bad_channels* function, the data is first re-referenced to the average, excluding EOG channels as defined previously. This is to have an approximation of the final data but includes the impact of bad channels. This average reference will not be used later on, but only to detect the bad channels. Here, we either make use of the previously added reference channel or, as a fallback, the *Full Rank Average Reference* EEGLAB plugin, to preserve the full rank of the data and thus as much information as possible. This is automatically detected in the *bemobil_avref* function.

Subsequently, we repeatedly run the *clean_artifacts* function of the *clean raw data* EEGLAB plugin, with the number of repetitions specified by the *bemobil_config.chan_detect_num_iter* parameter. This is necessary because *clean_artifacts* uses a random sample consensus (RANSAC) approach that does not necessarily converge to the same results when repeated. The function stores the sampling in a micro cache that will be accessed when restarting the function without restarting MATLAB or clearing the micro cache beforehand, resembling a stable result. But when using the function with a cleared micro cache, the detected channels might differ. To ensure a reproducible bad channel detection, we thus clear the micro cache and repeat the detection several times, with a recommended minimum of 10 iterations. Only channels that were flagged as ‘bad’ more than a given proportion of the processed data (specified in *bemobil_config.chan_detected_fraction_threshold*) are then detected for final removal. We exclude all EOG channels from the detected bad channels because their statistical properties will often lead to false positive detection.

Within the *clean_artifacts* function, the data is split into windows of five seconds, and robust interpolations of each channel are computed based on the RANSAC sampling of surrounding channels. We do not make use of the time-domain sample removal or the Artifact Subspace Reconstruction (ASR) options, as we are only interested in detecting bad channels at this point. In our detection, five parameters can be adjusted:

- *bemobil_config.chancorr_crit* is the main parameter. This is a correlation threshold. If a channel is correlated less than this value to its own robust estimate based on the surrounding channels, it is considered abnormal in the given time window. Recommended are correlation values of 0.75 (rather lax) to 0.85 (rather strict).
- *bemobil_config.chan_max_broken_time* sets the maximum proportion of time windows a given channel may be flagged as bad before it is detected as bad in the final output per iteration. Recommended are values from 0.2 (20% of the time, strict) to 0.5 (50% of the time, lax).

- *bemobil_config.flatline_crit* uses a criterion for detecting channels that are flat. This is recommended to be set to 'off' since i) flat channels will not correlate with their interpolation, and ii) sometimes, especially in MoBI, data may be lost without necessitating the removal of the complete channel.
- *bemobil_config.line_noise_crit* rejects channels that have increased noise. However, line noise was removed by Zapline-plus in a previous step and it is recommended to be kept 'off', too.
- *bemobil_config.num_chan_rej_max_target* determines the fraction of channels that can be maximally removed (e.g. 1/5). This is to ensure that even in the case of very noisy data or incorrect bad channel detection, the processing does not remove too many channels to reconstruct them.

Subsequently, the detected bad channels are interpolated in the dataset that is not yet re-referenced or filtered using spherical spline interpolation in EEGLAB in the *bemobil_interp_avref* function. When this is done, the rank of the data matrix is reduced by the number of interpolated channels and this information is stored within the data structure. As a final step, the data is re-referenced to the average of all scalp channels (excluding EOG channels; Delorme et al., 2012), maintaining the full rank, as within the bad channel detection: Either the reference channel was declared previously, which means it was added with zero-entries. In this case, it will now be filled and available for analysis. Or no reference was declared, in this case, we follow the approach of the *Full Rank Average Reference* EEGLAB plugin: a new dummy channel with zeros is added, the data is re-referenced, then the dummy channel is deleted again. In both options, the data rank remains intact. The final preprocessed dataset is then saved with the filename provided in *bemobil_config.preprocessed_filename*.

Independent component analysis

EEG measures not only brain signals, but always a mix of cortical, ocular, and muscular physiological sources, at times even cardiac activity in addition. Traditionally, any non-brain aspects of the data are considered artifacts and removed from the data, often in the time-domain. However, in MoBI data, physiological sources can carry important information and contribute to the interpretation of the data. Separating these sources and reconstructing their estimated activity throughout the experiment is thus an essential processing step, especially in mobile experiments where movement of the eyes and body is unrestricted. This can be done using blind source separation with ICA, which has become a staple in EEG analysis after having been introduced around three decades ago (Bell & Sejnowski, 1995; Hyvärinen & Oja, 1997). Different algorithms for ICA exist, but the Adaptive Mixture Independent Component Analysis (AMICA; Palmer et al., 2011) was shown to perform best in different comparisons, which is why we use it in this pipeline (Delorme et al., 2012; Leutheuser et al., 2013; Zakeri et al., 2014). The function *bemobil_process_all_AMICA* incorporates necessary steps from the preprocessed EEG dataset to the final dataset containing ICA information. The function stores the intermediate files of AMICA processing and accompanying plots in the folder specified in *bemobil_config.spatial_filters_folder* and its subfolder *bemobil_config.spatial_filters_folder_AMICA*. Exemplary visualizations of AMICA and subsequent processing can be seen in figure 4.

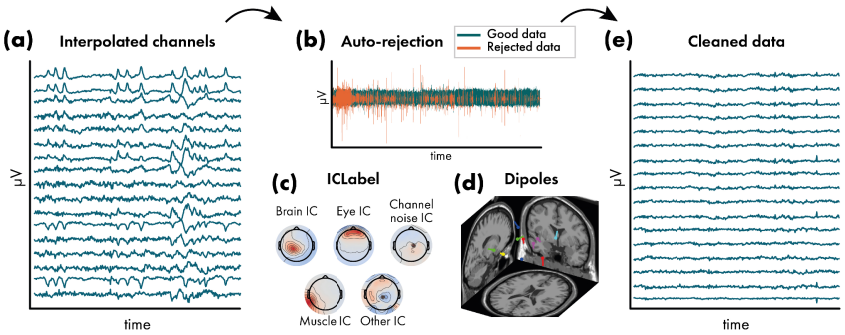


Figure 4: Exemplary impressions of the visualizations of the independent component analysis (ICA). During the computation of the ICA decomposition and subsequent source analysis, several plots are created to allow an inspection of the workflow. (a) The final preprocessed data is shown again here for visualization purposes (see Figure 3d). (b) An inspection plot of the automatic sample rejection of the AMICA algorithm is created (see section *Independent component analysis*), where rejected samples are denoted in red. (c) Exemplary visualization of the ICLabel classification of resultant independent components (ICs), see section *Original data reconstruction and independent component classification*. (d) Equivalent dipole models of the ICs that are classified as being of brain origin (subject 76 of the visual discrimination datasets that are available with the BeMoBIL Pipeline), see section *Equivalent dipole model reconstruction*. (e) Finally, the data is cleaned by removing unwanted ICs such as those stemming from non-brain sources, see section *Original data reconstruction and independent component classification*. The cleaned data is visualized in a plot similar to (a).

A vital step before computing ICA is to perform high-pass filtering with a suitable cutoff frequency. In earlier work, we showed that higher high-pass filters than commonly applied (> 0.5Hz cutoff) are often beneficial with diminishing returns, and that a higher cutoff frequency is recommended for data from mobile experiments and for higher channel density (Klug & Gramann, 2021). In addition, the improved decomposition also affects the signal-to-noise ratio of event-related potentials (ERPs) when cleaned with ICA, even when the ICA information was copied on an unfiltered dataset to make sure the effect was not due to the filter itself but only due to the improved ICA decomposition. Taken together, we found a filter of 1.25Hz cutoff to be a good overall option for data from both stationary and mobile experiments. However, for data containing strong movement, it might be suitable to use cutoff frequencies of 2Hz or even higher. It is important to note that the EEGLAB filter specifications do not use the cutoff frequency (the frequency where the filter starts taking noticeable effect, defined as -6db reduction in power) but instead, the user specifies the passband-edge (where the filter starts taking any effect). The transition bandwidth (two times the difference between passband edge and cutoff frequency) and resulting cutoff frequency are computed based on different heuristics depending on the frequency. It is recommended to specify the filter manually to ensure it leads to reproducible and comparable results (Widmann et al., 2015). We thus specify the filter order in the pipeline, and recommend using the same we used in our work comparing the high-pass filters (*bemobil_config.filter_AMICA_highPassOrder* parameter): The order of 1650 leads to a transition bandwidth of 0.5Hz, so to obtain a true cutoff frequency of 1.25Hz, a setting of 1.5Hz is required for the *bemobil_config.filter_lowCutoffFreqAMICA* parameter. A zero-phase Hamming window FIR filter is used. As with the high-pass filter for channel cleaning, this filter will only be applied to compute the AMICA and will not have any effect on the final dataset other than an improved decomposition. The final AMICA information will be copied to the unfiltered dataset.

Subsequently, the data is processed with AMICA. No time-domain cleaning is performed before doing so because AMICA includes a powerful cleaning option: At the beginning of the processing, samples with a log-likelihood estimation that is x standard deviations below the average (specified by *bemobil_config.AMICA_reject_sigma_threshold*, recommended 3 (Klug et al., 2022)) are removed repeatedly (as specified by *bemobil_config.AMICA_n_rej*, recommended 5 to 10 times (Klug et al., 2022)). This means that samples with suboptimal model fit are being rejected, as measured objectively by the algorithm itself. In our tests, this method yielded better or the same decomposition results than using our own time-domain cleaning option (see section *Time-domain cleaning based on epoch rejection* and supplementary material), which is why it is recommended and used in this pipeline. If desired, this feature can be disabled by setting the *bemobil_config.AMICA_autoreject* flag to 0. Note that these samples are not removed from the data itself, AMICA only disregards them internally when computing the spatial filters. A plot of the rejection is created and the rejection information is stored in the EEG data structure, thus being available for investigation or use in other circumstances. As it was shown, if stricter cleaning is being used (higher *bemobil_config.AMICA_n_rej* and/or lower *bemobil_config.AMICA_reject_sigma_threshold*), more samples will be discarded by AMICA (Klug et al., 2022). This can have a positive effect on the decomposition, which is important if subsequent source analysis is performed. However, in this case, the rejected samples will not contribute to the computation of the decomposition, which in turn means that the decomposition is not fit for these samples, and artifactual contributions might intersperse into brain components. Hence, it may be that even though the decomposition appears to be improved, the signal-to-noise ratio of the final measures (such as ERPs) is decreased or not changed in comparison to using more lax criteria when cleaning the data (Klug et al., 2022). We thus recommend using this time-domain cleaning conservatively: as much as necessary for the desired analysis, but as little as possible.

Equivalent dipole model reconstruction

To obtain estimates of the source location for the resultant independent components (ICs), the fitting of equivalent dipole models is done using the *DIPFIT* toolbox of EEGLAB with standard settings of a 3-layer boundary element model (an exemplary visualization can be seen in figure 4d). If a non-standard electrode layout with individual electrode locations is used, these locations need to be warped to the standard locations to enable the correct fit of dipoles. For this, a subset of electrodes where the corresponding closest standard electrode is known can be entered in *bemobil_config.warping_channel_names*. The final dipole model has an accuracy of 1-2cm for brain ICs (Acar & Makeig, 2010) and includes information about the component topography variance that is not explained by the physiological model (residual variance, RV). The RV value can serve as an estimate of the physiological plausibility of the component and its respective dipole, and inform decision-making about the removal of components at a later stage (Delorme et al., 2012). Note that as we are interested in the location of not only brain but also eye and muscle source locations, it would be desirable to use a head model that does not restrict the sources to brain tissue only. This could be done using HArtMuT, a new head volume conduction model that extends to the neck and includes brain sources as well as sources representing eyes and muscles that can be modeled as single dipoles, symmetrical dipoles, and tripoles (Harmening et al., 2022).

This model is currently under development and will be included in the pipeline as soon as it is available.

Original data reconstruction and independent component classification

The ICA processing thus far was performed on high-pass filtered data. Since final EEG measures may require the data to be filtered with a lower filter cutoff (e.g. ERPs), the computed AMICA information including rejections and dipole fitting is copied back to the initial preprocessed, but unfiltered, dataset (Hyvärinen et al., 2001; Klug & Gramann, 2021). In order to provide a directly usable dataset, a final zero-phase Hamming window FIR filter can be applied using the *bemobil_config.final_filter_lower_edge* (high-pass filter, recommended 0.2 Hz) and *bemobil_config.final_filter_higher_edge* (low-pass, recommended empty, i.e. not used). This filter will be applied to all sets in the *bemobil_config.single_subject_analysis_folder*. Using the recommended 0.2 Hz high-pass filter will remove slow drifts in the data, but leave all relevant information intact, even for ERP analysis. Information about the filter is stored in the EEG data structure and can easily be reported. The filter order can optionally be specified, otherwise, EEGLAB default filters are being used. If both a lower and higher edge for the filter are entered, the filters are being applied successively without the use of a band-pass filter, as suggested by (Widmann et al., 2015). This dataset will then be stored in the folder specified by *bemobil_config.single_subject_analysis_folder* with the name specified by *bemobil_config.preprocessed_and_ICA_filename*.

As a last step in the EEG processing, the data is cleaned with the ICLabel classifier (Pion-Tonachini et al., 2019) as specified in *bemobil_config.iclabel_classifier*. ICLabel classifies ICs into brain, eye, muscle, and heart sources as well as channel and line noise artifacts and a category of other, unclear, sources, based on a large database of expert labelings (an exemplary visualization can be seen in figure 4). Our experience with MoBI datasets is that the 'lite' classifier detects muscle components more reliably than the 'default' classifier (Klug & Gramann, 2021). If it is only important to detect brain ICs, the 'default' classifier is likely to be the better choice. Only classes that are specified in *bemobil_config.iclabel_classes* are kept in the data, all others are removed. The majority vote is used by default, meaning each component is assigned the class with the highest probability. It is possible to also set a different threshold using the *bemobil_config.iclabel_threshold* parameter. In that case, the summed probability of the classes specified in *bemobil_config.iclabel_classes* must be higher than this threshold to keep the component. The cleaned dataset is saved with the name specified in *bemobil_config.single_subject_cleaned_ICA_filename* alongside a plot of the kept IC topographies and dipoles. Note that this is a critical step in the cleaning process. We found that ICLabel often performs sufficiently well to justify using it, since it makes the ICA cleaning objective and reproducible, but there might be cases where it fails, especially in datasets containing many muscle sources. A reason for this is that the original datasets used to train the ICLabel classifier were taken mostly from stationary experiments, thus muscle components or components related to mechanical artifacts stemming from movement are underrepresented in the classifier. It is very important that the resulting plots are inspected and checked for misclassifications. Guidelines for this process can be found in (Chaumon et al., 2015), and especially for MoBI data, a training tool for ICA labeling can be found at

https://www.icmobi.org. On this website, experienced researchers can also label components in order to train a new classifier dedicated to MoBI data.

Motion data processing

As for EEG, the BeMoBIL Pipeline provides a fully automatic processing pipeline for motion data from the raw data to cleaned and filtered data including derivatives. If motion data is to be analyzed in sync with EEG data with non-experimental segments removed, the same segments need to be removed in the motion data, too, to maintain the synchronization and enable easy transfer of event markers or other features between the two modalities. Data of rigid body movement can be processed using the *bemobil_process_all_motion* function. This function takes a full set of motion data, containing one or more rigid body tracked points with six degrees of freedom (3D position and orientation values, orientation can be given in quaternion units or Euler angles), and creates a cleaned motion dataset with orientation in Euler angles, containing derivative channels for velocity and acceleration in addition. The function performs the following actions:

1. Split the complete motion data into individual sets for each tracked rigid body, where each of the rigid bodies undergoes processing steps 2.-8.
2. Clean the motion data, which includes removing excessively large jumps in the data as well as interpolation of samples with lost tracking (NaN samples). This process also extrapolates data to parts of the experiments containing no motion information (NaN entries inserted during import, see section *Data import and time synchronization*) by entering the nearest available value.
3. Unwrapping Euler angles to eliminate jumps between -pi and pi in radian values. This is necessary for low-pass filtering, as otherwise, the jumps will create ringing artifacts.
4. Low-pass filter the data with the filter frequency given in *bemobil_config.lowpass_motion.*
5. Wrap the angles to pi again.
6. Compute the first derivative (velocity), ignoring the jumps from -pi to pi.
7. As time derivatives effectively amplify the high frequencies, it is recommended to use another low-pass filter after each derivative, as can be set in *bemobil_config.lowpass_motion_after_derivative.*
8. Compute the second time-derivative (acceleration) and add another low-pass filter as in 7.
9. Merge all single rigid bodies into one complete motion dataset again.
10. The final processed motion set is then stored on disk in the folder specified in *bemobil_config.motion_analysis_folder*.

Event marker extraction

Experiments typically contain event markers to denote events within the experiment. This may be for example the beginning or end points of the experiment or experiment blocks, the presentation of stimuli, or responses of participants such as a button press. For stationary experiments in classical settings (seated participant, presentation of a stimulus on a computer monitor, no head movement), these event markers are sufficient for the

investigation of most experimental questions. However, mobile EEG or MoBI studies inherently contain movement that may be explicitly part of the relevant measures. Here, extracting event markers post-hoc from the collected data can play an important role, as it can reveal information about cognitive and motor processes, or other physiological states. Hence, the pipeline provides a set of functions that allow easy and fast creation of event markers from the most prominent MoBI data types: Motion (general motion as well as gait events), eye gaze (blink events), and heartbeats. If the multimodal data was loaded via the pipeline import pathways, they will be completely synchronized, allowing extracting event markers in one data stream and copying them into another for analysis of event-related activity. Additionally, it is possible to generate event markers from the EEG data alone, even in the absence of other data streams.

Motion event detection

Motion events can be an integral analysis aspect of MoBI experiments. For example, they can be used to detect reaction time and movement duration (Gehrke et al., 2022; Jungnickel & Gramann, 2016), they can serve as anchors for time-warping in spectral analysis (Gramann et al., 2021), they can help remove oscillatory gait artifacts (Gwin et al., 2010), or they can help to shed light on the neural basis of oscillatory gait generation (Wagner et al., 2016). To enable this functionality, the pipeline contains two functions: A basic movement onset and offset detector that requires only a single tracked element of any kind, and the advanced detection of relevant gait event markers.

Basic motion event detection

The *bemobil_detect_motion_startstops* algorithm detects motion starts and stops based on a coarse and a fine threshold of one or more given channels. The square root of the sum of squares of these channels is taken as the detection data, resulting in the absolute value (if a single channel was entered), or the absolute movement in more than one dimension (if more than one channel was entered). An overall movement is detected first based on a coarse threshold of a given quantile of the data (0.65 by default), then a fine threshold is applied based on a buffer (plus and minus two seconds by default) around the detected initial coarse movement onset. This fine threshold is the minimum within the buffer plus a proportion of the range of the data within the buffer (0.05 by default). From the detected coarse movement onset going back in time, the last sample exceeding the fine threshold is taken as the final movement start, and from this point going forward in time, the last sample exceeding the fine threshold is taken as the movement stop. In effect, the coarse movement quantile threshold can be regarded as related to the amount of time one expects the tracked object to be in motion overall, while the fine movement threshold describes the expected data variability during the rest phase before the movement onset. The detector assumes no trend in the data and thus works on data where the endpoint of a movement is the same as the start point in the relevant channels. This can be for example the yaw orientation of the head to detect rotation movements (Gramann et al., 2021), the position of the hand in a reaching task (Gehrke et al., 2022), or the up/down movement of a foot tracker to detect steps (see next section). The detected event markers and used parameters are stored in the data structure so they can be copied between synchronized datasets of different modalities.

Extracting gait parameters from motion data

Although computing final measures of specific movements is not in the scope of the BeMoBIL Pipeline, we would like to point out a few standard parameters that are commonly investigated in a number of mobile EEG and MoBI experiments. The aim of including this description is to enable researchers that are familiar with EEG, but not motion analysis, to obtain results comparable to biomechanics research. A large number of previous mobile EEG and MoBI studies investigated human brain dynamics in walking participants demonstrating the importance of gait in the research field. Here, one branch investigates the cortical dynamics accompanying gait control (e.g. Castermans et al., 2014; Gwin et al., 2010; Seeber et al., 2015; Wagner et al., 2012). A second branch of studies rather uses gait to investigate human brain dynamics in ecologically valid scenarios (e.g. De Sanctis et al., 2020; Jacobsen et al., 2021; Malcolm et al., 2015; Nenna et al., 2021; Protzak et al., 2021; Protzak & Gramann, 2021; Reiser et al., 2019, 2021). Here, the focus is less on gait itself but rather on the extraction of gait parameters in different walking scenarios to clean the signal from gait-related artifacts or to investigate the impact of walking difficulty on brain dynamics, respectively. Both approaches, however, require reliable extraction of gait parameters to inform the EEG analyses. To facilitate the extraction of reliable parameters and to provide guidelines for future mobile brain imaging studies, the definitions of a number of biomechanical gait parameters are given here.

As the main categorization, gait parameters could be divided into parameters of pace (including gait velocity or walking speed and referring step length), rhythm (including cadence, stride, and swing time), phase (including double support time), base of support (including step width) as well as variability (including the coefficients of variation for all parameters; Hollman et al., 2011). From a biomechanical point of view, an active heel-to-toe movement with associated ankle movement is necessary to maintain balance during the forward motion of walking. At the same time, with each step, the body's center of gravity shifts beyond the support surface, so that the pelvis must be stabilized in the period from heel strike to double stance phase (Perry & Burnfield, 2010). Therefore, the movement of the feet, as well as the pelvis and hip rotation, give insights into stable gait patterns. Nevertheless, gait is a whole-body movement, and therefore trunk rotation, head movements, and arm swing and their referring kinematics are oftentimes relevant aspects to consider, e.g. to detect pathological gait patterns.

To describe the susceptibility to disturbances of gait and its accompanying brain dynamics, a set of parameters has been established quantifying changes in various gait parameters. These include i) reduced stride length, defined as the distance that one part of the foot travels in front of the same part of the other foot during each step, we recommend using the distance from heel-strike to heel-strike (Hollman et al., 2011; Scott et al., 2015), ii) reduced walking speed (Verghese et al., 2009), iii) prolonged stance phases, e.g. expressed by the double support time (the time when both feet are in contact with the ground simultaneously), defined as the sum of the time elapsed during two periods of double support in the gait cycle (Hollman et al., 2011; Maki, 1997; Scott et al., 2015; Verghese et al., 2009), and iv) increased stride length variability, defined as the coefficient of variation (%CV), calculated as the average standard deviation of the gait parameter divided by the average mean (Hollman et al., 2011; Verghese et al., 2009).

An essential requirement for measuring and calculating spatiotemporal gait parameters (i.e step length and cadence) is the accurate spatiotemporal identification of heel-strike and toe-off events (Rudisch et al., 2021). We provide options to extract these event markers, but no additional gait parameters, as these can be derived from the data and event markers but may require knowledge about the measurement device or experimental paradigm that is impossible to anticipate in a generalized pipeline. Due to underlying differences in measurement devices and principles, it has to be ensured that the gait cycles can be accurately detected by using a standardized description of the axes in the Euclidean space. In biomechanical analyses, the gait parameters are commonly defined such that the x-axis describes the anterior-posterior direction, the y-axis describes the medial-lateral direction, and the z-axis refers to the vertical direction (distal and proximal, up and down). With these axes known, gait event markers can be extracted in a standardized way using the *bemobil_gait_analysis* function.

The BeMoBIL Pipeline extracts gait cycles defined as a sequence of events: 1) movement start (flat foot phase end), 2) toe-off, 3) heel-strike, and 4) movement stop (next flat foot phase start). For an ideal toe-off and heel-strike event detection, the heels and toes would require their own tracker. However, these events can be reasonably approximated by assuming that i) during the flat-foot phase the foot is moving backward in relation to the body, ii) as soon as the toe-off event occurs, the foot starts moving forward, and iii) as soon as a heel-strike event occurs, the foot stands still again, and is moving backward in relation to the body. Detecting these four events thus is possible with only one tracker on top of the feet and happens in two steps: First, foot movements, in general, are detected using the *bemobil_detect_motion_startstops* function on the z-Axis of the tracking (up-down movement). These mark the flat-foot end and the flat-foot start events, respectively. In a second step, toe-off and heel-strike events are then defined using the velocity in the x-Axis (forward-backward). This requires the foot movement measurement to be in relation to the body, i.e. not a continuous forward movement but a forward-and-back cycle. If the motion was measured on a treadmill, this is already the case (as the feet slide back under the body). In overground walking, a motion tracking of the torso or head of the participant is required in addition. If such tracking is provided, the values in the x- and y-axes are subtracted from those of the feet, such that the feet exhibit a forward-and-back cycle again, relative to the provided tracking. With this cyclic movement in the x- and y-direction, one additional issue has to be overcome: The tracking axes are not necessarily always aligned with the movement axes, but the movement x-axis is relevant for event extraction. Hence, a PCA analysis is computed using the provided x- and y-axes for each foot separately, and finally, the component with the higher variance is taken as the foot x-axis. In this oscillatory forward-backward movement of the feet, the zero-crossings of the first derivative (i.e. maxima/minima) are taken as the final two events: Such a zero-crossing after the foot movement start event is used as the toe-off event, and the same before the foot movement end event is used as the heel-strike. The final gait event markers are added to the EEGLAB data structure and can be copied to synchronized other datasets such as EEG, allowing further investigations. An example visualization of this gait event extraction can be seen in figure 5.

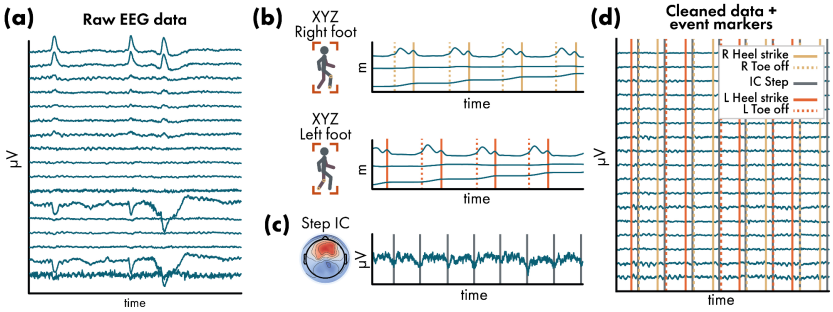


Figure 5: Exemplary impressions of the motion event extraction. (a) Example visualization of a 5-second data segment of raw data from subject 64 of the visual discrimination datasets that are available with the BeMoBIL Pipeline. (b) Motion data with gait event extraction (here, only toe-off and heel-strike events are plotted for visualization purposes). The x-axis of the feet is at the bottom of the two plots, respectively, and shows the forward movement. The y-axis represents the left-right movement of the foot, which is minimal in this case. The z-axis at the top of the two plots shows the lifting of the feet, with a lowering during the swing phase and a prolonged minimum during the flat foot phase. For details on the extraction of these events, see section *Extracting gait parameters from motion data*. (c) Exemplary detection of steps of both feet based on the independent component time series, see section *IC-based gait-related event detection*. (d) Exemplary final cleaned EEG data similar to (a) that contains all gait events for further analysis. Since the data modalities are synchronized during import (see section *Data import and time synchronization*), the events can easily be transferred between them.

Eye tracking-based blink event detection

Eye gaze data can be very informative for MoBI analysis, and especially using VR displays it is easy to be recorded. During blinks, the pupil is not detectable, which means that the eye tracker can not determine the gaze direction. For a basic eye gaze analysis, the *bemobil_clean_eye* function thus cleans eye gaze data by detecting blinks based on the pupil radius. It then interpolates all eye gaze data using *pchip* interpolation during the detected blink times. Eyeblink event markers are added to the event field of the dataset and blink extraction information is stored.

The blink detection is modified from code written by Ravi Chacko (Mitz et al., 2017) and processes the data as follows:

1. The pupil radius is used to determine blinks, as during closure of the eyelids the pupil radius is zero. To this end, the mean and SD of the radius throughout the recording are computed, disregarding samples below a radius of 0.2mm. The general threshold to define a blink is then defined in SDs below the mean, 3 by default.
2. Going through the data from the start, a blink is detected if a sample is below the threshold. The following first sample above the threshold is regarded as the end of the blink.
3. These coarse start and end values are now used to determine the exact starts and ends. Here, a search buffer around the detected timestamps is used (20ms by default). Within this period, large peaks in the absolute values of the first derivative of the pupil size are computed and the last peak denotes the true end of the blink. This allows for a brief period in which the pupil size can jitter after re-opening the eyes (e.g. several samples in which the pupil radius appears to be open, then closed again, then open again), but only the final opening is taken as the end of the blink.

4. A buffer around the detected start and end points of the blink is applied (30ms by default), the blink indices are stored, and the detector continues after the end index with 2.
5. After searching for blink indices in the entire dataset, the detected blink periods are interpolated using *pchip* interpolation in all eye tracking data. This also removes large jumps in the eye gaze position data that can occur during blink periods, when the pupil is not trackable.
6. Finally, blinks that were below a minimum duration (100ms by default) are discarded, and the final blink start and end event markers are added to the EEGLAB data structure. These can be copied to synchronized other datasets such as EEG, allowing further investigations. An example visualization of this blink event extraction can be seen in figure 6.

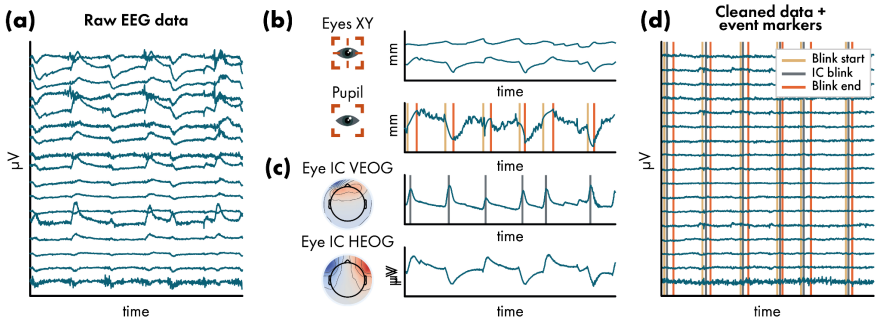


Figure 6: Exemplary impressions of the blink event extraction. (a) Example visualization of a 5-second data segment of raw data from subject 6 in the NeSitA example data that is available with the BeMoBIL Pipeline. (b) Eye tracking data with blink event extraction. Blinks are detected based on pupil size (see section *Eye tracking-based blink event detection*). The eye x- and y-axes are not used for event extraction but are shown for the visualization of the synchrony of eye tracking and independent component (IC) time series in (c). (c) Exemplary detection of blinks based on the IC time series, see section *IC-based blink and saccade event detection*. (d) Exemplary final cleaned EEG data similar to (a) that contains all blink events for further analysis. Since the data modalities are synchronized during import (see section *Data import and time synchronization*), the events can easily be transferred between them.

Event detection based on independent component time series

Paradigms including mobile EEG/MoBI can be comparably complex and time-consuming so it can be reasonable to reduce the data recording to EEG only. Even though this limits the extraction of event markers based on other modalities such as motion or eye gaze, there is another possible option to extract event markers for further analysis utilizing the decomposition of the data using ICA. Although ICA is commonly applied to remove non-brain activity from the data, it can also be used to extract event markers from components that stem from eyes, muscles, or mechanical artifacts like cable sway or electrode pressure from walking. Hence, these non-brain sources can now inform the analysis of brain activity by providing context events. Thus what is traditionally considered an artifact can become a signal and thus an integral part of the data analysis.

IC-based blink and saccade event detection

Following the approach published in Wunderlich & Gramann (2021), the BeMoBIL Pipeline provides two functions to extract blink, saccade, and step event markers from IC time series. The function *bemobil_detect_blinks_from_ICA* detects blinks and saccades based on the time series of two ICs representing horizontal and vertical eye movements, respectively (e.g., figure 6c). In the first step, these components are automatically detected based on their topographies and their spectral power below 5Hz. It is also possible to indicate specific ICs manually and hand them over as parameters when using the function. Blinks and saccades are detected using the *findpeaks* function. Default values for distance, prominence, and width of the peaks are provided based on EOG literature (Lins et al., 1993). Alternatively, the parameters can be set by the user. Before detecting peaks in the IC activation time course, the data is smoothed using a moving median filter, which preserves the steep edges while removing high frequency fluctuations (Bulling et al., 2011). The moving median filter *smoothdata* is used with a window length of 0.08s by default and can be defined by the user. Furthermore, *bemobil_detect_blinks_from_ICA* takes care of flipped IC activity and ensures that blinks are always positive peaks which is a requirement for the use of the *findpeaks* function. For saccade detection, the square root of the sum of vertical and horizontal eye movement is computed which is known in the electrooculogram (EOG) literature as EOG activity (Jia & Tyler, 2019). The squared derivative of the EOG time series allows for using peak detection to locate the quickest differences in the EOG time series equalling saccadic eye movements. To disentangle the blinks from saccades, all saccade detections in temporal proximity (by default +/- 100ms) of a blink are excluded. For all the remaining detected peak latencies, EEGLAB event markers are created using the type ‘blink’ or ‘saccade’, respectively.

An informed decision about the detector efficacy can be made by the provided figures. Here, one plot shows the activation of the two detected eye ICs and the newly created event markers. In addition, there are figures for blinks and saccades, respectively, depicting the whole dataset with the *findpeaks* parameters, allowing the inspection of the peak detection efficacy when zooming in. Below this plot are histograms of the prominences and peak widths (including those exceeding the thresholds). These histograms provide information about how well the used threshold fits this participant's data. In our tests and comparisons with eye tracking data in an experiment containing strong eye movements, we found the IC-based blink detector to be in correspondence with the eye tracking-based detector (see section *Eye tracking-based blink event detection*) around 80% of the time. This, however, depends on the nature of the eye movement, as e.g. strong vertical movements can appear almost like an eye blink in the vertical IC activity even in the absence of a blink. Saccades can thus falsely be detected as blinks, mask blinks, or go undetected because they can happen during a blink. Taken together, the event extraction should always be handled with care and especially saccades are not always reliable, which is why we offer an option to not extract saccade event markers. All parameters, the detected blink and saccade event latencies, as well as the prominences and widths of the detector, are stored in the EEG data structure. An example visualization of this IC-based blink detection can be seen in figure 6.

IC-based gait-related event detection

The function *bemobil_detect_steps_from_ICA* detects steps based on the time series of an indicated gait IC (e.g., figure 5c). The most indicative signature of a gait IC is that the time

series follows the same pattern as the upwards axis of a motion tracker device mounted to the head. As some data might contain steps only in parts of the entire duration, it is possible to specify the start and end points of the step detection to prevent false alarms during periods where the participant did not walk. Analogous to the eye-movement detector, the data is smoothed using the moving median and the function checks whether the stronger deflection is plotted upward. Steps are detected using *findpeaks* with parameters for distance, prominence, and minimal and maximal duration of the peaks. Finally, EEGLAB event markers are created at the respective latencies and stored in the EEG data. The parameters can be chosen freely and an informed decision about the detector efficacy can be made by the plots of the detection, including a histogram of the prominences and widths (including those exceeding the thresholds) akin to that of the blink detector. The detection can be repeated with varying search boundaries or IC indices. All parameters, the found step event markers, and latencies, as well as the prominences and widths of the detector, are stored in the EEG data structure. An example visualization of this IC-based step detection can be seen in figure 5.

Heartbeat event detection

Heartbeats and subsequent analyses such as heart rate variability can be of interest in a MoBI experiment, for example for assessments of workload and stress (Delliaux et al., 2019; Kim et al., 2018). We thus provide the widely used Pan-Tompkins algorithm to detect heartbeats from electrocardiography (ECG) data (Pan & Tompkins, 1985), either recorded from additional ECG sensors or derived from independent components reflecting cardiac activity. As the original algorithm is written in the C programming language, we use a modified version of a MATLAB implementation available online (Sedghamiz, 2014), additionally allowing the specification of the high and low-pass filter cutoff frequencies (1 Hz and 40 Hz by default, respectively). The MATLAB implementation by Sedghamiz (2014) makes use of the hard-coded frequency-dependent filters and computations defined in the original work if the data is given at a sampling rate of 200 Hz, but uses state-of-the-art MATLAB signal processing otherwise, which is our recommendation.

Single-subject and group-level post-processing

The BeMoBIL Pipeline focuses on the automatic processing and cleaning of EEG and other data but provides a selection of useful additional features regarding the next steps. With preprocessed EEG data as well as event markers available, this often is to analyze the data based on epochs that are centered around one or several events of interest. Using these epochs, it is then possible to perform analyses on either the cleaned sensor data or on the source-level, taking into account the location of the EEG equivalent dipole models of the ICs. To facilitate these steps, two options provided by the pipeline are helpful: First, we offer a way to reject epochs based on objective criteria and in a balanced fashion between conditions, and second, the pipeline includes a repeated IC clustering approach for reliable and reproducible group-level source analysis.

Epoch rejection and time-domain cleaning

After creating epochs either based on experiment event markers or based on the event markers created using the pipeline, these epochs might still contain non-brain signals even if the data was cleaned with ICA before. As a final option to improve the signal strength of the measure of interest, it is thus often necessary to reject epochs that are particularly noisy. This can be achieved by either manually selecting epochs to reject, or by using automated methods. One issue arising in many automated rejection tools, however, is that one cannot specify the amount of data to be removed, but only the threshold that leads to removal. Hence, one runs the risk of insufficient cleaning or the removal of an excessive amount of data when the threshold is not adjusted properly. More importantly, cleaning data from different movement conditions might lead to an imbalance in the removal of epochs, where significantly more epochs are rejected in the condition with more movement, potentially complicating the analysis or skewing the final results.

Automatic and balanced epoch rejection

We thus propose a method to rank epochs on their noise level and remove only a specified amount of the worst ranking epochs. Each epoch is evaluated using four measures that are normalized by their median across epochs: i) the mean of channel means, to catch epochs with high amplitude, ii) the SD of means, to catch epochs with inhomogeneous channel activity according to their mean, iii) the mean of SDs, to catch epochs with high variance within channels (e.g. strong leftover muscle activity), and iv) the SD of SDs, to catch epochs with inhomogeneous channel activity according to their variance. It is possible to weigh the measures separately, although the default of equal weights is recommended. Each epoch then receives a final summed score and the epochs are sorted according to that score. Then three options are available to determine the rejection threshold: i) a fixed number of epochs that should be left - this will guarantee an equal number of epochs for all conditions, ii) a fixed percentage threshold, e.g. the worst 10% of the epochs are removed - this will preserve the original ratio of epochs per condition, or iii) determine a “knee-point” of the score and use that as the threshold - this will lead to the removal of only outlier epochs. Downsides of the third method are that this can lead to an imbalance of the retained epochs between conditions, and in cases where very few very strong outliers exist, the “knee-point” can be shifted to a high threshold, while very clean datasets can exhibit an almost round curve with a “knee-point” that is shifted towards the center. Thus, we recommend using methods i) or ii).

Time-domain cleaning based on epoch rejection

The algorithm to reject epochs can be extended for use as general time-domain cleaning of continuous data. To this end, the data is first high-pass filtered and subsequently cut into epochs that are then cleaned as described above. If eye movements are to be ignored in this cleaning, it is recommended to use a high-pass filter of 10 Hz to remove the majority of eye contributions. This, however, is unnecessary if the cleaning is used on data where eye contributions were removed with ICA. To target bad segments more precisely, epochs can be specified to overlap such that a short burst of noise could be captured by one epoch rather than two adjacent ones. Additionally, when an epoch is marked for rejection, a buffer around the epoch is rejected as well to capture possible on- and offsets of the artifact. The

epoch length, overlap, and buffer can be specified according to the needs of the analysis. For example, when removing artifacts before running ICA, an epoch length of 500 ms with an overlap of 125 ms and an epoch buffer of 62.5 ms can be used, while longer epochs could be useful if it is important to retain longer contiguous data. We do not recommend using this method in our pipeline because we found no improvement when using it over the automated cleaning of AMICA itself.

Robust group-level source analysis for regions of interest

If source analysis is to be performed on the group level, it is necessary to find ICs of all participants that represent activity from the specific source region. To this end, k-means clustering can be used, which finds similar components in the complete study set containing ICs from all participants based on weighted measures such as dipole location, scalp topographies, spectrum, ERPs, or event-related spectral perturbations (ERSPs). Choosing the weights is subject to the analyst, but it is recommended to weigh the location highly, add topographies and spectra, and, depending on the situation, ERPs and ERSPs with lower weighting. However, when using ERPs or ERSPs, it can be argued that double-dipping happens in the selection of relevant ICs (meaning that the measure that is later used to compute statistics is also used to select the ICs). A counter-argument to this would be that the clustering uses average measures while the statistics are used to investigate condition differences. All in all, no final rule on how to choose the weights can be given. The standard k-means clustering, however, has one other strong limitation: the k-means results are not stable due to variation in the starting conditions. Repeating the clustering can result in different solutions, and depending on the location and the similarity of the ICs, the cluster of interest (COI, the cluster closest to your region of interest, ROI) can contain vastly different ICs (see figure 7a).

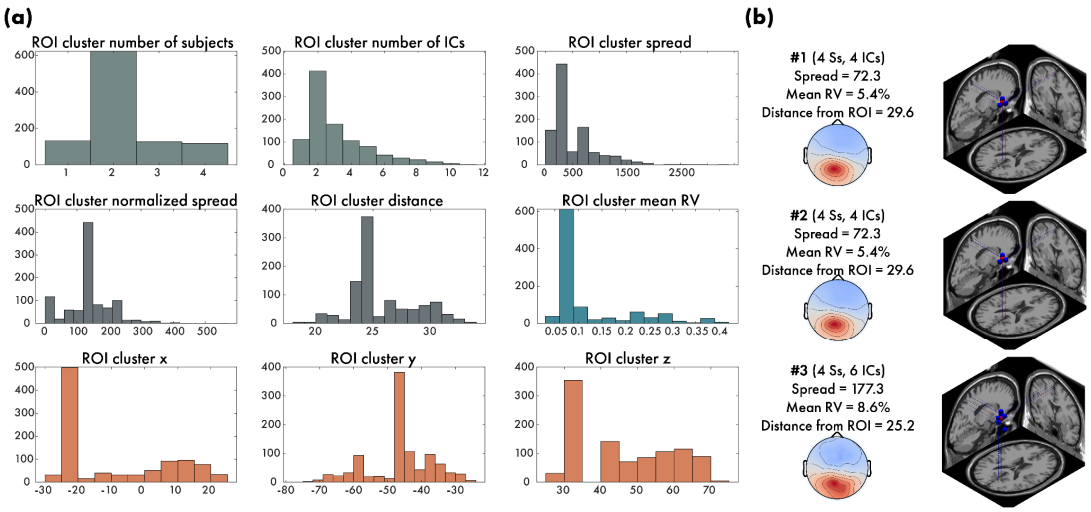


Figure 7. **Robust group-level clustering example.** In this example, we ran the robust clustering of independent components (ICs) on the group level of the four example participants from the visual discrimination study that is available with the BeMoBIL Pipeline. We weighted the dipoles with 3, the topographies with 1, the spectra with 1, and the ERPs of stimulus presentation with 1, and ran 1000 repetitions of the clustering. The region of interest (ROI) was set to the posterior parietal cortex (MNI coordinates of [0,-48,39]) and the weights for the ROI cluster quality measures were

chosen as: subjects = 3, ICs/subjects = -1, normalized spread = -1, mean RV = -1, distance from ROI = -2, and mahalanobis distance from median of multivariate distribution = -1. See section *Robust group-level source analysis for regions of interest* for details. (a) The resultant distributions of the various quality measures show the variability of the outcome of the clustering. The unnormalized spread and the X/Y/Z coordinates are not used for the final selection but shown for visualization purposes only. (b) The resulting best three clusters and their respective measures show that, although they are not identical, they are similar, and the top 2 clusters are identical, indicating a stable result of the repeated clustering.

To alleviate this issue, we implemented a repeated clustering approach that clusters several hundred or thousand times and selects the COI for each clustering solution, based on the distance from a given ROI in MNI coordinates (Evans et al., 1993). For each of these COIs, a set of quality measures is derived: the number of subjects in that cluster, the average number of ICs per subject, the normalized Euclidean spread of the cluster (distance of the individual IC locations from the cluster centroid divided by the number of ICs), the mean RV of the ICs in the cluster, the distance of the cluster centroid from the ROI, and the Mahalanobis distance from the median of the multivariate distribution of all cluster solutions. The last measure shows how normal, or representative for the entire distribution, the given cluster solution is. Ideally, we are looking for a solution that contains as many subjects as possible (so the final measures are representative of the group), few ICs per subject (because it is difficult to interpret several ICs per subject in an identical cortical area), a small distance from the ROI, a low spread (tight cluster around the ROI restricting it to one “functional” cortical area), a low mean RV (reflecting physiologically plausible ICs), and a low Mahalanobis distance from the median (attenuating outlier clusters). To this end, the quality measures are assigned a weight and the clustering solutions are sorted according to their summed score. The solution with the highest combined score is then taken as the final clustering solution that can be used for further analysis.

To make sure that no outlier solution is taken as the final solution, on the one hand, one can weigh the Mahalanobis distance more negatively, on the other hand, we provide plots of the locations and average scalp topographies of the five highest-ranking solutions (figure 7b). These should look very similar, which indicates that the results are stable. Depending on the ROI it might be possible to achieve a stable solution with only 100 repetitions (e.g. in the visual cortex), but deeper ROIs like the retrosplenial complex may require several thousand repetitions. If one is interested not only in one ROI but several, two options are possible: i) Optimize separately for all ROIs and create different STUDY files accordingly. A limitation of this approach is that the same IC may be present in two or more ROIs if they are too close together. ii) If one ROI is more important than the others, it might be better to only optimize for that one ROI and use this cluster solution for all subsequent analyses (Gramann et al., 2021).

Miscellaneous functions

As a final element, the pipeline comes with a set of post-processing functionalities not pertaining to event-related EEG analysis: Different spatial filtering techniques are implemented that do not rely on blind source separation but instead make use of the additional data modalities or other information, and scripts to visualize motion and eye gaze data in an intuitive way are provided.

Additional spatial filtering approaches

While blind source separation methods such as ICA can be beneficial in disentangling the mix of electrical sources in EEG data in general, other spatial filtering methods that make use of additional information may prove to be more powerful in circumstances where knowledge or expectations about the data are already available before the analysis. We thus provide several such options in the function *bemobil_signal_decomposition_extended*: Source Power Comodulation (SPoC; Dähne, Meinecke, et al., 2014) allows the separation of data subspaces that describe the modulation of a given target value and can, for example, be used to extract motion-related information from EEG when motion data is available, allowing either the removal or the interpretation of the data (Gehrke et al., 2019). Canonical Correlation Analysis (CCA) can be used to find common subspaces between EEG and other data, allowing the investigation of their relationships such as the interplay of EEG and functional magnetic resonance imaging (fMRI) data (Biebmann et al., 2010), or the removal of motion artifacts in EEG data (Safieddine et al., 2012). Lastly, Spatiospectral Decomposition (SSD; Nikulin et al., 2011) is a possible preprocessing step to reduce dimensionality before applying SPoC or CCA (Dähne, Nikulin, et al., 2014) but can also be used standalone to extract spatial filters that enhance specific frequencies such as the theta or alpha band in the EEG. As a final element when using the above described spatial filtering techniques, the function *bemobil_distributed_source_localization* allows the inspection of the computed spatial patterns on the source level by using previously found source locations of the ICA, a method that was originally intended to visualize the sources of brain-computer interface classifiers (Krol et al., 2018; Zander et al., 2016).

Visualizations of motion and eye gaze data

Motion datasets can in particular be difficult to visualize without neglecting parameters that could lead to serendipitous discoveries. We have thus focused on developing multiple informative plots for both eye-tracking and motion datasets that ensure an accessible, coherent, and rapid inspection of the data. As both eye-tracking and motion data are best examined using several parameters in a single plot, we have developed plots that readily and intuitively visualize velocities and positions in space over time (figure 8).

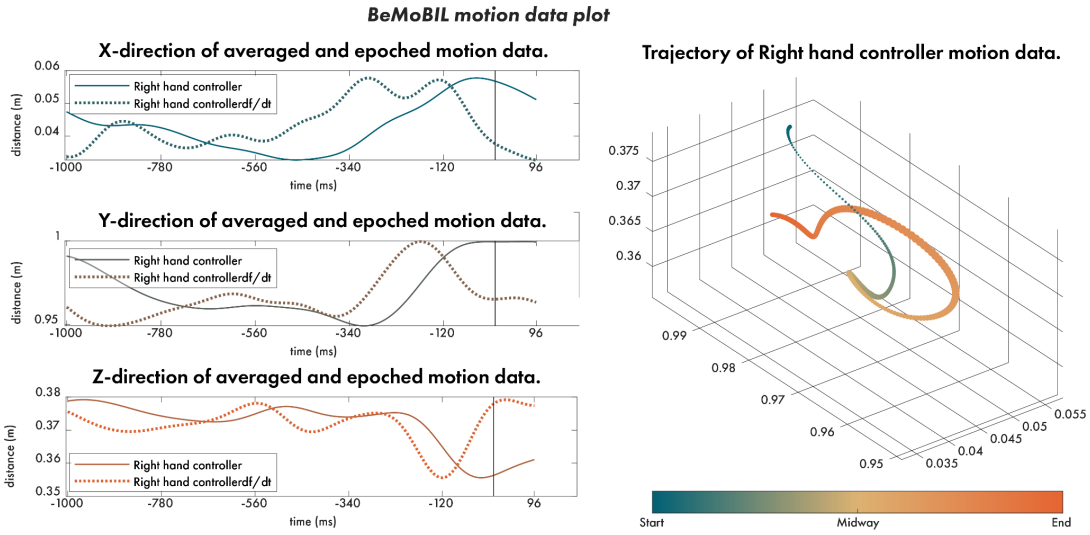


Figure 8. **BeMoBIL motion data plot example.** In this example, we visualize the controller movement in the time up to a touch of an object in a physical reaching task (data from subject 1 in the NeSitA example data that is available with the BeMoBIL Pipeline). On the left side of the plot, we visualize the synchronized XYZ coordinates separately with their velocity measures plotted on top. On the right side of the plot, we visualize the three-dimensional movement as a trajectory through space. Besides illustrating the position in space, we also show the temporal dimension using colors and the velocity of the movement using the thickness of the line as inspired by the kinematics of handwriting, supporting an intuitive and accessible reading. All the information used in the plots is available as the function output upon plotting the figure.

The pipeline provides three such functions for plotting. The first, *bemobil_plot_motion,* serves to inspect each XYZ coordinate as well as their three-dimensional trajectory. On the left side of the plot, the function divides the XYZ coordinates of the motion data into three separate plots that visualize the distance against the time as well as the velocity of the movement. The right side of the plot visualizes the three-dimensional trajectory in space without neglecting velocity or time. To visualize velocity, we have been inspired by the everyday kinematics of handwriting, i.e. strokes that are thicker relative to the rest of the line represent a slower hand movement as opposed to the thinner part of the line, representing much faster movements. To visualize the temporal dimension, we have been inspired by the techniques in plots of imaginary numbers and Riemann topology, i.e. we use a gradient of color to depict the end and beginning of the trajectory. All the information used in the plots is available as the function output upon plotting the figure.

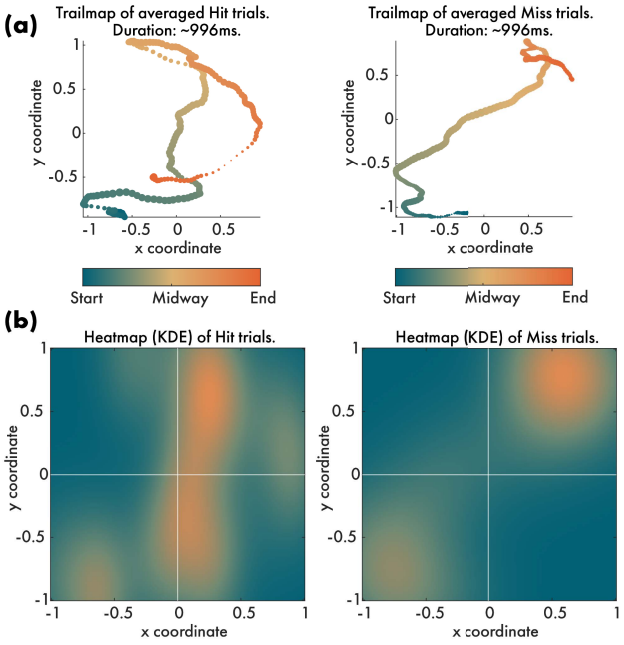


Figure 9. Eye tracking data plot example. (a) In this example, we compare two conditions in which the participant either hit or missed the target in a physical reaching task (data from subject 1 in the NeSitA example data that is available with the BeMoBIL Pipeline). The plot is based on the same principles as the motion data plot (see figure 8). (b) The data can readily be transferred to a heatmap kernel density estimation (KDE) plot that can be exported and further processed using image-based analyses.

The last two functions are dedicated to eye-tracking. Following the same principles as above, the function *bemobil_plot_trail* compares the trails of two conditions in two dimensions, using again the thickness of the trail to represent the velocity and the colors for the temporal dimension (figure 9a). As an additional plot that enables image-based analyses, we have included a heat-map function, *bemobile_plot_heatmap*, that plots the areas in which the trail spent the longest time (figure 9b). For custom colormaps, we offer the *bemobil_makecmap* which generates a gradient between given colors. The output here can be used with all our plot functions.

Summary

The combination of electrophysiological data and other body measures such as motion or eye tracking becomes more prevalent as a tool in neuroscientific studies investigating human brain dynamics in more ecologically valid scenarios such as the workplace (Ayaz & Dehais, 2018; Mehta & Parasuraman, 2013; Parasuraman & Rizzo, 2007; Wascher et al., 2014) or urban environments (Aspinall et al., 2015; Djebbara et al., 2019). These large multimodal datasets require special treatment during the analysis to ensure the reliability of the results. While there are powerful solutions available for individual aspects of EEG data preprocessing (e.g. the PREP pipeline (Bigdely-Shamlo et al., 2015)) or automated EEG analysis (e.g. the HAPPE pipeline (Gabard-Durnam et al., 2018)), they do not cater to the specific needs of multimodal MoBI data. For example, these pipelines do not include the synchronized import of multiple data streams, the processing of data not stemming from

EEG, the treatment of EEG data from mobile experiments specifically, or additional functionalities for extracting event markers, reliable source analysis, or early-fusion analysis that can greatly benefit MoBI research.

Our proposed BeMoBIL Pipeline thus seeks to fill this gap and provides automatic, transparent, replicable, and easy-to-use data processing for multimodal datasets of human participants. It is a MATLAB pipeline based on EEGLAB (Delorme et al., 2011; Delorme & Makeig, 2004) and fieldtrip (Oostenveld et al., 2011), comprising wrappers and new functions to i) import the raw multimodal dataset to obtain BIDS-compatible shareable data and fully synchronized multimodal EEGLAB files that allow easy computations of multimodal analysis or transfer of event markers, ii) preprocess and clean EEG data, including line-noise removal, channel rejection and interpolation, and artifact rejection using ICA and epoch cleaning, iii) process motion data and extract event markers from motion, eye tracking, ECG, and IC activity, iv) robust clustering of ICs for group-level source analysis, v) allow early-fusion analysis of EEG and other data combined, and vi) visualize motion and eye gaze data intuitively. This set of features helps to reduce experimenter bias when analyzing data, allows the easy replication of data processing, and lowers the threshold of entry into EEG data analysis for researchers from other fields. Importantly, while the pipeline is designed around the requirements of MoBI studies, it can be applied to datasets from stationary studies with minimal changes.

The pipeline can be used with a small set of wrapper functions and a configuration file, but the steps can also be called individually to allow a modular setup of self-defined pipelines. For reporting the methods, all steps are documented within the EEG data structure itself, allowing precise descriptions of all data processing such as exact filter specifications, removed channels, or the amount of data removed by AMICA, to simplify replications or similar investigations in future studies. Furthermore, once set up, the entire processing from the raw files to the cleaned datasets and even the final study with clusters of interest can easily be reproduced using the pipeline scripts. This will yield identical final datasets and results though minor deviations due to suboptimal parameter selection in the configuration files (e.g., number of iterations for specific processing steps) are possible. The pipeline is documented in depth in a wiki section on the public GitHub repository, available online at https://github.com/BeMoBIL/bemobil-pipeline/wiki. Here, a comprehensive user guide on all steps including the installation can be found alongside explanations of the chosen default values and other practical considerations when running the functions.

Limitations and conclusions

A major limitation of largely automated data processing is that researchers have no insight into the actual data and its subtleties. This can lead to overlooking processing errors, such as mishandling artifacts or unrealistic results. Therefore, we strongly believe that the visualization and close inspection of the data is an essential element of electrophysiological data analysis. To this end, analytics plots of the data during and after processing steps, as well as figures for the processing itself are created throughout the pipeline in order to keep the experimenters informed about the analysis. A comprehensive guide on the interpretation of the analytics plots is available on the wiki. Additionally, Even though the pipeline is as flexible as possible and has been extensively tested, it is in the nature of MoBI experiments that they sometimes raise unprecedented analysis issues that are difficult to predict and may

require a customized pipeline. This problem can be addressed by taking modular elements of the pipeline and setting up custom analyses or adding entirely new elements as appropriate. Since the BeMoBIL Pipeline is an open source project, new analysis approaches from the community as well as contributions to the code repository are strongly encouraged. Our aim is to continuously provide the best possible pipeline that takes into account algorithmic advances or new insights while providing a stable and reliable analysis that can be easily used.

Taken together, our two main goals were to make the processing and analysis of (mobile) EEG data more reliable and independent of the researcher and to open up EEG research to other fields as a method to answer their own research questions. We provide a flexible and powerful open-source toolset for multimodal data processing and cleaning that paves the way for fast and reliable research using mobile EEG and complementary body measures.

Acknowledgments

We gratefully acknowledge all users and testers of earlier versions of the pipeline in recent years. They provided invaluable feedback that allowed the pipeline to grow to its current state. This work has been supported by the DFG (grant number GR2627/8-1) and USAF (grant number ONR 10024807).

Data and Code Availability Statement

The source code of the BeMoBIL Pipeline is available at https://github.com/BeMoBIL/bemobil-pipeline. The data and scripts that were used in the examples are available for download as described in the repository.

Ethics Statement

The experimental procedures of the example data were approved by the local ethics committee (Technische Universität Berlin, Germany), the research was performed in accordance with the ethics guidelines of the Declaration of Helsinki, and all participants signed a written informed consent prior to participation.

Conflict of Interest

The authors declare no conflict of interest.

References

Acar, Z. A., & Makeig, S. (2010). Neuroelectromagnetic Forward Head Modeling Toolbox.

Journal of Neuroscience Methods, *190*(2), 258–270.

Aspinall, P., Mavros, P., Coyne, R., & Roe, J. (2015). The urban brain: analysing outdoor

physical activity with mobile EEG. *British Journal of Sports Medicine*, *49*(4), 272–276.

Ayaz, H., & Dehais, F. (Eds.). (2018). *Neuroergonomics - The Brain at Work and in Everyday Life* (1st ed., pp. 59–63). Elsevier Academic Press.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159.

Biebmann, F., Meinecke, F. C., Gretton, A., Rauch, A., Rainer, G., Logothetis, N. K., & Müller, K. R. (2010). Temporal kernel CCA and its application in multimodal neuronal data analysis. *Machine Learning*, *79*(1-2), 5–27.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, *9*, 16.

Büchel, D., Sandbakk, Ø., & Baumeister, J. (2021). Exploring intensity-dependent modulations in EEG resting-state network efficiency induced by exercise. *European Journal of Applied Physiology*, *121*(9), 2423–2435.

Bulling, A., Ward, J. A., Gellersen, H., & Tröster, G. (2011). Eye movement analysis for activity recognition using electrooculography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(4), 741–753.

Castermans, T., Duvinage, M., Cheron, G., & Dutoit, T. (2014). About the cortical origin of the low-delta and high-gamma rhythms observed in EEG signals during treadmill walking. *Neuroscience Letters*, *561*, 166–170.

Chaumon, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, *250*, 47–63.

Cohen, M. X. (2017). Rigor and replication in time-frequency analyses of cognitive electrophysiology data. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *111*, 80–87.

da Cruz, J. R., Chicherov, V., Herzog, M. H., & Figueiredo, P. (2018). An automatic pre-processing pipeline for EEG analysis (APP) based on robust statistics. *Clinical*

Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, *129*(7), 1427–1437.

Dähne, S., Meinecke, F. C., Haufe, S., Höhne, J., Tangermann, M., Müller, K. R., & Nikulin, V. V. (2014). SPoC: A novel framework for relating the amplitude of neuronal oscillations to behaviorally relevant parameters. *NeuroImage*, *86*, 111–122.

Dähne, S., Nikulin, V. V., Ramírez, D., Schreier, P. J., Müller, K. R., & Haufe, S. (2014). Finding brain oscillations with power dependencies in neuroimaging data. *NeuroImage*, *96*, 334–348.

de Cheveigné, A. (2020). ZapLine: A simple and effective method to remove power line artifacts. *NeuroImage*, *207*, 116356.

Delliaux, S., Delaforge, A., Deharo, J.-C., & Chaumet, G. (2019). Mental Workload Alters Heart Rate Variability, Lowering Non-linear Dynamics. *Frontiers in Physiology*, *10*, 565.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Delorme, A., Mullen, T., Kothe, C., Akalin Acar, Z., Bigdely-Shamlo, N., Vankov, A., & Makeig, S. (2011). EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. *Computational Intelligence and Neuroscience*, *2011*, 130714.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PloS One*, *7*(2), e30135.

De Sanctis, P., Malcolm, B. R., Mabie, P. C., Francisco, A. A., Mowrey, W. B., Joshi, S., Molholm, S., & Foxe, J. J. (2020). Mobile Brain/Body Imaging of cognitive-motor impairment in multiple sclerosis: Deriving EEG-based neuro-markers during a dual-task walking study. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *131*(5), 1119–1128.

Djebbara, Z., Fich, L. B., Petrini, L., & Gramann, K. (2019). Sensory-motor brain dynamics reflect architectural affordances. *Proceedings of the National Academy of Sciences*, 1–31.

Evans, A. C., Collins, D. L., & Mills, S. R. (1993). 3D statistical neuroanatomical models from 305 MRI volumes. *1993 IEEE*. https://ieeexplore.ieee.org/abstract/document/373602/

Gabard-Durnam, L. J., Leal, A. S. M., Wilkinson, C. L., & Levin, A. R. (2018). The harvard automated processing pipeline for electroencephalography (HAPPE): Standardized processing software for developmental and high-artifact data. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00097

Gehrke, L., Guerdan, L., & Gramann, K. (2019). Extracting Motion-Related Subspaces from EEG in Mobile Brain/Body Imaging Studies using Source Power Comodulation. *International IEEE/EMBS Conference on Neural Engineering, NER*, *2019-March*, 344–347.

Gehrke, L., Lopes, P., Klug, M., Akman, S., & Gramann, K. (2022). Neural sources of prediction errors detect unrealistic VR interactions. *Journal of Neural Engineering*, *19*(3), 036002.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 1–9.

Gramann, K., Ferris, D. P., Gwin, J., & Makeig, S. (2014). Imaging natural cognition in action. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *91*(1), 22–29.

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T. P., Lin, C. T., Liao, L. D., & Makeig, S. (2011). Cognition in action: Imaging brain/body dynamics in mobile humans. *Reviews in the Neurosciences*, *22*(6), 593–608.

Gramann, K., Hohlefeld, F. U., Gehrke, L., & Klug, M. (2021). Human cortical dynamics during full-body heading changes. *Scientific Reports*, *11*(1), 18186.

Gwin, J. T., Gramann, K., Makeig, S., & Ferris, D. P. (2010). Removal of movement artifact from high-density EEG recorded during walking and running. *Journal of*

Neurophysiology, *103*(6), 3526–3534.

Harmening, N., Klug, M., Gramann, K., & Miklody, D. (2022). HArtMuT - Modeling eye and muscle contributors in neuroelectric imaging. In *bioRxiv* (p. 2022.08.19.504507). https://doi.org/10.1101/2022.08.19.504507

Hollman, J. H., McDade, E. M., & Petersen, R. C. (2011). Normative spatiotemporal gait parameters in older adults. *Gait & Posture*, *34*(1), 111–118.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent Component Analysis*. John Wiley & Sons.

Hyvärinen, A., & Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural Computation*, *9*, 1483–1492.

Jacobsen, N. S. J., Blum, S., Witt, K., & Debener, S. (2021). A walk in the park? Characterizing gait-related artifacts in mobile EEG recordings. *The European Journal of Neuroscience*, *54*(12), 8421–8440.

Jia, Y., & Tyler, C. W. (2019). Measurement of saccadic eye movements by electrooculography for simultaneous EEG recording. *Behavior Research Methods*, *51*(5), 2139–2151.

Jungnickel, E., Gehrke, L., Klug, M., & Gramann, K. (2019). Chapter 10 - MoBI—Mobile Brain/Body Imaging. In H. Ayaz & F. Dehais (Eds.), *Neuroergonomics* (pp. 59–63). Academic Press.

Jungnickel, E., & Gramann, K. (2016). Mobile Brain/Body Imaging (MoBI) of Physical Interaction with Dynamically Moving Objects. *Frontiers in Human Neuroscience*, *10*(June), 306.

Kappenman, E. S., & Keil, A. (2017). Introduction to the special issue on recentering science: Replication, robustness, and reproducibility in psychophysiology. *Psychophysiology*, *54*(1), 3–5.

Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., & Koo, B.-H. (2018). Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investigation*, *15*(3), 235–245.

Klug, M., Berg, T., & Gramann, K. (2022). No need for extensive artifact rejection for ICA - A multi-study evaluation on stationary and mobile EEG datasets. In *bioRxiv* (p. 2022.09.13.507772). https://doi.org/10.1101/2022.09.13.507772

Klug, M., & Gramann, K. (2021). Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *The European Journal of Neuroscience*, *54*(12), 8406–8420.

Klug, M., & Kloosterman, N. A. (2022). Zapline-plus: A Zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG. *Human Brain Mapping*, *43*(9), 2743–2758.

Krol, L., Mousavi, M., De Sa, V., & Zander, T. (2018). Towards Classifier Visualisation in 3D Source Space. *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 71–76.

Larson, M. J., & Moser, J. S. (2017). Rigor and replication: Toward improved best practices in human electrophysiology research. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *111*, 1–4.

Leutheuser, H., Gabsteiger, F., Hebenstreit, F., Reis, P., Lochmann, M., & Eskofier, B. (2013). Comparison of the AMICA and the InfoMax algorithm for the reduction of electromyogenic artifacts in EEG data. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 6804–6807.

Lins, O. G., Picton, T. W., Berg, P., & Scherg, M. (1993). Ocular artifacts in EEG and event-related potentials. I: Scalp topography. *Brain Topography*, *6*(1), 51–63.

Makeig, S., Gramann, K., Jung, T.-P., Sejnowski, T. J., & Poizner, H. (2009). Linking brain, mind and behavior. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *73*(2), 95–100.

Maki, B. E. (1997). Gait changes in older adults: predictors of falls or indicators of fear. *Journal of the American Geriatrics Society*, *45*(3), 313–320.

Malcolm, B. R., Foxe, J. J., Butler, J. S., & De Sanctis, P. (2015). The aging brain shows less

flexible reallocation of cognitive resources during dual-task walking: A mobile brain/body imaging (MoBI) study. *NeuroImage*, *117*, 230–242.

Mehta, R. K., & Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Frontiers in Human Neuroscience*, *7*(December), 1–10.

Mitz, A. R., Chacko, R. V., Putnam, P. T., Rudebeck, P. H., & Murray, E. A. (2017). Using pupil size and heart rate to infer affective states during behavioral neurophysiology and neuropsychology experiments. *Journal of Neuroscience Methods*, *279*. https://doi.org/10.1016/j.jneumeth.2017.01.004

Miyakoshi, M., Schmitt, L. M., Erickson, C. A., Sweeney, J. A., & Pedapati, E. V. (2021). Can We Push the “Quasi-Perfect Artifact Rejection” Even Closer to Perfection? *Frontiers in Neuroinformatics*, *14*(January), 1–5.

Nenna, F., Do, C. T., Protzak, J., & Gramann, K. (2021). Alteration of brain dynamics during dual-task overground walking. *The European Journal of Neuroscience*, *54*(12), 8158–8174.

Nikulin, V. V., Nolte, G., & Curio, G. (2011). A novel method for reliable and fast extraction of neuronal EEG / MEG oscillations on the basis of spatio-spectral decomposition. *NeuroImage*, *55*(4), 1528–1535.

Ojeda, A., Bigdely-Shamlo, N., & Makeig, S. (2014). MoBILAB: an open source toolbox for analysis and visualization of mobile brain/body imaging data. *Frontiers in Human Neuroscience*, *8*(March), 121.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Palmer, J. A., Kreutz-delgado, K., & Makeig, S. (2011). *AMICA : An Adaptive Mixture of Independent Component Analyzers with Shared Components*. 1–15.

Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Transactions*

on Bio-Medical Engineering, *32*(3), 230–236.

Parasuraman, R., & Rizzo, M. (Eds.). (2007). *Neuroergonomics - The brain at work*. Oxford University Press.

Pedroni, A., Bahreini, A., & Langer, N. (2019). Automagic: Standardized preprocessing of big EEG data. *NeuroImage*, *200*(December 2018), 460–473.

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, *6*(1), 1–5.

Pernet, C. R., Martinez-Cancino, R., Truong, D., Makeig, S., & Delorme, A. (2021). From BIDS-Formatted EEG Data to Sensor-Space Group Results: A Fully Reproducible Workflow With EEGLAB and LIMO EEG. *Frontiers in Neuroscience*, *14*(January), 1–7.

Perry, J., & Burnfield, J. (2010). *Gait analysis. Normal and pathological function* (2nd ed.). SLACK Inc.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*(May), 181–197.

Protzak, J., & Gramann, K. (2018). Investigating established EEG parameter during real-world driving. *Frontiers in Psychology*, *9*(NOV), 1–11.

Protzak, J., & Gramann, K. (2021). EEG beta-modulations reflect age-specific motor resource allocation during dual-task walking. *Scientific Reports*, *11*(1), 16110.

Protzak, J., Wiczorek, R., & Gramann, K. (2021). Peripheral visual perception during natural overground dual-task walking in older and younger adults. *Neurobiology of Aging*, *98*, 146–159.

Reiser, J. E., Wascher, E., & Arnau, S. (2019). Recording mobile EEG in an outdoor environment reveals cognitive-motor interference dependent on movement complexity. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-49503-4

Reiser, J. E., Wascher, E., Rinkenauer, G., & Arnau, S. (2021). Cognitive-motor interference in the wild: Assessing the effects of movement complexity on task switching using

mobile EEG. *The European Journal of Neuroscience*, *54*(12), 8175–8195.

Richer, N., Downey, R. J., Hairston, W. D., Ferris, D. P., & Nordin, A. D. (2020). Motion and Muscle Artifact Removal Validation Using an Electrical Head Phantom, Robotic Motion Platform, and Dual Layer Mobile EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, *28*(8), 1825–1835.

Robbins, K. A., Touryan, J., Mullen, T., Kothe, C., & Bigdely-Shamlo, N. (2020). How Sensitive Are EEG Results to Preprocessing Methods: A Benchmarking Study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society*, *28*(5), 1081–1090.

Rodrigues, J., Weiß, M., Hewig, J., & Allen, J. J. B. (2021). EPOS: EEG Processing Open-Source Scripts. *Frontiers in Neuroscience*, *15*, 660449.

Rudisch, J., Jöllenbeck, T., Vogt, L., Cordes, T., Klotzbier, T. J., Vogel, O., & Wollesen, B. (2021). Agreement and consistency of five different clinical gait analysis systems in the assessment of spatiotemporal gait parameters. *Gait & Posture*, *85*, 55–64.

Safieddine, D., Kachenoura, A., Albera, L., Birot, G., Karfoul, A., Pasnicu, A., Biraben, A., Wendling, F., Senhadji, L., & Merlet, I. (2012). Removal of muscle artifact from EEG data: Comparison between stochastic (ICA and CCA) and deterministic (EMD and wavelet-based) approaches. *EURASIP Journal on Advances in Signal Processing*, *2012*(1). https://doi.org/10.1186/1687-6180-2012-127

Scott, D., McLaughlin, P., Nicholson, G. C., Ebeling, P. R., Stuart, A. L., Kay, D., & Sanders, K. M. (2015). Changes in gait performance over several years are associated with recurrent falls status in community-dwelling older women at high risk of fracture. *Age and Ageing*, *44*(2), 287–293.

Sedghamiz, H. (2014). *Matlab Implementation of Pan Tompkins ECG QRS detector*. ResearchGate. https://www.researchgate.net/publication/313673153_Matlab_Implementation_of_Pan_Tompkins_ECG_QRS_detector

Seeber, M., Scherer, R., Wagner, J., Solis-Escalante, T., & Müller-Putz, G. R. (2015). High and low gamma EEG oscillations in central sensorimotor areas are conversely modulated during the human gait cycle. *NeuroImage*, *112*, 318–326.

Short, M. R., Damiano, D. L., Kim, Y., & Bulea, T. C. (2020). Children With Unilateral Cerebral Palsy Utilize More Cortical Resources for Similar Motor Output During Treadmill Gait. *Frontiers in Human Neuroscience*, *14*, 36.

Verghese, J., Holtzer, R., Lipton, R. B., & Wang, C. (2009). Quantitative gait markers and incident fall risk in older adults. *The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences*, *64*(8), 896–901.

Wagner, J., Makeig, S., Gola, M., Neuper, C., & Muller-Putz, G. (2016). Distinct Band Oscillatory Networks Subserving Motor and Cognitive Control during Gait Adaptation. *Journal of Neuroscience*, *36*(7), 2212–2226.

Wagner, J., Solis-Escalante, T., Grieshofer, P., Neuper, C., Müller-Putz, G., & Scherer, R. (2012). Level of participation in robotic-assisted treadmill walking modulates midline sensorimotor EEG rhythms in able-bodied subjects. *NeuroImage*, *63*(3), 1203–1211.

Wascher, E., Heppner, H., & Hoffmann, S. (2014). Towards the measurement of event-related EEG activity in real-life working environments. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*, *91*(1), 3–9.

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *Journal of Neuroscience Methods*, *250*, 34–46.

Wunderlich, A., & Gramann, K. (2021). Eye movement-related brain potentials during assisted navigation in real-world environments. *The European Journal of Neuroscience*, *54*(12), 8336–8354.

Zakeri, Z., Assecondi, S., Bagshaw, A. P., & Arvanitis, T. N. (2014). Influence of Signal Preprocessing on ICA-Based EEG Decomposition. *IFMBE Proceedings*, *41*, 563–566.

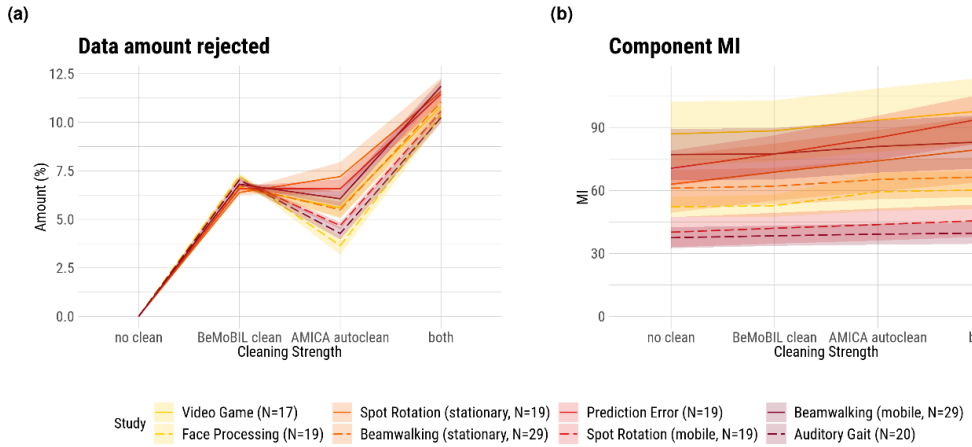
Zander, T. O., Krol, L. R., Birbaumer, N. P., & Gramann, K. (2016). Neuroadaptive technology enables implicit cursor control based on medial prefrontal cortex activity.

Proceedings of the National Academy of Sciences, *113*(52), 14898–14903.

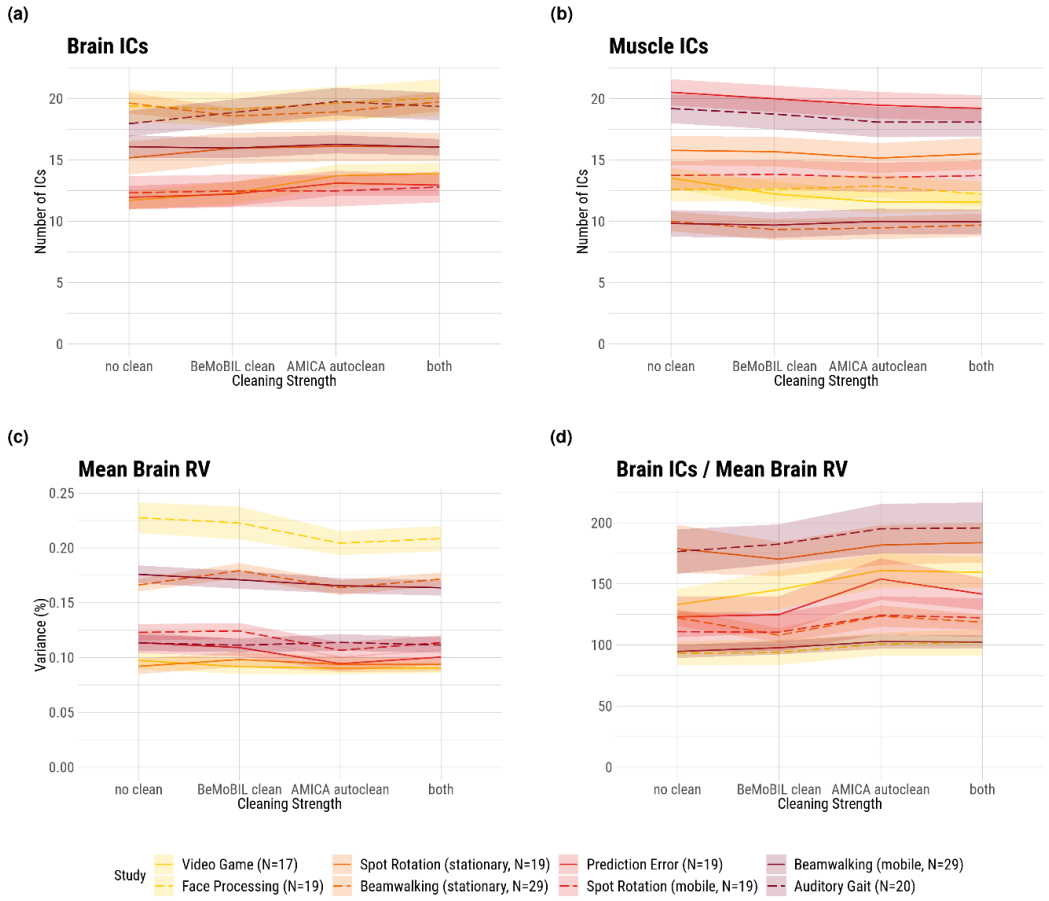
Supplementary material

Comparison of the BeMoBIL time-cleaning with the automatic AMICA sample rejection

We decomposed 8 studies with AMICA and investigated the decomposition quality. We compared no time cleaning with the two different methods and another using both methods in succession. The data and processing are otherwise identical to that in Klug et al. (2022). For the BeMoBIL time cleaning, we used a threshold of 5%, resulting in around 6% of data removed due to added buffering. For the AMICA autoclean, we used 5 iterations and 3 standard deviations as parameters.



1: Results for the data rejection amount and component mutual information (MI). Shaded areas depict the standard error of the mean (SE). "No clean" refers to no sample rejection being applied when computing AMICA. The colors denote the movement intensities: yellow - low, orange - low-to-medium, red - medium-to-high, violet - high.



2: Results for the ICLabel classification and residual variance (RV). Shaded areas depict the standard error of the mean (SE). "No clean" refers to no sample rejection being applied when computing AMICA. The colors denote the movement intensities: yellow - low, orange - low-to-medium, red - medium-to-high, violet - high.

The results indicate no relevant difference between the two decomposition methods. On average, a similar amount of data was rejected, and the change in brain ICs and RV was mostly within the SE range. The small trend towards a better decomposition in the automatic AMICA sample rejection resulted in this method being chosen for the BeMoBIL Pipeline.

References

Klug, M., Berg, T., & Gramann, K. (2022). No need for extensive artifact rejection for ICA - A multi-study evaluation on stationary and mobile EEG datasets. In *bioRxiv* (p. 2022.09.13.507772). https://doi.org/10.1101/2022.09.13.507772

Discussion

"It does not matter that we will never reach our ultimate goal. The effort yields its own rewards."—This quote by Lt. Cmdr. Data was introduced in the beginning of this dissertation because it carries meaning pertaining to the work presented here in several ways. The fist aspect points to the philsophical background of MoBI and is hidden in where this quote is embedded: It is the answer Lt. Cmdr. Data—a robot—gives when being asked why he tries to emulate humans, knowing he can't ever fully succeed. In a way this is both similar and the inverse to what MoBI tries to achieve: We already are human, conscious beings that experience emotions and perceive the world. However, we do not understand how exactly we work, and in our quest to increase this understanding, we, too, will never fully succeed. Yet, the effort we make yields its own rewards. As laid out in the introduction, it is this quest that is at the core of MoBI research, and although this dissertation is concerned with extremely zoomed in details on this path, the philosophical struggle is the motivation behind all efforts made here.

The second aspect is more pragmatic: No matter how much effort we put into reaching our ultimate goal of a 'perfect' data analysis, it can never be achieved. Analyzing neural time series data and linking it to functional interpretations about cognition is a task so rich and complex (Cohen, 2014) that the amount of variables that need to be investigated to obtain the 'perfect' analysis pipeline vastly exceeds the number of available researchers and their time, and a pipeline universally accepted as 'perfect ´ might not even be possible in principle. Adding mobility and new data modalities from MoBI experiments to this undertaking further complicates those already existing challenges, and introduces entirely new ones: the reduced signal-to-noise ratio of mobile EEG studies caused by non-brain contributions to the data (e.g. Debener et al., 2012; Jungnickel & Gramann, 2016), the emphasis and dependence on decomposing the data with ICA (already pointed out in the defining paper by Makeig et al., 2009), and the synchronized analysis of additional body data streams (e.g. Gramann et al., 2021). Yet, notwithstanding the futility of the attempt to obtain the 'perfect' pipeline that solves all challenges, the insights we gain when walking down that path will move the field forward and allow us to better address the complexities of MoBI data analysis. Understanding and improving the methods for brain and body imaging in and of itself thus already rewards us with new insights that, albeit not the focus of MoBI research in general, advance the field. In the remainder of this chapter, it will first be discussed, how the combined presented tools and studies contribute to the advancement of the four aspects of MoBI data analysis in light of the previously identified challenges, before diving deeper into the topic of automation in EEG data analysis, and finally giving an outlook on future directions of MoBI methods in the conclusions.

The first new tool in this regard is Zapline-plus, introduced in chapter 1 (Klug & Kloosterman, 2022). It addresses aspect iii) of MoBI data analysis, the preprocessing of EEG data, by removing frequency-specific artifacts without hampering downstream analysis. This includes the fully automatic removal of not only line noise but all other strong spectral peaks, while maintaining a high level of flexibility through a variety of parameters, and delivering comprehensive information via diagnostic plots. The algorithm was shown to substantially reduce noise levels such that the noise/surroundings power ratio (which was in the range of 2 to 7 in the raw EEG data and even several hundreds in the raw MEG data) approached 1 in the cleaned data, meaning the noise was removed without introducing a notch in the spectrum. It was also shown that all features were necessary for the successful cleaning of MoBI data (EEG study II of the

performed tests). Notably, the splitting of the data into chunks, in combination with the individual peak frequency detection, addressed the issue of non-stationary noise, which is a potentially problematic limitation of MoBI experiments. In addition, it was shown that MoBI datasets can exhibit a variety of unusual spectral peaks that are unlikely to be of physiological origin, but can lie within the range of interest for functional analyses, which emphasizes the importance of the fine automated detection of such peaks. Although we can only speculate on the origin of the high level of contamination in EEG study II, it is not uncommon in MoBI studies[18] and also EEG study I (which was recorded by another group and also included a mobile condition) exhibited an unanticipated spectral peak at a low frequency of around 7 Hz. Importantly, we found a corollary benefit of the search for individual peaks per chunk: This feature allows to clean recordings where artifacts are completely absent in parts of the data without risking serious negative effects from overcleaning, because in these chunks Zapline (de Cheveigné, 2020) is only applied minimally. In summary, the created tool is highly effective in removing frequency-specific artifacts in MoBI data, has a minimal effect on the non-artifactual data, and can be integrated into any kind of analysis pipeline, including those that are entirely automatic, thus laying an important foundation for the BeMoBIL Pipeline introduced in chapter 3.

In chapter 2, two studies investigate the important task of obtaining an optimal ICA decomposition, thus further addressing aspect iii) of MoBI data analysis. As laid out in the introduction, the use of ICA is a major aspect of MoBI studies because it allows the separation of brain and non-brain contributions in the EEG data. Yet, the effects of commonly applied measures to improve it—such as using high-density EEG electrode montages, using high-pass filters before computing ICA, and cleaning the data in the time-domain—are not well investigated. Examining specifically the impact of channel density and high-pass filtering in a mobile and a stationary condition of the same dataset, the first study (Klug & Gramann, 2021) made important observations: First, the study confirmed that ICA is a powerful tool also for mobile EEG experiments despite the increased level of non-brain signals. Removing non-brain sources from the data generally resulted in an increase of the exemplary signal-to-noise ratio (SNR) of sensor-level ERPs. This was the case in the stationary, but more strongly so in the mobile condition, suggesting that this step might even be considered a vital element of data cleaning in mobile EEG studies. The study also found a positive effect of channel density on SNR for montages up to 64 channels, from which on the effect leveled off, and on the number and quality of brain ICs for up to 128 channels. Although this result was not a surprise, this finding is important in order to substantiate the use of high-density EEG in MoBI studies especially when analyses are to be performed on the source level. A subsidiary finding regarding channel count was that even low-density montages could achieve a substantial improvement in the SNR on the sensor level. This was especially the case if a more dorsal layout was chosen (see the supplementary material of this study), which supports the use of such montages if source-level analysis is not a concern and the cost, in terms of both money and time, of high-density setups is too high to warrant their use. Another main finding of the study was that high-pass filtering does improve the ICA decomposition as anticipated, but this effect is dependent on the channel density and movement. Here, a higher-than-usual[19] high-pass filter cutoff is required to reach an optimal decomposition with both

[18] Also other, unpublished, analyses of datasets from the same laboratory indicated a similar contamination.

[19] E.g. in a twitter survey with 121 answers, the majority of voters used a high-pass filter cutoff <0.5 Hz, with exactly 0.5 Hz, and 0.5–1 Hz closely following. Only 15% of the voters selected the >1 Hz option. See

increased density and mobility. While a cutoff frequency of 1 Hz can be considered a common ground for most studies, high-density MoBI studies may require a cutoff up to 2 Hz, corresponding with previous studies (Dimigen, 2020; Winkler, Debener, Müller, & Tangermann, 2015). As an incidental effect, this study also principally validated the use of the ICLabel 'lite' classifier for ICA components (Pion-Tonachini et al., 2019) even in MoBI studies, as it was chosen for the automatic analysis of the data, and selecting only ICs classified as brain by this classifier proved to be beneficial for the SNR on the sensor level.

The second study in chapter 2 followed up on these findings and extended the generalizability by making use of additional open-access datasets with varying degrees of mobility. Using a fixed channel montage of around 60 channels, and a high-pass filter cutoff frequency of 1 Hz, this study varied the intensity of movement and time-domain cleaning. To this end, the AMICA algorithm (e.g. Palmer, Kreutz-Delgado, & Makeig, 2011) was used, as it includes an estimation of the model fit for each sample and can automatically remove samples below a data-driven threshold for which the strictness can be set in the function. As a first result, the study found an ambiguous effect of movement intensity on the ICA decomposition. Here, confirming the result of the first investigation, comparisons within two studies each containing two conditions with different intensity of movement yielded the expected result of a negative impact of movement on the decomposition. This effect, however, was not found across different studies, indicating that the effect of the difference between laboratory setups and experimental paradigms (Melnik et al., 2017) exceeds that of movement. An important consequence of this result is that mobility is not to be equated with negative impact on the data. For example, although participants who walk slowly through a city are more mobile than seated participants performing an arm reaching task, data recorded from the latter experiment might be more affected by sudden movements and especially by neck muscle activity or small shifts of the EEG cap, resulting in a worse ICA decomposition. The second investigated parameter was the application of the data-driven sample rejection with varying strength. Contradicting the expectations, no strong effects were found here, indicating that the AMICA algorithm is more robust to artifacts than initially assumed. Although a small trend was found towards an improvement of a moderate cleaning strength (removing around 4–8% of the data), the effect was weaker than anticipated and, more importantly, it did not vary systematically between movement intensities. This means that more movement did not require more cleaning when decomposing the data with ICA. This can be explained by the fact that physiological sources such as muscles and eyes are active in general and throughout the recording. Removing parts of this data hence is not sufficient to prevent the AMICA from taking them into account for the unmixing matrix, and thus has little effect on the final model. Summarizing the two studies of chapter 2 in terms of their insights on best practices before computing ICA in MoBI, it can be concluded that: i) Higher movement intensity requires more channels for source analysis, ideally 128, but ICA can already be applied with fewer channels, even as low as 16, for the purpose of data cleaning. ii) A high-pass filter cutoff of 1.25 Hz before running ICA can be regarded as a good overall selection for data from MoBI experiments, while being unproblematic for stationary experiments. iii) A moderate amount of time-domain cleaning using the automatic AMICA sample rejection is likely to be optimal for data from both mobile and stationary settings.

With the insights gained in these studies and the lack of a comprehensive MoBI

https://twitter.com/marius_s_klug/status/1314614106926833664

analysis pipeline in mind, the creation of the BeMoBIL Pipeline, presented in chapter 3, was the consequent next step to address the challenges of MoBI data analysis identified before. This pipeline integrates Zapline-plus, as well as the decomposition of EEG data with ICA. It is not limited to these aspects, however, as it comprises all necessary steps from the raw recording (as an XDF file containing time-stamped data streams) to a synchronized and cleaned multimodal dataset with events extracted for contextual analysis, and visualizations of motion and eye gaze data. In doing so, the pipeline presents itself as a solution in particular to aspects i) to iii) of MoBI data analysis identified in the introduction (synchronization, body data analysis, and EEG data preprocessing). In addition to that, it provides a solution to the robust clustering of ICs that is necessary for aspect iv), and it integrates late-fusion algorithms such as SPoC to allow a functional interpretation of the combined data.

Similar to MOBILAB, the BeMoBIL Pipeline addresses aspect i) by synchronizing the data using interpolation of other data streams to match the timestamps of the EEG, but it improves this synchronization in several ways: First, instead of directly importing the data into EEGLAB, it creates files in the BIDS data structure (Gorgolewski et al., 2016), which facilitates the standardized sharing of neural imaging data including EEG data (Pernet et al., 2019) and is in the process of being extended to include motion and other data with timestamps[20]. Additionally, when importing the data from BIDS to EEGLAB, the pipeline allows resampling using interpolation as well as filter-based methods, enhancing the functionality for data with highly precise measurements. The last improvement is the fact that the pipeline generates completely synchronized final merged datasets of all modalities, including those which were not recorded in all conditions (such as motion in a seated condition). To this end, the data is filled with not-a-number (NaN) values or cut where necessary, so the final data of all conditions have the exact same amount of samples. This is highly important for the easy extraction of events and their application to other modalities, such as the extraction of body events to be applied in EEG analysis.

With the obtained synchronized dataset, the second important feature of the BeMoBIL Pipeline is aspect ii), the processing of body data stemming from motion capture and eye tracking. For motion, this includes the interpolation of occluded samples, low-pass filters, and the calculation of derivatives, and for eye gaze, this includes the interpolation of blink periods. The pipeline also allows the extraction of motion start and stop events, detailed gait events, and blink events. This is done within the EEGLAB environment, meaning it can be applied at any time after the data import and events can be freely interchanged between the modalities, allowing the parallel processing of EEG and body data instead of relying on a fixed order as it is the case in MOBILAB. In addition to extracting events from body data modalities, the pipeline also allows for the extraction of events from the EEG data using 'artifactual' components of the ICA decomposition, such as eye blinks or gait ICs, thereby effectively re-evaluating these ICs as being signals of body imaging instead of artifacts in brain imaging (Wunderlich & Gramann, 2021).

The core of the BeMoBIL Pipeline, however, is the fully automatic, reproducible, transparent, and easy-to-use preprocessing of EEG data, addressing aspect iii) of MoBI data processing. This includes all steps from the raw imported data up to the final data that is prepared for the computation of single-subject measures: The removal of frequency-specific noise (using Zapline-plus), the robust detection and interpolation of bad channels (using a repeated automatic detection approach based on the *clean raw*

[20] See "BEP029" at https://bids.neuroimaging.io/get_involved.html

data plugin of EEGLAB), ICA decomposition (using AMICA with bad sample rejection as investigated in chapter 2), and a dipole model estimation (using the *dipfit* plugin of EEGLAB). Hidden in this process are numerous details, including, but not limited to i) the use of double data precision as suggested by Bigdely-Shamlo et al. (2015), ii) the automated generation of ur-data to allow the reconstruction of events, iii) the use of a full rank average reference (based on the identically named EEGLAB plugin) or reconstruction of the original reference EEG channel, iv) side-loops for steps that require a high-pass filter that might be inappropriate for downstream analysis (Widmann et al., 2015), such as the detection of bad channels and the application of AMICA, with their results being applied to the original dataset, and v) the separation of band-pass filters into successive high-pass and low-pass filters, which allows the separate specification of transition bandwidths. Once the data is cleaned and epochs for, for example, ERP or time-frequency analysis are extracted, the pipeline also offers an option to automatically reject bad epochs based on a fixed percentage threshold, thus enabling the balanced removal of epochs from different conditions.

In addition to these single-subject level processing tools, the pipeline also includes an important feature for robust group-level source imaging, on the border of aspects iii) and iv) of MoBI data processing: The repeated clustering of ICs and selection of an optimal solution for a given region of interest to counteract the need for post-hoc cluster cleaning. It optimizes the clusters directly according to the chosen weights of different quality measures, and it provides stable clustering results which otherwise would need to be examined, for example, by post-hoc permutation tests (Ehinger et al., 2014).

Importantly, the processing is continuously documented. To this end, the used parameters and results of the individual steps are stored in the *EEG.etc* structure of the data, which allows the easy inspection and report of processing elements such as the frequencies that were cleaned with Zapline-plus, the interpolated channels, the amount of samples removed by AMICA auto-cleaning (note that these samples are not automatically removed from the final data, but their indices are stored), or the applied filter settings. This transparency is supported by the visualization of the data at several milestones during the cleaning, as well as plots that directly inform the researcher about parts of the automated processes. In order to facilitate their interpretation, a detailed guide is available on the public repository[21]. In summary, the BeMoBIL Pipeline applies the insights gained from previous studies, and provides a reliable and fully automatic processing toolbox for MoBI data analysis, thus facilitating the application of MoBI methods for experts and novices alike.

Automation in EEG data cleaning

As laid out above, the BeMoBIL Pipeline emphasizes automation in data processing. However, this is a controversial topic (Gramann, personal communication; Cohen, 2017a) and cannot simply be considered an advantage without further discussion. This section will detail the arguments against and in favor of automated data processing in general and point out the solutions presented in this dissertation to address the reservations against automation in EEG preprocessing in particular. In doing so, it hopes to convince the reader that the high degree of automation in the BeMoBIL Pipeline is, in fact, an important advantage, especially considering the requirements of MoBI studies and the goals formulated in this dissertation.

[21] https://github.com/BeMoBIL/bemobil-pipeline/wiki/A-guide-to-the-analytics-figures

To introduce the premise: In the last years it was discovered that a significant proportion of results from experiments, not only, but also in psychology and neuroscience, were not replicable, leading to what is now called the "replication crisis" (see e.g. Baker, 2016; Larson & Moser, 2017; Open Science Collaboration, 2015). The reasons for this lack of replicability are manifold and too intricate to be discussed in detail in this work[22], but the main points are: i) Publication bias, i.e. studies that failed to replicate another study are not published and their information is lost (the "file drawer phenomeneon"), ii) selection bias, which means that the high degree of experimenter freedom allows the selection of features in the data (for example of ERPs), until one obtains a statistically significant result (Luck & Gaspelin, 2017), and iii) processing complexity, regarding both the final values, for example of time-frequency analysis, and the preprocessing (Cohen, 2017a). Point iii) is related to point ii), but different insofar as that point ii) indicates an active selection towards a significant result whereas point iii) refers to the overwhelming number of choices that have to be made, which are bound to include minor selections of preferences or randomness that can add up over the entire processing and ultimately lead to the failure of replicating the study.

The BeMoBIL Pipeline naturally cannot affect point i), and it is also not concerned with point ii), as the final single-subject analysis is not part of the pipeline. However, it does improve point iii) as is takes complexity off the shoulders of the researchers and performs a solid analysis of MoBI data based on automatic methods. Cohen (2017a) argues that in the preprocessing phase, most steps are not problematic regarding experimenter bias as they are unlikely to affect replication, with the exception of artifact rejection. Here, two main steps are identified: trial rejection (i.e. time-domain cleaning) and pruning the data with ICA. This can be extended to include channel rejection, as that is an important part of MoBI preprocessing due to the added movement. These three steps are often performed manually, and Cohen (2017a) argues that this is the the preferred option. The main argument in favor of manual cleaning is that automated rejection is suffering from both: misses of elements that are regarded as artifacts by the researcher, and false alarms, the rejection of elements that are not regarded as artifacts by the researcher. Thus, especially when guided by automated pre-assessment (see for example Chaumon et al., 2015, for a guide on IC interpretation), "carefully and meticulously … inspecting the data is the only way to ensure high data quality" (Cohen, 2017a, p. 83). The last, yet highly important, argument against automated processes is that they are often opaque and lead to their blind use, and in turn prevent both their evaluation and the gathering of experience during their use.

Evidently, my opinion is a different one: I see automated processing as the preferred option. It is undoubtedly true that experts with years, or decades, of experience can likely see intricacies in the data that will allow them to pick artifactual elements with high precision. Yet, I do not think this is a sufficient reason to prefer manual cleaning over sophisticated and validated automated tools, so long as those tools are sufficiently reliable and provide the researchers with a high degree of feedback. As introduced in the second quote by Sherlock Holmes in the beginning of this dissertation, "it is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts." I would like to adapt this to fit the current context: We need to avoid theorizing before analyzing the data, as otherwise we might twist the analysis process to obtain results that suit our theories. To be more precise: I do not trust myself to be incontestably objective in my assessments

[22] For an comprehensive special issue on rigor and replication in psychophysiological research, see https://www.sciencedirect.com/journal/international-journal-of-psychophysiology/vol/111/suppl/C.

of what is an artifact and what is not, and I do not want to carry that burden, nor put it on others[23]. The issue here is that we most likely all theorize in some way or another before we collect our data. In fact, the theorization before data collection is a key element that prevents data misuse, as otherwise we fall in the danger of hypothesizing after the results are known (harking). Additionally, other elements such as mood, sleep, or prior experience of the researcher, can influence the cleaning process (Cohen, 2017a). Even expert raters can argue about their assessments and the rater-algorithm agreement might exceed inter-rater agreement (Pion-Tonachini et al., 2019). It can be argued that this issue can be alleviated by having independent raters and finding consensus (e.g. when assessing ICs or identifying bad channels), or by having the person performing the cleaning being blind to the experiment conditions (e.g. when performing time-domain cleaning). Yet, this is not always possible, because, for example, mobile conditions most likely contain more muscle activity than stationary conditions and are thus easily identifiable as such. Also, finding researchers who are actually uninformed about an experiment's hypotheses and theories, yet willing to clean data of said experiment, might prove difficult in practice. Hence, the preferred solution to the issue of subjectivity in artifact rejection is, in my opinion, the use of objective, automated methods.

There are more arguments to be made in favor of automation, though, that are independent of the issue of objectivity and more related to practical considerations: First, manual cleaning takes a lot of time. This time spent cannot be spent on other parts of the scientific work, and thus limits the amount of new insights that can be gained per time. In some way, automated data processing is akin to a process of industrializing science—notwithstanding the issues and limitations this entails, it can lead to great advancements and an acceleration of the creation of knowledge, for example by lowering the threshold to replicate previous studies or to quickly evaluate new theories. Second, a high level of expertise is required for a sophisticated and reliable manual data cleaning. Especially in MoBI, it can take years until the details are understood, and not everyone has this time. MoBI in particular benefits from being accessible to researchers from outside the neuroscience sphere, and opening the gates to these applications by providing a solid automated analysis pipeline can be highly beneficial to all sides (e.g. the study of dance by Barnstaple et al., 2021, who applied the BeMoBIL Pipeline). Lastly, code sharing is an important step in ensuring replicability (Baker, 2016; Cohen, 2017a), and having one comprehensive pipeline that is still easy to use can greatly facilitate this, as otherwise, creating documented and sound analysis code can be a major undertaking (Baker, 2016).

Thus far, only the principle argumentation around automation was discussed. However, there is one argument hidden in the main argument of the prevalence of misses and false alarms in automated cleaning: This is a technical limitation that can, at least in parts, be overcome by improving the algorithms and in turn improve the trust in the automation of the cleaning process. In the BeMoBIL Pipeline, several key issues regarding the points of channel cleaning, time-domain cleaning, ICA cleaning, and IC clustering in the automated methods available thus far were addressed: First, although an effective objective method for detecting bad channels was available by evaluating the channels on their correlation with their own robust interpolation (introduced in the PREP pipeline, Bigdely-Shamlo et al., 2015, and implemented in the *clean raw data* plugin of EEGLAB), this method suffered from a random sampling compo-

[23] I call this method the KonMari method of data cleaning—data that does not spark joy is removed, following the famous tidying-up method of Marie Kondo, https://konmari.com/about-the-konmari-method

nent that could lead to unstable results, thus making the method unreliable and undermining trust. This was accounted for in the pipeline by repeating the process and selecting only channels for removal that are flagged as bad in an specifiable proportion of the repetitions, thus greatly increasing the chances for a reproducible channel cleaning without user intervention. The second point of time-domain cleaning before ICA was difficult to automate especially for EEG data from MoBI experiments, because methods that use amplitude or spectral power of the time-course may trigger an excessive amount of cleaning in mobile conditions. As it was discussed in the second study of chapter 2, AMICA proved to be more stable to this time-domain cleaning than anticipated when using its own automated sample rejection. We also showed that this approach was equal or superior to the removal of time-segments based on our own cleaning method (see supplementary material of chapter 3), which in combination leads us to believe that using the AMICA auto-cleaning is close to an optimal approach to time-domain cleaning when computing ICA. The BeMoBIL Pipeline also allows the balanced removal of epochs before computing ERPs, which is also especially important in MoBI studies, as otherwise the classic approach of removing data that "looks artifactual" may lead to an excessive removal of epochs in mobile conditions. Addressing the third point, removing ICs from the data, the pipeline makes use of ICLabel, a validated algorithm with high reliability (Pion-Tonachini et al., 2019). Yet, while being generally reliable, this is the weakest link in the automated processes in the pipeline, as the classifier was mostly trained on data from stationary experiments and classification can be sub-par to an experienced MoBI researcher's assessment. This is an issue currently not solved, which is why the pipeline provides both an option to use the ICLabel classifer for direct cleaning and a dataset that contains the information of ICLabel but still includes all ICs, thus allowing the guided manual selection of ICs as suggested by Cohen (2017a). Further advancements regarding a MoBI-specific classifier are in progress[24], and in combination with a physiological head model that includes eyes and neck muscles as sources (Harmening, Klug, Gramann, & Miklody, 2022), the classification of ICs will be even more reliable in the future. The last point of automation addressed by the pipeline is the issue of unstable IC clustering results for group-level source analysis. This, too, is solved by a repeated approach and depending on the region of interest, a high number of iterations is necessary to reach a stable result (e.g. 10.000 iterations were necessary to investigate the retrosplenial complex; Gramann et al., 2021), emphasizing the importance of this step. To increase trust in this process, the pipeline visualizes the distribution of the clustering solutions, as well as the best five final clusters of interest to assess their stability.

Taken together, there is a price to pay in both manual and automated data cleaning, but especially in light of the methodological advancements presented here, the benefits of automation exceed the cost of potential mistakes by the algorithms in my view. Supporting the use of automated approaches by visualizing the data and the relevant information of the automated processing for each milestone also further reduces the disadvantages. It is important to keep the processing of the data transparent in every step, as black box processing can lead to blind trust of the automation and loss of skill and insight in the underlying processes. The BeMoBIL Pipeline thus strongly emphasizes the visualization and documentation of the cleaning process and in doing so attempts to achieve the best of both worlds—a high degree of reliability of the automated data cleaning process, and a high degree of feedback to the researcher in order to maintain the option to intervene and learn.

[24] https://www.icmobi.org

Summary, outlook, and conclusions

This dissertation aimed at realizing two major goals for MoBI data analysis: increasing the reliability and the usability of MoBI as a research method by enabling automated processing, in order to make the method more applicable in general and more accessible to researchers from other fields. First, the conceptual background of MoBI was introduced before laying out the applied instrumentation, data analysis methods, and challenges thereof, where the analysis of MoBI data was categorized into the four aspects of data synchronization, body data analysis, sound mobile EEG preprocessing, and the functional interpretation of the combined multimodal data. Chapter 1 then presented Zapline-plus, which removes frequency-specific artifacts from EEG recordings, thus reducing the impact of electrical artifacts in aspect iii). Chapter 2 presented two studies that investigated the question how to obtain an optimal ICA decomposition in aspect iii), with the parameters mobility, channel density, high-pass filter, and time-domain cleaning under investigation. Chapter 3 presented the BeMoBIL Pipeline, a comprehensive MoBI analysis pipeline that addresses all four aspects of MoBI data processing by applying insights from the previous studies and allowing the fully automated processing of multimodal MoBI data while maintaining a high degree of flexibility and insight into the process. The presented works are finally discussed with regards to their contributions to MoBI as a research method, in particular addressing the use of automation when processing MoBI data.

Naturally, this work is not exhaustive regarding the investigations and limitations of MoBI. Further studies should, for example, establish a taxonomy of movement and its effect on EEG data to allow detailed validations. These validations, akin to chapter 2, should then further investigate the effect of different cleaning elements on ICA, for example the effect of Zapline-plus, the effect of the robust channel cleaning, or the use of artifact subspace reconstruction (ASR; Blum, Jacobsen, Bleichner, & Debener, 2019; Mullen et al., 2015), as this was suggested to improve the decomposition (Chang, Hsu, Pion-Tonachini, & Jung, 2020). A limitation not addressed in this work is the stationarity assumption of ICA. It is possible that by the increased complexity of MoBI data, other methods for source imaging and data cleaning could be beneficial, notably recursive sparse Bayesian learning (Ojeda, Klug, Kreutz-Delgado, Gramann, & Mishra, 2019; Ojeda, Kreutz-Delgado, & Mullen, 2018), but this remains to be tested. To further investigate the reliability of the employed automation of the channel and epoch cleaning of the BeMoBIL Pipeline, these methods should be evaluated on their overlap with the assessment of expert raters. Also, the robust clustering of the BeMoBIL Pipeline does not allow analysis without having a region of interest, which should be addressed by future improvements. Other methodological advancements that should be added to the pipeline are, for example, the analysis of motion data from IMUs, as these can be important, and are available in some EEG amplifiers already, or additional options to analyze eye tracking data in light of 3D environments. Also, the BeMoBIL pipeline should be compared with other options to evaluate its effect on downstream analysis. Finally, previous MoBI studies could be replicated using the BeMoBIL Pipeline in order to investigate the consistency of the obtained results (Cohen, 2017a).

I would like to finish this dissertation by adding a last interpretational aspect to the quote by Lt. Cmdr. Data: "It does not matter that we will never reach our ultimate goal. The effort yields its own rewards." When beginning my journey of MoBI research, the original plan was to delve into the field of neuroergonomics (Ayaz & Dehais, 2018; Parasuraman & Rizzo, 2007) to investigate situational awareness (Endsley, 1995). However, I became increasingly unsatisfied with the methods at hand: While

general guidelines were available for analyzing MoBI data, this analysis was time-consuming, as performed manually, or involved stitching together a large set of individual processing steps in which I was unsure if the chosen parameters were optimal. As a consequence, in order to ensure the right settings are chosen and to save several months of manual data cleaning, several years were spent to create a pipeline that does it for us. The irony of this choice is in your hands: The ultimate goal of doing neuroergonomics research was never reached, as the path became the goal, and the reward of the spent effort is this dissertation.

To conclude, as King Julien remarks so pointedly in the last quote introduced in the beginning of this work: We like to move it. Understanding this movement and its impact on our cognition, how we perceive the world and interact with it, is a never ending endeavor, for which this dissertation provides a set of guidelines and tools. So we can move one step closer to answering the question: What goes on in our minds when we do something?

References

Acar, Z. A., & Makeig, S. (2010). Neuroelectromagnetic forward head modeling toolbox. *J. Neurosci. Methods*, *190*(2), 258–270.

Aricò, P., Borghini, G., Di Flumeri, G., Colosimo, A., Pozzi, S., & Babiloni, F. (2016). A passive brain–computer interface application for the mental workload assessment on professional air traffic controllers during realistic air traffic control tasks. *Prog. Brain Res.*, *228*, 295–328.

Ayaz, H., & Dehais, F. (Eds.). (2018). *Neuroergonomics - the brain at work and in everyday life* (1st ed.). London: Elsevier Academic Press.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452–454.

Banaei, M., Hatami, J., Yazdanfar, A., & Gramann, K. (2017). Walking through architectural spaces: The impact of interior forms on human brain dynamics. *Front. Hum. Neurosci.*, *11*, 477.

Barnstaple, R., Protzak, J., DeSouza, J. F. X., & Gramann, K. (2021). Mobile brain/body imaging in dance: A dynamic transdisciplinary field for applied research. *Eur. J. Neurosci.*, *54*(12), 8355–8363.

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Comput.*, *7*(6), 1129–1159.

Biebmann, F., Meinecke, F. C., Gretton, A., Rauch, A., Rainer, G., Logothetis, N. K., & Müller, K. R. (2010). Temporal kernel CCA and its application in multimodal neuronal data analysis. *Mach. Learn.*, *79*(1-2), 5–27.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., & Robbins, K. A. (2015). The PREP pipeline: standardized preprocessing for large-scale EEG analysis. *Front. Neuroinform.*, *9*.

Blanke, O., & Metzinger, T. (2009). Full-body illusions and minimal phenomenal selfhood. *Trends Cogn. Sci.*, *13*(1), 7–13.

Blum, S., Jacobsen, N. S. J., Bleichner, M. G., & Debener, S. (2019). A riemannian modification of artifact subspace reconstruction for EEG artifact handling. *Front. Hum. Neurosci.*, *13*, 141.

Bohbot, V. D., Copara, M. S., Gotman, J., & Ekstrom, A. D. (2017). Low-frequency theta oscillations in the human hippocampus during real-world and virtual navigation. *Nat. Commun.*, *8*, 14415.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neurosci. Biobehav. Rev.*, *44*, 58–75.

Boto, E., Holmes, N., Leggett, J., Roberts, G., Shah, V., Meyer, S. S., … Brookes, M. J. (2018). Moving magnetoencephalography towards real-world applications with a wearable system. *Nature*, *555*(7698), 657–661.

Buzsaki, G. (2006). *Rhythms of the brain*. New York: Oxford University Press.

Buzsáki, G. (2019). *The brain from inside out*. New York: Oxford University Press.

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes. II: Arm flexion and extension have differential effects on attitudes. *J. Pers. Soc. Psychol.*, *65*(1), 5–17.

Castermans, T., Duvinage, M., Cheron, G., & Dutoit, T. (2014). About the cortical origin of the low-delta and high-gamma rhythms observed in EEG signals during treadmill walking. *Neurosci. Lett.*, *561*, 166–170.

Chang, C. Y., Hsu, S. H., Pion-Tonachini, L., & Jung, T. P. (2020). Evaluation of artifact subspace reconstruction for automatic artifact components removal in Multi-Channel EEG recordings. *IEEE Transactions on Biomedical Engineering*, *67*(4), 1114–1121.

Chaumon, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *J. Neurosci.*

Methods, *250*, 47–63.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Pers. Soc. Psychol. Bull.*, *25*(2), 215–224.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*(1), 7–19.

Cohen, M. X. (2014). *Analyzing neural time series data: Theory and practice*. MIT Press.

Cohen, M. X. (2017a). Rigor and replication in time-frequency analyses of cognitive electrophysiology data. *Int. J. Psychophysiol.*, *111*, 80–87.

Cohen, M. X. (2017b). Where does EEG come from and what does it mean? *Trends Neurosci.*, *40*(4), 208–218.

Cruz-Neira, C., Sandin, D. J., DeFanti, T. A., Kenyon, R. V., & Hart, J. C. (1992). The CAVE: audio visual experience automatic virtual environment. *Commun. ACM*, *35*(6), 64–72.

Curtin, A., & Ayaz, H. (2018). The age of neuroergonomics: Towards ubiquitous and continuous measurement of brain function with fNIRS. *Jpn. Psychol. Res.*, *60*(4), 374–386.

Dähne, S., Nikulin, V. V., Ramírez, D., Schreier, P. J., Müller, K. R., & Haufe, S. (2014). Finding brain oscillations with power dependencies in neuroimaging data. *Neuroimage*, *96*, 334–348.

Debener, S., Minow, F., Emkes, R., Gandras, K., & de Vos, M. (2012). How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology*, *49*(11), 1617–1621.

de Cheveigné, A. (2020). ZapLine: A simple and effective method to remove power line artifacts. *Neuroimage*, *207*, 116356.

de Cheveigné, A., & Arzounian, D. (2015). Scanning for oscillations. *J. Neural Eng.*, *12*, 066020.

de Cheveigné, A., & Parra, L. C. (2014). Joint decorrelation, a versatile tool for multichannel data analysis. *Neuroimage*, *98*, 487–505.

Dehais, F., Duprès, A., Blum, S., Drougard, N., Scannella, S., Roy, R., & Lotte, F. (2019). Monitoring pilot's mental workload using ERPs and spectral power with a Six-Dry-Electrode EEG system in real flight conditions. *Sensors*, *19*(6), 1324.

Dehais, F., Roy, R. N., Durantin, G., Gateau, T., & Callan, D. (2018). EEG-Engagement index and auditory alarm misperception: An inattentional deafness study in actual flight condition. In C. Baldwin (Ed.), *Advances in neuroergonomics and cognitive engineering* (pp. 227–234). Cham: Springer International Publishing.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods*, *134*(1), 9–21.

Delorme, A., Mullen, T., Kothe, C., Akalin Acar, Z., Bigdely-Shamlo, N., Vankov, A., & Makeig, S. (2011). EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. *Comput. Intell. Neurosci.*, *2011*, 130714.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., & Makeig, S. (2012). Independent EEG sources are dipolar. *PLoS One*, *7*(2), e30135.

De Sanctis, P., Butler, J. S., Malcolm, B. R., & Foxe, J. J. (2014). Recalibration of inhibitory control systems during walking-related dual-task interference: a mobile brain-body imaging (MOBI) study. *Neuroimage*, *94*, 55–64.

Dimigen, O. (2020). Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. *Neuroimage*, *207*, 116117.

Dimigen, O., & Ehinger, B. V. (2021). Regression-based analysis of combined EEG and eye-tracking data: Theory and applications. *J. Vis.*, *21*(1), 3.

Djebbara, Z., Fich, L. B., Petrini, L., & Gramann, K. (2019). Sensorimotor brain dynamics reflect architectural affordances. *Proceedings of the National Academy of Sciences*, *116*(29), 14769–14778.

Ehinger, B. V., & Dimigen, O. (2019). Unfold: an integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *PeerJ*, *7*, e7838.

Ehinger, B. V., Fischer, P., Gert, A. L., Kaufhold, L., Weber, F., Pipa, G., & König, P. (2014). Kinesthetic and vestibular information modulate alpha activity during spatial navigation: a mobile EEG study. *Front. Hum. Neurosci.*, *8*, 71.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64.

Engel, A. K., Maye, A., Kurthen, M., & König, P. (2013). Where's the action? the pragmatic turn in cognitive science. *Trends Cogn. Sci.*, *17*(5), 202–209.

Feinberg, I. (1978). Efference copy and corollary discharge: implications for thinking and its disorders. *Schizophr. Bull.*, *4*(4), 636–640.

Fischer, H. (2011). *A history of the central limit theorem* (Buchwald, J, Z, Berggren, J, L, & J. Lützen, Eds.). New York: Springer.

Gabard-Durnam, L. J., Leal, A. S. M., Wilkinson, C. L., & Levin, A. R. (2018). The harvard automated processing pipeline for electroencephalography (HAPPE): Standardized processing software for developmental and high-artifact data. *Front. Neurosci.*, *12*, 97.

Gateau, T., Ayaz, H., & Dehais, F. (2018). In silico vs. over the clouds: On-the-fly mental state estimation of aircraft pilots, using a functional near infrared spectroscopy based passive-BCI. *Front. Hum. Neurosci.*, *12*, 187.

Gehrke, L., Akman, S., Lopes, P., Chen, A., Singh, A. K., Chen, H.-T., … Gramann, K. (2019). Detecting Visuo-Haptic mismatches in virtual reality using the prediction error negativity of Event-Related brain potentials. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–11). New York, NY, USA: Association for Computing Machinery.

Gehrke, L., Guerdan, L., & Gramann, K. (2019). Extracting Motion-Related subspaces from EEG in mobile Brain/Body imaging studies using source power comodulation. In *International IEEE/EMBS conference on neural engineering, NER* (Vol. 2019-March, pp. 344–347). IEEE Computer Society.

Gehrke, L., Iversen, J. R., Makeig, S., & Gramann, K. (2018). The invisible maze task (IMT): Interactive exploration of sparse virtual environments to investigate action-driven formation of spatial representations. In S. Creem-Regehr, J. Schöning, & A. Klippel (Eds.), *Spatial cognition XI* (pp. 293–310). Cham: Springer International Publishing.

Gerjets, P., Walter, C., Rosenstiel, W., Bogdan, M., & Zander, T. O. (2014). Cognitive state monitoring and the design of adaptive instruction in digital environments: Lessons learned from cognitive workload assessment using a passive brain-computer interface approach. *Front. Neurosci.*, *8*, 385.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human–computer interaction. *Theoretical Issues in Ergonomics Science*, *4*(1-2), 113–131.

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., … Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 1–9.

Gramann, K., Ferris, D. P., Gwin, J., & Makeig, S. (2014). Imaging natural cognition in action. *Int. J. Psychophysiol.*, *91*(1), 22–29.

Gramann, K., Gwin, J. T., Bigdely-Shamlo, N., Ferris, D. P., & Makeig, S. (2010). Visual evoked responses during standing and walking. *Front. Hum. Neurosci.*, *4*, 202.

Gramann, K., Gwin, J. T., Ferris, D. P., Oie, K., Jung, T. P., Lin, C. T., … Makeig, S. (2011). Cognition in action: Imaging brain/body dynamics in mobile humans. *Rev. Neurosci.*, *22*(6), 593–608.

Gramann, K., Hohlefeld, F. U., Gehrke, L., & Klug, M. (2021). Human cortical dynamics during full-body heading changes. *Sci. Rep.*, *11*(1), 18186.

Gramann, K., Onton, J., Riccobon, D., Mueller, H. J., Bardins, S., & Makeig, S. (2010). Human brain dynamics accompanying use of egocentric and allocentric reference frames during

navigation. *J. Clin. Psychol.*, *22*(12), 2836–2849.

Gwin, J. T., Gramann, K., Makeig, S., & Ferris, D. P. (2010). Removal of movement artifact from high-density EEG recorded during walking and running. *J. Neurophysiol.*, *103*(6), 3526–3534.

Harmening, N., Klug, M., Gramann, K., & Miklody, D. (2022). HArtMuT - modeling eye and muscle contributors in neuroelectric imaging. *bioRxiv*.

Hoogenboom, N., Schoffelen, J.-M., Oostenveld, R., Parkes, L. M., & Fries, P. (2006). Localizing human visual gamma-band activity in frequency, time and space. *Neuroimage*, *29*(3), 764–773.

Hyvärinen, A., Karhunen, J., & Oja, E. (2001). *Independent component analysis*. New York: John Wiley & Sons.

Jungnickel, E., Gehrke, L., Klug, M., & Gramann, K. (2019). MoBI-Mobile Brain/Body imaging. In H. Ayaz & F. Dehais (Eds.), *Neuroergonomics: The brain at work and in everyday life* (1st ed., pp. 59–63). London: Elsevier.

Jungnickel, E., & Gramann, K. (2016). Mobile Brain/Body imaging (MoBI) of physical interaction with dynamically moving objects. *Front. Hum. Neurosci.*, *10*, 306.

Kirsh, D. (1995). The intelligent use of space. *Artif. Intell.*, *73*(1), 31–68.

Klug, M., & Gramann, K. (2021). Identifying key factors for improving ICA-based decomposition of EEG data in mobile and stationary experiments. *Eur. J. Neurosci.*, *54*(12), 8406–8420.

Klug, M., & Kloosterman, N. A. (2022). Zapline-plus: A zapline extension for automatic and adaptive removal of frequency-specific noise artifacts in M/EEG. *Hum. Brain Mapp.*, *43*(9), 2743–2758.

Kohlmorgen, J., Dornhege, G., Braun, M, L, Blankertz, B., Müller, K.-R., Curio, G., … Kincses, W, E (2007). Improving human performance in a real operating environment through Real-Time mental workload detection. In G. Dornhege, J. del R. Millan, T. Hinterberger, D. J. McFarland, & K.-R. Müller (Eds.), *Toward brain-computer interfacing* (pp. 409–422). Cambridge: MIT Press.

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., & Ietswaart, M. (2017). Understanding minds in Real-World environments: Toward a mobile cognition approach. *Front. Hum. Neurosci.*, *10*, 694.

Ladouce, S., Donaldson, D. I., Dudchenko, P. A., & Ietswaart, M. (2019). Mobile EEG identifies the re-allocation of attention during real-world activity. *Sci. Rep.*, *9*, 15851.

Larson, M. J., & Moser, J. S. (2017). Rigor and replication: Toward improved best practices in human electrophysiology research. *Int. J. Psychophysiol.*, *111*, 1–4.

Lin, C.-T., Chiu, T.-C., & Gramann, K. (2015). EEG correlates of spatial orientation in the human retrosplenial complex. *Neuroimage*, *120*, 123–132.

Luck, S. J. (2014). *An introduction to the Event-Related potential technique, second edition*. Cambridge: MIT Press.

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157.

Makeig, S., Gramann, K., Jung, T. P., Sejnowski, T. J., & Poizner, H. (2009). Linking brain, mind and behavior. *Int. J. Psychophysiol.*, *73*(2), 95–100.

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2015). The psychometrics of mental workload: Multiple measures are sensitive but divergent. *Hum. Factors*, *57*(1), 125–143.

McKendrick, R., Parasuraman, R., Murtza, R., Formwalt, A., Baccus, W., Paczynski, M., & Ayaz, H. (2016). Into the wild: Neuroergonomic differentiation of Hand-Held and augmented reality wearable displays during outdoor navigation with functional near infrared spectroscopy. *Front. Hum. Neurosci.*, *10*, 216.

Mehta, R. K., & Parasuraman, R. (2013). Neuroergonomics: a review of applications to physical and cognitive work. *Front. Hum. Neurosci.*, *7*, 889.

Melnik, A., Legkov, P., Izdebski, K., Kärcher, S. M., Hairston, W. D., Ferris, D. P., & König, P. (2017). Systems, subjects, sessions: To what extent do these factors influence EEG data? *Front. Hum. Neurosci.*, *11*, 150.

Miklody, D., Moessmer, P., Dettmann, T., Klinkenberg, K., & Blankertz, B. (2017). Multi-timescale spectra as features for continuous workload estimation in realistic settings. In G. R. Müller-Putz, S. C. Wriessnegger, & R. Scherer (Eds.), *Proceedings of the 7th graz Brain-Computer interface conference.* Graz: Verlag der Technischen Universität Graz.

Mitz, A. R., Chacko, R. V., Putnam, P. T., Rudebeck, P. H., & Murray, E. A. (2017). Using pupil size and heart rate to infer affective states during behavioral neurophysiology and neuropsychology experiments. *J. Neurosci. Methods*, *279*.

Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., … Cauwenberghs, G. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Transactions on Biomedical Engineering*, *62*(11), 2553–2567.

Neisser, U. (2014). *Cognitive psychology: Classic edition*. New York: Psychology press.

Nenna, F., Do, C. T., Protzak, J., & Gramann, K. (2021). Alteration of brain dynamics during dual-task overground walking. *Eur. J. Neurosci.*, *54*(12), 8158–8174.

Nordin, A. D., Hairston, W. D., & Ferris, D. P. (2019). Human electrocortical dynamics while stepping over obstacles. *Sci. Rep.*, *9*(1), 1–12.

Ojeda, A., Bigdely-Shamlo, N., & Makeig, S. (2014). MoBILAB: an open source toolbox for analysis and visualization of mobile brain/body imaging data. *Front. Hum. Neurosci.*, *8*, 121.

Ojeda, A., Klug, M., Kreutz-Delgado, K., Gramann, K., & Mishra, J. (2019). A bayesian framework for unifying data cleaning , source separation and imaging of electroencephalographic signals. *bioRxiv*.

Ojeda, A., Kreutz-Delgado, K., & Mullen, T. (2018). Fast and robust Block-Sparse bayesian learning for EEG source imaging. *Neuroimage*, *174*, 449–462.

Onton, J., & Makeig, S. (2006). Information-based modeling of event-related brain dynamics. *Prog. Brain Res.*, *159*, 99–120.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251).

Palmer, J. A., Kreutz-Delgado, K., & Makeig, S. (2011). AMICA : An adaptive mixture of independent component analyzers with shared components. , 1–15.

Parasuraman, R., & Rizzo, M. (Eds.). (2007). *Neuroergonomics - the brain at work*. New York: Oxford University Press.

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behavior*. Cambridge: MIT Press.

Pernet, C. R., Appelhoff, S., Gorgolewski, K. J., Flandin, G., Phillips, C., Delorme, A., & Oostenveld, R. (2019). EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data*, *6*(1), 1–5.

Pernet, C. R., Martinez-Cancino, R., Truong, D., Makeig, S., & Delorme, A. (2021). From BIDS-Formatted EEG data to Sensor-Space group results: A fully reproducible workflow with EEGLAB and LIMO EEG. *Front. Neurosci.*, *14*, 610388.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage*, *198*, 181–197.

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology*, *40*(1), 187-195.

Protzak, J., & Gramann, K. (2021). EEG beta-modulations reflect age-specific motor resource

allocation during dual-task walking. *Sci. Rep.*, *11*(1), 16110.

Protzak, J., Wiczorek, R., & Gramann, K. (2021). Peripheral visual perception during natural overground dual-task walking in older and younger adults. *Neurobiol. Aging*, *98*, 146–159.

Raez, M. B. I., Hussain, M. S., & Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: detection, processing, classification and applications. *Biol. Proced. Online*, *8*, 11–35.

Reiser, J. E., Wascher, E., & Arnau, S. (2019). Recording mobile EEG in an outdoor environment reveals cognitive-motor interference dependent on movement complexity. *Sci. Rep.*, *9*, 13086.

Revonsuo, A. (2010). *Consciousness: The science of subjectivity*. New York: Psychology Press.

Robbins, P., & Aydede, M. (2008). A short primer on situated cognition. In P. Robbins & M. Aydede (Eds.), *The cambridge handbook of situated cognition* (pp. 3–10). Cambridge: Cambridge University Press.

Sebastiani, M., Di Flumeri, G., Aricò, P., Sciaraffa, N., Babiloni, F., & Borghini, G. (2020). Neurophysiological vigilance characterisation and assessment: Laboratory and realistic validations involving professional air traffic controllers. *Brain Sciences*, *10*(1), 1–22.

Seeber, M., Scherer, R., Wagner, J., Solis-Escalante, T., & Müller-Putz, G. R. (2015). High and low gamma EEG oscillations in central sensorimotor areas are conversely modulated during the human gait cycle. *Neuroimage*, *112*, 318–326.

Skinner, B. F. (2011). *About behaviorism*. New York: Random House.

Snider, J., Plank, M., Lynch, G., Halgren, E., & Poizner, H. (2013). Human cortical theta during free exploration encodes space and predicts subsequent memory. *Journal of Neuroscience*, *33*(38), 15056–15068.

Snyder, K. L., Kline, J. E., Huang, H. J., & Ferris, D. P. (2015). Independent component analysis of gait-related movement artifact recorded using EEG electrodes during treadmill walking. *Front. Hum. Neurosci.*, *9*, 639.

Thompson, E., & Varela, F. J. (2001). Radical embodiment: neural dynamics and consciousness. *Trends Cogn. Sci.*, *5*(10), 418–425.

Wagner, J., Makeig, S., Gola, M., Neuper, C., & Muller-Putz, G. (2016). Distinct band oscillatory networks subserving motor and cognitive control during gait adaptation. *Journal of Neuroscience*, *36*(7), 2212–2226.

Wagner, J., Martínez-Cancino, R., & Makeig, S. (2019). Trial-by-trial source-resolved EEG responses to gait task challenges predict subsequent step adaptation. *Neuroimage*, *199*, 691–703.

Wagner, J., Solis-Escalante, T., Grieshofer, P., Neuper, C., Müller-Putz, G., & Scherer, R. (2012). Level of participation in robotic-assisted treadmill walking modulates midline sensorimotor EEG rhythms in able-bodied subjects. *Neuroimage*, *63*(3), 1203–1211.

Wahn, B., Ferris, D. P., Hairston, W. D., & König, P. (2016). Pupil sizes scale with attentional load and task experience in a multiple object tracking task. *PLoS One*, *11*(12), 9–22.

Waldert, S., Preissl, H., Demandt, E., Braun, C., Birbaumer, N., Aertsen, A., & Mehring, C. (2008). Hand movement direction decoded from MEG and EEG. *J. Neurosci.*, *28*(4), 1000–1008.

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychol. Rev.*, *20*(2), 158–177.

Widmann, A., Schröger, E., & Maess, B. (2015). Digital filter design for electrophysiological data – a practical approach. *J. Neurosci. Methods*, *250*, 34–46.

Wilson, G. F., & Russell, C. A. (2007). Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. *Hum. Factors*, *49*(6), 1005–1018.

Wilson, M. (2002). Six views of embodied cognition. *Psychon. Bull. Rev.*, *9*(4), 625–636.

Winkler, I., Debener, S., Müller, K.-R., & Tangermann, M. (2015). On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *Conf. Proc. IEEE Eng. Med. Biol. Soc.*,

2015, 4101–4105.

Wunderlich, A., & Gramann, K. (2021). Eye movement-related brain potentials during assisted navigation in real-world environments. *Eur. J. Neurosci.*, *54*(12), 8336–8354.

Yeo, L. G., Sun, H., Liu, Y., Trapsilawati, F., Sourina, O., Chen, C.-H., … Ang, W. T. (2017). Mobile eeg-based situation awareness recognition for air traffic controllers. In *2017 ieee international conference on systems, man, and cybernetics (smc)* (p. 3030-3035).

Zander, T. O., Andreessen, L. M., Berg, A., Bleuel, M., Pawlitzki, J., Zawallich, L., … Gramann, K. (2017). Evaluation of a dry EEG system for application of passive Brain-Computer interfaces in autonomous driving. *Front. Hum. Neurosci.*, *11*, 78.