Signal Processing Methods for Reliable Extraction of Neural Responses in Developmental EEG

Velu Prabhakar Kumaravel

Prof Massimo Turatto, Program Coordinator
Dr Elisabetta Farella, Thesis Advisor
Dr Marco Buiatti, Thesis Co-Advisor
Prof Judit Gervain, Reviewer
Dr Silvia Orlandi, Reviewer

Trento, University of Trento, 2023

You are what you believe in.
You become that which you believe you can become.
— Bhagavad Gita

To my mother,
my father,
all my teachers,
and
God of all religions...

Acknowledgements

When I left India in 2015 to pursue a master's in Italy, I had only one professional goal: Pursuing PhD in an interdisciplinary field integrating Neuroscience and Engineering. My Thesis Advisor, Dr Elisabetta Farella (affiliated with Fondazione Bruno Kessler, Italy), has been instrumental in making this happen. Dr Farella took my passion seriously and found a new collaboration with Dr Marco Buiatti (affiliated with the University of Trento, Italy). When this thesis project was proposed, I remember how exciting it was. Indeed, the path was challenging, but after 4 years of research with continued support from my advisors, I am more than satisfied today. My sincere gratitude to my advisors, Dr Farella and Dr Buiatti, for providing me with a challenging and fulfilling PhD experience in Italy.

I thank Fondazione Bruno Kessler (FBK) for funding my research and offering a rich research environment. The members of the E3DA unit, though a small, exclusive group of individuals, perform research on diverse and exciting topics (AI at the edge for Speech/Video Processing, Sensor Localization). As such, there was always something to learn from my group members. I thank my colleagues Amy Murphy, Rajeev Piyare, Davide Giovanelli, Gianmarco Cerutti, Francesco Paissan, Elia Leoni, and Alberto Ancilotto for their valuable formal and informal comments and companionship.

Among my collaborators, I'd like to thank Francesco Paissan from FBK, Victor Kartsch and Simone Benatti from the University of Bologna. I thank Prof Dr David Atienza for accepting me as a visiting PhD student at the ESL laboratory, EPFL, Switzerland. Prof Atienza has provided a friendly environment in the lab, organizing various lab activities and events like marathons, dinners, ski trips, and summer activities. It was a rich learning experience working with Tomas Teijeiro (who respected my independence in research and supported me with his excellent ideas) and Una Pale (who was actively involved in brainstorming sessions in facing the technical challenges) in Machine Learning algorithms for epileptic seizure detection.

During these years, I enjoyed the friendship with Balaji, Yamini, Gianmarco, Elia, Raktim, Stefano, Vincent, Pierre-yves, Rafael, Ruben, Dimitrios, and Christodoulos in Italy and Switzerland.

Finally, I am very grateful for abundant love and support from my family in India, my lovely parents, my supportive younger brother, and my wife.

Trento, 27 February 2023 VP. K.

Abstract

Studying newborns in the first days of life prior to experiencing the world provides remarkable insights into the neurocognitive predispositions that humans are endowed with. First, it helps us to improve our current knowledge of the development of a typical brain. Secondly, it potentially opens new pathways for earlier diagnosis of several developmental neurocognitive disorders such as Autism Spectrum Disorder (ASD). While most studies investigating early cognition in the literature are purely behavioural, recently there has been an increasing number of neuroimaging studies in newborns and infants.

Electroencephalography (EEG) is one of the most optimal neuroimaging technique to investigate neurocognitive functions in human newborns because it is non-invasive and quick and easy to mount on the head. Since EEG offers a versatile design with custom number of channels/electrodes, an ergonomic wearable solution could help study newborns outside clinical settings such as their homes. Compared to adult EEG, newborn EEG data are different in two main aspects: 1) In experimental designs investigating stimulus-related neural responses, collected data is extremely short in length due to the reduced attentional span of newborns; 2) Data is heavily contaminated with noise due to their uncontrollable movement artifacts. Since EEG processing methods for adults are not adapted to very short data length and usually deal with well-defined, stereotyped artifacts, they are unsuitable for newborn EEG. As a result, researchers manually clean the data, which is a subjective and time-consuming task. This thesis work is specifically dedicated to developing (semi-) automated novel signal processing methods for noise removal and for extracting reliable neural responses specific to this population. The solutions are proposed for both high-density EEG for traditional lab-based research and wearable EEG for clinical applications.

To this end, this thesis, first, presents novel signal processing methods applied to newborn EEG: 1) Local Outlier Factor (LOF) for detecting and removing bad/noisy channels; 2) Artifacts Subspace Reconstruction (ASR) for detecting and removing or correcting bad/noisy segments. Then, based on these algorithms and other preprocessing functionalities, a robust preprocessing pipeline, Newborn EEG Artifact Removal (NEAR), is proposed. Notably, this is the first time LOF is explored for EEG bad channel detection, despite being a popular outlier detection technique in other kinds of data such as Electrocardiogram (ECG). Even if ASR is already an established artifact real algorithm originally developed for mobile adult EEG, this

thesis explores the possibility of adapting ASR for short newborn EEG data, which is the first of its kind. NEAR is validated on simulated, real newborn, and infant EEG datasets. We used the SEREEGA toolbox to simulate neurologically plausible synthetic data and contaminated a certain number of channels and segments with artifacts commonly manifested in developmental EEG. We used newborn EEG data ($n = 10$, age range: 1 and 4 days) recorded in our lab based on a frequency-tagging paradigm. The chosen paradigm consists of visual stimuli to investigate the cortical bases of facelike pattern processing, and the results were published in 2019. To test NEAR performance on an older population with an event-related design (ERP) and with data recorded in another lab, we also evaluated NEAR on infant EEG data recorded on 9-months-old infants ($n = 14$) with an ERP paradigm. The experimental paradigm for these datasets consists of auditory stimulus to investigate the electrophysiological evidence for understanding maternal speech, and the results were published in 2012. Since authors of these independent studies employed manual artifact removal, the obtained neural responses serve as ground truth for validating NEAR's artifact removal performance. For comparative evaluation, we considered the performance of two state-of-the-art pipelines designed for older infants. Results show that NEAR is successful in recovering the neural responses (specific to the EEG paradigm and the stimuli) compared to the other pipelines.
In sum, this thesis presents a set of methods for artifact removal and extraction of stimulus-related neural responses specifically adapted to newborn and infant EEG data that will hopefully contribute to strengthening the reliability and reproducibility of developmental cognitive neuroscience studies, both in research laboratories and in clinical applications.

Key words: Developmental Cognitive Science, EEG, Signal Processing, Artifacts, Frequency-tagging (FT), Stead-State Visually Evoked Potentials (SSVEP), Local Outlier Factor (LOF), Artifacts Subspace Reconstruction (ASR), Canonical Correlation Analysis (CCA), Wearable EEG, Neural Networks.

Contents

List of Figures

List of Tables

Acronyms

ADC Analog-to-Digital Converter

AFE Analog Front End

ASD Autism Spectrum Disorder

ASR Artifacts Subspace Reconstruction

BLE Bluetooth Low Energy

CCA Canonical Correlation Analysis

CMRR Common Model Rejection Ratio

CNN Convolutional Neural Network

CRD Clean Raw Data

EEG Electroencephalogram

ERP Event-Related Potentials

FASTER Fully Automated Statistical Thresholding for EEG artifact Rejection

FFT Fast Fourier Analysis

FT Frequency-tagging

FTA Frequency-tagging Analyis

FTR Frequency-tagged Response

HAPPE Harvard Automated Processing Pipeline for Electroencephalography

ICA Independent Component Analysis

IMU Inertial Measurement Unit

LOF Local Outlier Factor

MEG Magnetoencephalogram

MADE Maryland Analysis of Developmental EEG

NCCA Normalized Canonical Correlation Analysis

NEAR Newborn EEG Artifact Removal

NN Neural Network

NFC Near-field Communication

PCA Principal Component Analysis

PSD Power Spectral Density

PULP Parallel Ultra-Low Power

SEREEGA Simulating Event-Related EEG Activity

SNR Signal-to-Noise Ratio

SOBI Second-Order Blind Identification

SOC System-on-Chip

SPI Serial Peripheral Interface

SSVEP Steady-State Visually Evoked Potentials

1 Introduction

Studying early brain development and functioning has significant theoretical implications as it provides unique opportunities to understand the origins of human cognitive functions (Reynolds and Richards, 2005, Buiatti et al., 2019). Progress in this research area could be helpful in two main aspects: 1) We acquire more knowledge on the neurocognitive predispositions humans are endowed with by studying typical human brains. 2) Understanding how typical and atypical brains differ in early cognition and perception could help us discover new pathways for early diagnosis and treatment of various cognitive developmental disorders such as aphasia, dyslexia, attention deficits, or autism (Johnson, 2014).

1.1 Importance of studying newborns

Let us consider Autism Spectrum Disorder (ASD) as a use case to motivate the objectives set for this thesis. ASD is a neurodevelopmental disorder highlighted by impairments in social communication skills and the presence of restricted and repetitive behaviour and interests (American Psychiatric Association, 2013). Some critical behavioural traits of ASD have been linked to impairments in the social brain, a network of cortical areas and subcortical structures responsible for processing social stimuli. Several behavioural studies demonstrated substantial evidence of social communication deficits in ASD-developing infants. For example, deficits in the initiation of social interaction are observed in infants between 12 and 14 months of age, which include: decreased showing and pointing, fewer gestures, lower rate of joint attention initiation, and less gaze alteration (Barbaro and Dissanayake, 2012; Landa, 2007; Macari et al., 2012; Sheppard et al., 2015; Talbott et al., 2015; Zwaigenbaum et al., 2004). Further, between 6 and 12 months of age, infants who develop ASD begin to show decreased attention to faces and reduced response to their own names (Chawarska et al., 2013; Feldman, 2012; Nadig et al., 2007). Such deficits in attention become more pronounced by 12 months of age when high-risk infants are identified with reduced social smiling, poor gazing to faces, less directed vocalisation, and less attention to their mothers (Hutman et al., 2010; Luyster et al., 2009; Ozonoff et al., 2010; Wan et al., 2012). It can be concluded from this brief review that social communication deficits are evident by 12 months of age, while language deficits begin

to emerge. In sum, these studies show that robust behavioural differences are noticeable by the second year of life in high-risk infants, with less certainty in the first year. This means that a reliable clinical diagnosis is only possible in the second year of age.

Despite this evidence, the developmental origins of such abnormalities were unclear. Were the impairments in the social brain already present during birth or developed during infancy? A critical hypothesis is that the unusual development of the social brain in ASD individuals may be due to a modification and delay in the early activation of subcortical orienting mechanisms (Johnson, 2014). However, this hypothesis was challenged by two major studies, where infants of 2 months old demonstrated normal visual orientation to the presented social stimulus but were later diagnosed with ASD (Jones and Klin, 2013); and autistic adults have a functional subcortical orienting mechanism as typical adults. These studies were conducted in infants as old as two months of age or adults and, therefore, cannot be conclusive as the participants already had plenty of visual exposure to social stimuli. Therefore, direct evidence by studying newborns and young infants before their exposure to the outside world is crucial. To sum up, investigating early social cognition in newborns (before their exposure to the outside world) will bring invaluable knowledge of the underlying developmental origins of autism and also possibly help in earlier diagnosis (i.e., within the first year of life, when the developing neural system is more plastic for intervention).

According to the infant-sibling model of ASD, infants with one older sibling with ASD are at 20% risk of developing ASD, while infants with more than one older sibling with ASD are at 33-50% risk of ASD (Messinger et al., 2013; Ozonoff et al., 2010). Following this model, familial risk newborns can be studied from birth which can provide a remarkable opportunity to study the earliest manifestation of atypical brain development. Such studies can help identify the predictive markers of social communication deficits, guide the appropriate implementation of earlier behavioural interventions, and shed light on the developmental origins of the disorder. To this end, recently, Italian scientists conducted a behavioural study to investigate the inborn predisposition to pay attention to multiple classes of social stimuli, including face preference (Di Giorgio et al., 2016). The study, performed in 17 high-risk and 17 low-risk newborns, provided the first evidence that visual attention to social stimuli impairments are present very early in newborns with high familial risk for ASD.

While all the research discussed so far are purely behavioural, neural-based biomarkers for ASD in newborns would enable an additional screening test facilitating a reliable earlier diagnosis system. Thanks to the recent emergence of high-quality Electroencephalography (EEG) systems with innovative paradigms suitable for newborns paved the way for an increasing number of investigations on the neural bases of such predispositions with EEG (Beauchemin et al., 2010; Buiatti et al., 2019; Fifer et al., 2010; Ronga et al., 2021). EEG applied to developmental research may provide an irreplaceable source of knowledge to understand the temporal dynamics of cognitive processes in typical and atypical populations (Hervé et al., 2022). However, preprocessing newborn EEG data is challenging, especially when a visual stimulus is involved in the study. Until this work, there was no standard and/or automated

preprocessing procedure for this data. Currently, developmental researchers perform time-consuming manual processing of data, which hinders progress in this field. This thesis aims to resolve this problem by presenting an automated pipeline comprising robust artifact removal methods adapted to newborn EEG. The proposed methods are validated both on high-density (aimed for lab-based research) and low-density EEG data (useful for clinical applications). In the next section, we will expand on current challenges in newborn EEG studies.

1.2 Challenges in Newborn/Infant EEG Studies

1.2.1 Short Recording Time

Newborns and infants have limited visual attention within the first year of life (Atkinson and Braddick, 2012), and they have reduced tolerance for testing compared to healthy adults. For these reasons, the experimental paradigms/protocols designed for these populations have far shorter EEG acquisition times than the ones designed for adults. Usually, the researchers present repeated trials (each comprised of short duration) as long as the infant is attentive (see Table 1.1 (Column "Stimulus Duration")). Moreover, the experimental stimuli are carefully chosen to keep their attention as long as possible (Parise and Csibra, 2012; Tran et al., 2004).

Study	Mean Age	Usable Files	Stimulus Duration	Unusable Files
Beauchemin et al., 2010	21 hours	16	212 ms	NS
Fifer et al., 2010	40 hours	30	1 s	4
Ronga et al., 2021	52.5 hours	25	200 μs	NS
Buiatti et al., 2019	60 hours	10	50 s	44
Hoehl et al., 2008	108 days	15	3 s	NS
Parise and Csibra, 2012	277 days	28	2 s	21
Pomiechowska and Csibra, 2022[1]	369 days	16	10 s	7
Pomiechowska and Csibra, 2022[2]	375 days	16	12 s	8
Pomiechowska and Csibra, 2022[3]	368 days	16	10 s	3
Pomiechowska and Csibra, 2022[4]	375 days	16	10 s	10
Percio et al., 2022	69.6 years	145	3-5 min.	NS

Table 1.1: Descriptive Statistics in EEG-based Cognitive Studies. For newborns and young infants, the stimulus duration of each experimental trial is shorter compared to a randomly chosen recent adult study (last row). The number of participants is significantly higher in the adult study compared to newborn/infant studies. Moreover, the unusable EEG files due to poor data quality and inattention of participants are higher in developmental EEG compared to the adult study. NS = Not Specified.

Most infant studies are based on the Event-Related Paradigm (ERP), in which the neural responses to specific events are extracted from EEG. According to the presented stimuli and the experimental objectives, relevant ERP components (e.g., N170 is a face-specific

ERP component; P300 represents higher cognitive functions such as working memory and decision-making) are extracted for concluding the experimental results. Since ERP data segments also contain the background EEG activities, the stimulus-specific trials are usually averaged to remove them. The background EEG noise decays approximately as $1/\sqrt{N}$, where N is the number of trials. Therefore, the higher the number of trials, the better the EEG background cancellation. However, acquiring many trials is not always practically feasible in developmental EEG studies.

Recently, a novel EEG paradigm called Frequency-tagging (FT) was introduced to newborns' EEG (Buiatti et al., 2019). A visual or auditory stimulus is presented periodically at a specific temporal frequency in the FT design. The resultant neural response is observed as a sharp peak in the EEG power spectrum at the same frequency. Since the response is narrowband, and the background EEG and artifacts are broadband in frequency, the FT response can be easily discriminated from the background brain activity. This implies that a higher Signal-to-Noise Ratio (SNR) is possible even with a short amount of data compared to the one obtained with ERPs. Therefore, the challenge of extracting a reliable stimulus-related neural response from shorter data is partially addressed.

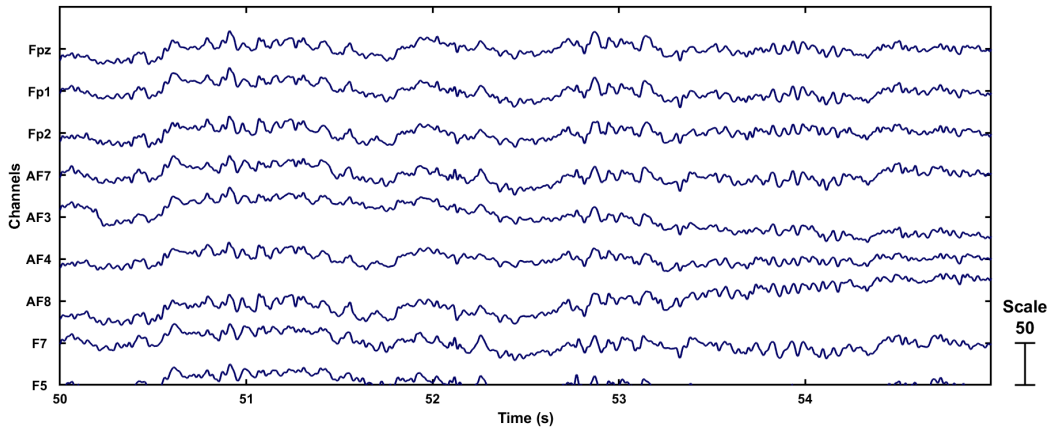
1.2.2 Artifacts Removal

Newborns and infants cannot follow instructions to refrain from making movements during data acquisition. Therefore, the developmental EEG data have the highest levels of artifact contamination. The most frequent artifacts are caused by various movements (head, arms, frowning, sucking), which generate non-stereotyped artifacts that constantly vary in topography and temporal dynamics (see Figure 1.1 for a comparison between a typical adult and newborn EEG segment). For this reason, quite often, developmental researchers exclude a certain number of collected EEG files as they contain more noise than stimulus-specific neural information (see Table 1.1 (last column)).

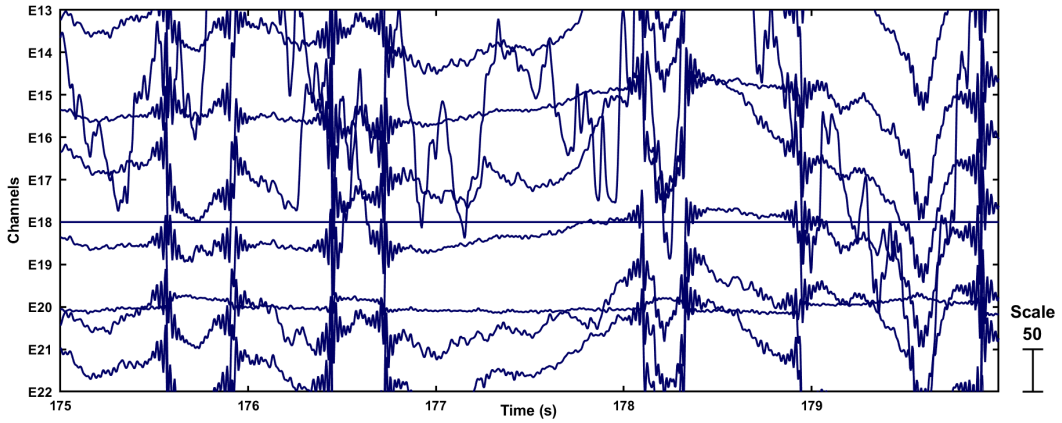
A widely used algorithm for artifact removal is Independent Component Analysis (ICA), which efficiently removes stereotypical artifacts (Makeig et al., 1995). While ICA is an automated artifacts removal solution for adults and older infants (Leach et al., 2020; Mognon et al., 2011), considering the dominant presence of non-stereotypical artifacts in newborns, ICA is unreliable. Further, ICA requires a considerable amount of samples for a reliable decomposition (Onton et al., 2006), which is generally not possible in newborn EEG (as discussed in the above Section 1.2.1). For these reasons, processing newborns' EEG for artifact removal is an arbitrary and time-consuming task.

1.2.3 Need for Standardization

One might motivate the need for standardisation in processing newborn data by considering two pillars for reliable science: Reproducibility and Replicability. *Reproducibility* involves



(a) Adult EEG.



(b) Newborn EEG.

Figure 1.1: Comparison of a sample adult and newborn EEG.

sharing data and the relevant code for analysis such that other researchers can reproduce the same scientific results obtained by the original authors. As a scientific community, we can be confident that our knowledge base is built on the contributions of several research groups. By demonstrating reproducibility, we show integrity, build trust and encourage reuse. *Replicability* is finding similar results by repeating a previously established study on a new data set. By proving that the results are consistent for a given experimental hypothesis, researchers build confidence in the scientific merit of the results. In sum, reproducibility focuses on replicating the same results using the same data and encourages transparency; replicability focuses on proving the generalizability robustness of scientific outcomes on newly collected data.

As we can imagine, the subjective nature of analysing the data leads to important scientific issues. Firstly, the results obtained through manual data correction are sometimes not reproducible by other researchers. Second, since the thresholds (derived by visual inspection)

applied to each dataset may be too specific for the considered data, using the same thresholds on new datasets might not lead to replicable outcomes. This has been quite an issue, especially in cognitive science/psychology, including the developmental community (Klapwijk et al., 2021). This thesis aims to address this challenge by proposing a standardised, objective and automated pipeline for preprocessing newborn/infant EEG. This promises both replicability and reproducibility, thereby ensuring reliable scientific outcomes.

1.2.4 Mobile EEG for Developmental Cognitive Neuroscience

Neurocognitive Developmental Disorders such as ASD are known to impact the brain's cognition and perception across the lifespan. As motivated earlier, identifying sensitive and specific neural-based biomarkers of atypical brain developments could help in earlier diagnosis and more efficient intervention. This could mitigate the life-long challenges typically associated with these disorders.

Understanding developmental neural signatures of typical and atypical brains require periodical, longitudinal, large-scale EEG studies. Such studies can be interfered with for several reasons, including the absence of some participants for some sessions. Thanks to the high versatility and flexibility of modern EEG systems and tremendous improvement in signal processing strategies, it is nowadays possible to record high-quality EEG data outside the clinical/research settings (for example, newborns' homes or children's schools). However, most of the EEG devices available on the market do not have adapted shapes for newborns' heads despite an emerging interest from the industry to develop custom EEG devices designed for neonatal populations. This motivates the need for developing custom EEG devices explicitly designed for neurodevelopmental populations.

Further, given the short attention span of newborns, the mobile EEG devices should be tailored to optimal experimental paradigms (e.g., FT). Depending on the experimental paradigm (e.g., visual or auditory), the location of the electrodes should be decided. Dry and soft electrodes can quicken the setup time and provide comfort to the sensitive skin of newborns. Remarkably, such devices must ensure a long recording time such that several successful trials (where success implies appropriate attention of newborns) can be acquired once the EEG cap is worn. Given all these constraints, novel signal processing techniques might be required as traditional methods that work in high-density EEG and in adults might be sub-optimal or unsuitable.

1.3 Contributions

As mentioned earlier, this thesis work attempts to address the challenges in newborn EEG studies, namely, artifacts removal and mobile EEG solutions, with much more emphasis on the former than the latter.

1.3.1 Novel Methods for Artifacts Removal in High-Density EEG

Traditionally, artifacts are removed in two subsequent steps. First, bad channels are detected and removed. Second, bad portions of the data are rejected or corrected. My preliminary analysis proved that the existing approaches (developed for adult or older infant EEG) are over-sensitive and less specific in identifying artifacts (therefore, high false positive rate) in newborn EEG. These results motivated the need for novel approaches deliberately designed or adapted for newborn EEG characteristics (shorter data + heavier artifacts contamination). Towards this end, I introduced, for the first time, the Local Outlier Factor (LOF), a robust outlier detection algorithm to the context of EEG for bad channel detection problems (Kumaravel, Buiatti, et al., 2022). The algorithm is described in Section 2.3.1, and a robust technique for finding optimal user-defined parameter is presented in Section 2.5.1.1. LOF is evaluated in newborn, infant and adult EEG, proving its adaptability in different EEG settings. The validation results are presented in Section 2.5.1.2.

For the first time, I explored and successfully adapted Artifacts Subspace Reconstruction (ASR), a component-based artifacts reduction technique, to removing/correcting bad portions of newborn EEG data (Kumaravel, Farella, et al., 2022). The algorithm is described in detail in Section 2.2.2. ASR is primarily intended for mobile adult EEG data that usually contain severe motion artifact contamination, which is conceptually similar to artifacts present in newborn EEG. I implemented a data-oriented automatic method to calibrate crucial user-defined parameters such as ASR parameter k and ASR processing mode (Correction or Rejection). Finally, with the calibrated parameters, I validated the adapted version of ASR on newborn and infant EEG datasets, and the results are presented in Section 2.5.2.2.

LOF and ASR algorithms are integrated with other preprocessing functionalities, such as data import, filtering, etc., to compose a fully automated pipeline for newborn/infant EEG preprocessing. Chapter 3 is dedicated to presenting the details of the proposed pipeline (NEAR: Newborn EEG Artifact Removal) and its validation results on simulated and real EEG datasets.

1.3.2 Portable Low-Density Hardware-Software Design for Newborns EEG

Once proposed artifact removal methods for high-density EEG data are proven successful, we attempt to implement them in a portable low-density EEG device. This part of the thesis is executed in collaboration with the University of Bologna as a part of the European ERC Proof-of-Concept Grant NeuroSoNew (https://cordis.europa.eu/project/id/842243/it). We first identified the BioWolf platform (Kartsch et al., 2019) as an optimal hardware solution for developmental populations because of its high versatility and quality of acquired signals using dry and skin-friendly electrodes. Then, we acquire data from adults with clean and artificially contaminated trials using the FT paradigm to validate artifact removal techniques. The hardware architecture of BioWolf, including the choice of electrodes, is described in Section 4.1. In addition to the disadvantages of ICA mentioned earlier, ICA is also unreliable for low-density systems such as BioWolf. A possible solution to clean artifacts in such systems

is ASR. For effective artifact removal using ASR, a standard EEG system with at least 20 channels is (empirically) recommended (Chang et al., 2018). For a systematic evaluation, I explored ASR for the first time on data collected using an ultra-low density portable and wireless EEG setup (8 electrodes, (Kumaravel, Kartsch, et al., 2021)). Further, I calibrated the ASR parameters (ASR cut-off parameter k, and processing mode) on this dataset. The validation results and optimal ASR user-defined parameters are presented in Section 4.2.3. In parallel, we also proposed preliminary NN-based solutions for EEG bad channel detection (Kumaravel, Paissan, and Farella, 2021; Paissan et al., 2022). The network architectures and the results are presented in Section 4.3.

While it has been found that ASR is a reliable artifact removal even if the number of electrodes is far less (< 32 electrodes in which ASR was previously validated), ASR is a computationally intensive algorithm to be implemented in embedded hardware such as BioWolf. As our initial efforts in simplifying the model did not yield promising results, we focused on developing a novel feature extraction algorithm to detect the stimulus-specific neural response automatically. Importantly, we aimed to develop an algorithm that is robust to noise and requires minimal preprocessing (Kartsch et al., 2022). Our ideal choice was Canonical Correlation Analysis (CCA), the current state-of-the-art algorithm for online frequency-tagging BCI applications (Hakvoort et al., 2011). We propose a novel frequency-normalized CCA-based index to correctly and rapidly identify the tagged response without any artifacts removal or channel selection. The validation results on data acquired using BioWolf are presented in Section 4.4.6. Indeed, much progress is needed to develop a fully online portable setup suitable for investigating cognition and perception in human newborns and young infants. The conclusions chapter discusses some of the open challenges and possible solutions.

2 Artifact Removal Methods for Developmental EEG

Generally, artifact preprocessing in EEG comprises two main steps: bad channel detection/removal and bad segment rejection or correction. First, we review existing algorithms and we discuss why they do not produce optimal results when applied to newborn/infant EEG. Then, we introduce novel methods for bad channel detection (LOF) and bad segment detection (Adapted ASR) specifically designed to clean artifacts in developmental EEG. Evaluation of the LOF algorithm for bad channel detection is straightforward - it requires bad channel labels annotated by EEG experts (ground truth). The list of detected/predicted bad channels can then be compared against these ground truth labels. However, validating ASR is tricky because it is uncommon to find datasets with annotated bad segments of data. Therefore, indirect yet efficient ways of validating ASR include analysing the quality of data (in terms of SNR) and comparing the statistical effect of recovered neural responses with that of obtained in the original paper after ASR preprocessing. In this chapter, we describe the complete validation of LOF on newborn, infant, and adult EEG. Instead, we present qualitative results achieved using ASR cleaning. Detailed validation including the statistical evaluation of experimental effects is presented in the next chapter, where we will compare a preprocessing pipeline based on LOF and Adapted ASR with the state-of-the-art preprocessing pipelines validated on developmental EEG data.

This chapter is organised as follows: Sections 2.1, and 2.2 present an overview of bad channel and segments removal methods along with their implementation details, respectively. Then, in the section 2.3, we introduce the proposed algorithms explicitly designed or adapted for newborn EEG. The experimental datasets and validation results are presented in sections 2.4, and 2.5 respectively. The materials used in this chapter are published in (Kumaravel, Farella, et al., 2022) and (Kumaravel, Buiatti, et al., 2022), under a CC license.

2.1 Overview of Bad Channel Removal Methods

EEG channels/sensors that have a poor signal-to-noise ratio (SNR) due to biological or technical artifacts contaminating a larger portion of the recording are commonly termed "bad

channels". In a broad sense, EEG bad channel detection is an anomaly detection problem. It is the process of finding records that significantly deviate from the regular or usual data. In most cases, the total number of anomalies is lower than the regular ones in a given dataset. Bad channel detection is crucial in removing artifacts for the following reasons:

(i) Removing noisy segments of EEG in the presence of bad channels can lead to severe data loss due to a misleading overall rejection threshold.

(ii) The presence of bad channels can produce a strong bias on the overall statistics of the extracted neural features leading to the wrong interpretation of the experiments.

(iii) Further, bad channels can also bias the source level analysis as they often suppress the information from the adjacent good channels, resulting in a wrong source reconstruction.

Here, we discuss the widely used bad channel detection methods in the literature.

2.1.1 Kurtosis

Kurtosis is a higher-order statistical measure that reflects the Gaussianity of a distribution. Positive kurtosis indicates a super-Gaussian distribution, while negative kurtosis denotes a sub-Gaussian distribution. Despite being a simple measure, it has been widely used as a reliable feature for several artifact removal methods in EEG (Delorme et al., 2007; Greco et al., 2006; Mahajan and Morshed, 2015). For the implementation, we used the EEGLAB function *pop_rejspec* to detect bad channels with default parameter settings. In particular, the kurtosis values computed for each channel were normalised to have zero mean and unit standard deviation (i.e. using Z-score). Channels with a Z-score of more than 5 were identified as bad channels.

2.1.2 Fully Automated Statistical Thresholding for EEG artifact Rejection (FASTER)

FASTER is a fully automated pipeline that transforms raw EEG data into processed files for further analysis, with artifact rejection steps implemented at three levels (channels, segments, independent components) (Nolan et al., 2010). FASTER detects bad channels by first computing the temporal correlation between channels, followed by their variance and a score based on the Hurst exponent. Usually, heavily contaminated channels tend to demonstrate poor correlation with their closest neighbours. Sometimes, a group of noisy channels can exhibit a higher correlation among themselves. To capture them, the second metric, variance, is used. The final measure Hurst exponent explains the biological phenomenon (a value of at least 0.5 indicates the presence of time-series components in the data). Once all these measures are computed, FASTER applies a Z-score threshold of 3 to detect the list of bad channels.

2.1.3 Clean_RawData EEGLAB Plugin (CRD)

The default EEGLAB function *clean_rawdata* (CRD, https://github.com/sccn/clean_rawdata) is adopted from BCILAB (Kothe and Makeig, 2013; Mullen et al., 2015). It contains several methods to detect bad channels; in particular, it deals with flat-line channels, channels contaminated with high-frequency noise and channels uncorrelated with their neighbours. CRD considers a given channel as bad if either one of the following conditions is met: i) A channel recorded a flat-line (no signal) for at least 5 s; ii) A channel that contains predominant electrical power-line interference (50 or 60 Hz) more than 4 noise standard deviations; iii) A channel that is poorly correlated with its neighbours (correlation threshold < 0.8).

2.1.4 Harvard Bad Channel Detection (HAPPE)

A preprocessing step within the Harvard Automated Processing Pipeline for EEG (HAPPE; Gabard-Durnam et al., 2018) finds bad channels by computing the joint probability of the average logarithmic power between 1 and 125 Hz across channels (Gabard-Durnam et al., 2018). Implementation-wise, HAPPE bad channel detection uses EEGLAB's function *pop_rejchan* (Delorme and Makeig, 2004) but with a change in Z-score threshold (= 3, instead of the original threshold of 5). Since HAPPE is designed for highly contaminated infant datasets, this process is run twice on the same file to ensure complete channel artifact removal.

2.1.5 Drawbacks of Existing Bad Channel Detection Approaches

The aforementioned bad channel detection methods and their corresponding thresholds are initially developed for adult EEG. Applying them with their default rejection threshold to a noisier infant or newborn data will more likely reject good channels along with bad ones (leading to a higher false positive rate), leading to unfair data analysis. This is because the default thresholds designed for relatively cleaner adult data are generally too strict for developmental data. This is because the statistical structure of the data is generally different. Also, in some cases, the measures used on infants are more robust to artifacts than the ones used to calibrate the thresholds on adults. Indeed, in HAPPE, where the bad channel detection method *pop_rejchan* (developed for adult EEG) is adapted to infant EEG, the authors reduced the threshold from 5 to 3 and incorporated a two-stage thresholding approach to avoid unnecessary data removal. Reducing the detection thresholds is a possible way to adapt infant/newborn data methods. But, there is a risk of missing out on bad channels (leading to a higher misdetection rate) in comparatively less noisy datasets. More importantly, all these methods assume a normal distribution for EEG data and therefore impose thresholds based on distribution statistics such as mean, standard deviation, or Z-score. First, this assumption can no longer hold given the short amount of newborn data with high-level noise contamination. Secondly, as we know, such distribution-based thresholds are sensitive to the outliers in the data; hence, they are less resilient to outliers. Given these, we might all agree that a novel bad channel technique is required for identifying bad channels in newborn and infant EEG data.

Notably, this method should not assume an underlying data distribution, and the derived outlier measure should be robust against outliers.

2.2 Overview of Bad Segment Rejection/Correction Methods

The most common approach to deal with artifacts in adult data is to

(i) identify and remove paroxysmally artifacted segments by visual inspection of the raw data;

(ii) correct the remaining data by removing the residual artifacts with Independent Component Analysis (ICA; Delorme and Makeig, 2004; Makeig et al., 1995), assuming that those artifacts are spatiotemporally stereotypical.

However, ICA may not deal with some types of artifacts, especially non-biological artifacts that are difficult to model and caused by head/cable movements (Chang et al., 2018) commonly present in newborn/infant EEG. To deal with such artifacts, a recently emerging successful technique is Artifacts Subspace Reconstruction (ASR). ASR is an unsupervised component-based algorithm to remove high-amplitude, transient samples that may manifest in EEG due to the contamination of artifacts of any kind. Further, ASR is an online cleaning algorithm validated on high-density mobile EEG (Mullen et al., 2015), whereas, ICA is a computationally demanding algorithm primarily designed for offline use.

In this section, we will review both algorithms in detail.

2.2.1 Independent Component Analysis (ICA)

ICA is one of the Second-Order Blind Identification (SOBI) approaches (Bell and Sejnowski, 1995; Lee et al., 1999), which considers relationships between multiple time points to identify maximally independent source distributions. ICA was successfully applied to multi-channel EEG for noise reduction for the first time in 1995 (Makeig et al., 1995) with the assumption that EEG may be plausibly modelled as a linear mixture of the activities of multiple brain (and non-brain such as eyes) sources with (quasi-) independent time courses. Therefore, the obtained independent components can be further classified as "brain" or "non-brain" depending on their electrical activities. The implementation of ICA for EEG data is available within EEGLAB (Delorme and Makeig, 2004) and MNE-Python (Gramfort, 2013) platforms.

Once sufficiently high-pass filtered (see Onton et al., 2006 for filtering recommendations) EEG data of n channels and t samples are fed into the ICA algorithm, it performs a blind separation of the data (A) based only on the criterion that the resulting source time courses (B) are maximally independent. To be specific, like other component-based methods, ICA finds an 'unmixing' component (W) that, when multiplied by the original data (A), yields the independent components (B):

$$B = WA \tag{2.1}$$

where A and B are $c \times t$ matrices that represent input EEG and independent component activated EEG, respectively, and W is $c \times c$ representing the 'mixing' matrix whose columns contain the relative weights of the component projects to each of the channels c. By simple linear algebra, the equation 2.1 implies that

$$A = W^{-1}B \tag{2.2}$$

The portion of the input EEG data (A) that generates the ith independent component (IC) is the outer product of two vectors, the ith column of W and the ith row of B:

$$A_i = {W_i}^{-1}B_i \tag{2.3}$$

The original data (A) can be retrieved back by summing the back-projected ICs (B_i) as follows:

$$X = \sum X_i \qquad \text{where i = 1, 2,...n} \tag{2.4}$$

In theory, each IC is representative of a specific signal generator producing electric fields with a stable spatial projection pattern across the recording channels. However, perfect separation of source signals is only sometimes possible in practice, and it is difficult to verify without simultaneous invasive recordings. Traditionally, the independent components ICs are classified as "brain" and "artifactual" by visually inspecting the temporal and spatial patterns of the decomposed IC activations. Thanks to the recent efforts in the community, there have been a few successful automated methodologies to solve the IC classification problem. Here, we will have a brief overview of these techniques.

1. **FASTER:** As we saw earlier, FASTER is a fully-automated preprocessing pipeline that cleans raw EEG data. Within the FASTER pipeline, there exists an implementation of IC labelling, which considers a given IC as "artifactual" if any of the computed features deviate from the dataset average by more than 3 standard deviations (Nolan et al., 2010).

2. **ADJUST: A**utomatic EEG artifact **D**etection based on the **J**oint **U**se of **S**patial and **T**emporal features (ADJUST) computes stereotyped artifact-specific spatial and tempo-

ral features on the decomposed ICs (Mognon et al., 2011). These features are optimised to capture the most common artifacts in the data, such as blinks, eye movements, and generic discontinuities. Data-centric automatic thresholding is derived using the Expectation-Maximization technique for each feature, and bad components are detected.

3. **Adjusted-ADJUST:** Since ADJUST is configured for typical adult EEG electrode montage configuration (a standard international 10-20 system), a modification was required to make it suitable for high-density EEG setup (geodesic nets) that is commonly used for pediatric data. The adjusted-ADJUST classification method is a part of the MADE pipeline, which we will see in the section 3.1.2. The changes with respect to the original methodology ADJUST are related to features for detecting blinks and horizontal eye movements (Leach et al., 2020).

4. **MARA: M**ultiple **A**rtifact **R**ejection **A**lgorithm (MARA) is a machine-learning algorithm that classifies the ICA-derived components into the brain and artifactual groups (Winkler et al., 2014). MARA exploits six features based on spatiotemporal and spectral information to assign artifact probability to each IC. Since MARA has been trained on manual component rejections, it aims to capture a wide range of artifact categories, whereas ADJUST focuses on specific artifacts like eye movement and signal discontinuities. Even though MARA was trained on adult data, several features of MARA support its application to infant data (such as MARA computes spatial features for a 10-20 electrode system, which is highly consistent across adult and infant brains (Kabdebon et al., 2014)). Therefore, MARA was used in HAPPE preprocessing pipeline (which we will see in the section 3.1.1).

5. **ICLabel:** ICLabel is a recent machine learning-based technique for IC component classification. ICLabel learns based on a crowd-sourced labelling (https://labeling.ucsd.edu/tutorial), where researchers across the globe can visualise the IC spatiotemporal and spectral patterns and categorise them as one of seven different classes. To provide a generalised solution, the training data contains EEG acquired using a multitude of paradigms. The pretrained model is open-source software within the EEGLAB environment (Pion-Tonachini et al., 2019).

2.2.2 Artifacts Subspace Reconstruction (ASR)

ASR is an adaptive spatial filtering technique to automatically detect/remove transient high-amplitude artifacts in continuous EEG data (Kothe and Jung, Google Patent WO2015047462A9, Jun. 2014). The approach is loosely connected to Principal Component Analysis (PCA), a popular dimensionality-reduction technique. Initially, ASR is proposed as an online artifacts removal method in the BCILAB software (Kothe and Makeig, 2013), but it is also available as an EEGLAB plug-in function (*clean_rawdata*) for offline processing. ASR was validated extensively on simulated data, and real EEG was acquired using a mobile setup from adult

participants (Kumaravel, Kartsch, et al., 2021; Mullen et al., 2015). Thanks to its efficient artifact removal, ASR is now considered one of the defaults in preprocessing algorithms within the EEGLAB framework. Here, we will see how ASR processes artifacts:

(i) ASR requires calibration data (X_c), which is a portion of original data (X), ideally free from artifacts. In cases where it is not possible (for example, from clinically-ill patients or newborns), ASR automatically finds cleaner portions of data (i.e., calibration data) by fitting input data chunks into a predefined Gaussian distribution representative of EEG-like data. The data chunks with a tolerable fitting error are concatenated to form the calibration data.

(ii) ASR applies PCA on the obtained calibrated data (X_c) to decompose it into principal components, each explaining a certain level of variance in the time series data.

(iii) A spatially varying threshold is then computed using the mean and standard deviation of the individual principal components as follows:

$$T_i = \mu_i + k.\sigma_i \tag{2.5}$$

where T_i the threshold, μ_i is the mean and σ_i is the standard deviation of the component i, respectively and k is a user-defined parameter also known as ASR cut-off parameter.

(iv) Since the threshold T is computed from the calibrated data; it is representative of artifact-free EEG data. As such, a subspace of artifact components in a given chunk of data from X is removed if its variance exceeds T.

(v) Finally, ASR imputes each removed component with a linear combination of activity of the remaining non-artifact components and back-projects the signal from the component space to channel space.

2.2.3 Drawbacks of Existing Bad Segment Rejection Approaches

All ICA-based methods might not apply to newborn or infant EEG for the following reasons: 1) ICA is suitable for detecting the artifacts/noise components that are stereotypical (and easy to model), such as eye blinks or eye movements and heart activities. 2) There is an empirical limitation of applying ICA to short-duration EEG: ICA demands more data for high-density EEG setup – which is the norm for the developmental population. For example, EEG acquired using a 256-channel setup with a sampling rate of 256 Hz, ICA requires at least 85 minutes of recording for a reliable decomposition (Onton et al., 2006). Since developmental EEG data are short and contaminated with non-stereotypical artifacts that are difficult to model, ICA is not an ideal solution for these data.

ASR is a real-time artifacts cleaning algorithm for mobile EEG acquisitions recorded outside laboratory conditions. As such, it deals with non-stereotypical artifacts and aims to remove transient high-amplitude artifacts even in a small segment of data (i.e., 1 s). Even if both features of ASR fit very well for developmental EEG, it is unexplored in developmental data to prove its effectiveness. Further, it needs to be made clear which are the optimal user-defined parameters adapted for developmental EEG. To answer these questions, an extensive evaluation is required.

2.3 Proposed Methods for Newborn/Infant EEG

This section presents a novel method for bad channel detection and an adapted version of an existing algorithm for bad segment correction or rejection.

2.3.1 Local Outlier Factor (LOF)

LOF is a local density-based outlier detection approach (Breunig et al., 2000) that measures the degree of isolation of a given EEG channel with respect to its local neighbourhood (where the neighbourhood is defined using the k-neighbours algorithm (Fix and Hodges, 1989) computed from the activity vectors associated with each channel). As a result, LOF assigns an outlier score for each channel by computing its local density, where the k-neighbours algorithm defines locality. Thanks to this property, LOF is a robust technique compared to traditional methods that employ global measures of uncertainty.

2.3.1.1 LOF Algorithm

The LOF algorithm quantifies the outlierness of each electrode in the multidimensional activity space where each electrode is associated with a vector representing its EEG activity (not to be confounded with its physical location on the scalp). The algorithm is described as follows:

1. The optimal k value (i.e., the number of nearest neighbours) is first computed using the Natural Neighbors algorithm (NaN; Q. Zhu et al., 2016), a data-centric non-parametric approach.

2. For a given channel p, the LOF algorithm identifies k neighbour channels based on the predefined distance metric (e.g., Euclidean) using the k-nearest neighbours algorithm (Fix and Hodges, 1989).

3. Then, a reachability distance is computed between channels. For example, let us consider two channels, namely p and o. The reachability distance between p and o is computed as follows:

$$\text{reach-dist}_{k}(p,o) = \max\{k\text{-distance}(o), d(p,o)\} \tag{2.6}$$

where k-distance (o) is computed using the *knnsearch* function (MATLAB, 2018) and $d(p,o)$ is the Euclidean distance between two channel vectors. Intuitively, if channel p is far from o, the reachability distance is their actual Euclidean distance. Instead, if they are sufficiently close, the Euclidean distance is replaced by the k-distance of channel o (see Figure 2.1). Considering the k-distance rather than the actual distance reduces the statistical fluctuations for the points existing within the k neighbourhood.

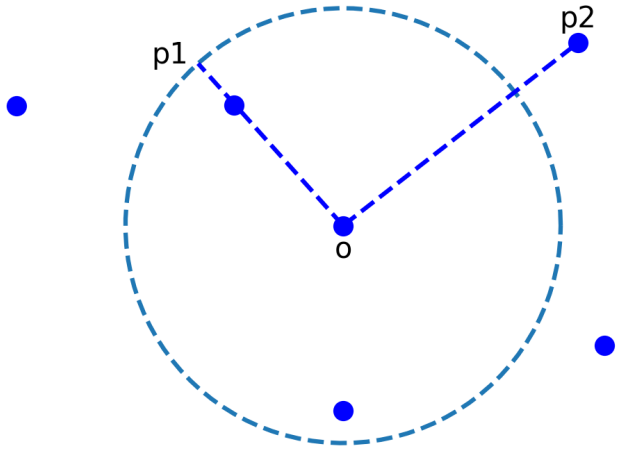


Figure 2.1: An example scenario for the computation of reachability distance using $k = 3$. The dotted circle represents k neighbourhood of point o. Blue points represent the data samples. For the demonstration, let us consider two points $p1$ (that exists within the k neighbourhood) and $p2$ (that exists outside the k neighbourhood). The reachability distance between the point $p1$ and o will be the k-distance (*knnsearch*, MATLAB, 2018), whereas the reachability distance between the point $p2$ and o will be the Euclidean distance between them.

4. Once the reachability distance of each channel with respect to its neighbours is computed, then the local reachability density (LRD) is determined as follows:

$$\text{LRD}_{\text{k}}(p) = 1/\left(\frac{\sum_{o\in N_{\text{k}}(p)} \text{reach-dist}_{\text{k}}(p,o)}{\left|N_{\text{k}}(p)\right|}\right) \tag{2.7}$$

where $\text{N}_{k}(p)$ refers to the total number of k neighbors of p. To put it in words, the LRD of the channel p is the inverse of the average reachability distance based on the k-nearest

neighbours of p. Intuitively, channel p will have a lower LRD if it were an outlier (i.e., bad) channel because it is not easily "reachable" by most of its neighbours.

5. As a final step, the local outlier factor (LOF) is computed as follows:

$$\mathrm{LOF}_{\mathrm{k}}(p) = \frac{\sum_{o \in N_{\mathrm{k}}(p)} \frac{\mathrm{LRD}_{\mathrm{k}}(o)}{\mathrm{LRD}_{\mathrm{k}}(p)}}{\left|N_{\mathrm{k}}(p)\right|} \tag{2.8}$$

The LOF of channel p is the ratio of the average LRD of k neighbours of p to the LRD of p. The lower p's LRD is, and the higher the LRD of p's k-nearest neighbours are, the higher the LOF value of p is (and, therefore, possibly an outlier). In other words, an outlier channel would display a lower LRD (thus, widely separated in the distance) than its neighbours (on average). If channel p has a similar LRD value compared to its k neighbours, the LOF score would be approximately 1.

2.3.1.2 LOF Threshold Computation

In an ideal scenario where the objects (or samples) form a uniform or a Gaussian cluster, inliers would yield LOFs approximately equal to 1, as can be inferred from Equation (2.8). As such, any object (or sample) that exceeds a LOF score of 1 can be considered an outlier. However, this criterion might vary in real-world data, where the distribution of objects is unknown and less likely to be uniform or Gaussian. A thorough investigation of the decision boundary is required as there are different EEG settings (populations, experimental paradigms, and so on) and the definition of outliers varies according to the settings. Therefore, it is necessary to consider the LOF Threshold ($\mathrm{LOF}_{\mathrm{thr}}$) as a hyperparameter to be optimised using the conventional supervised methods such as train/test split or k-fold cross-validation.

2.3.2 Adapted ASR

As mentioned in the Section 2.2.3, while ASR seems a promising technique to deal with non-stereotypical artifacts manifested due to movements, it is still unexplored in developmental EEG. For the first time, this thesis work attempted to evaluate the sensitivity and specificity of artifact removal processing using ASR on EEG acquired from developmental populations. The bad channel detection method (CRD, described in the Section 2.1.3) suggested by EEGLAB as a preliminary step to ASR processing removes nearly half of the channels in newborn EEG. Therefore, in this work, CRD bad channel detection is replaced by the proposed LOF algorithm. Further, an empirical analysis showed that the predefined (or default) ASR parameters (that work well for adult EEG) are too aggressive for this population. Therefore, this work is focused on calibrating ASR parameters before a systematic validation. There are two crucial user-defined parameters of ASR, which are described as follows:

2.3.2.1 ASR Cut-off Parameter (k)

ASR defines an upper-bound threshold for a PC representing EEG-like components based on the mean and variance of PCs extracted from the cleaner portion of the data (see Equation (2.5)). Therefore, the components exceeding this threshold are most likely artifactual. From the Equation (2.5), it can be observed that a lower k implies a lower threshold and, therefore, a strict artefact detection (i.e. more artifacts are detected); a higher k indicates a looser cleaning of the data (i.e. less artifacts are detected). The optimal k values for adult EEG lie between 20 and 30 (Chang et al., 2018). However, the ASR parameter k has never been studied on developmental data.

2.3.2.2 Processing Mode

Using the *clean_rawdata* plugin, ASR can be operated in two distinct modes: ASR Correction (hereafter indicated as ASR_C throughout this thesis), in which the bad portions of the data are corrected to 'EEG-like' data, and ASR Removal (indicated as ASR_R) in which the detected bad portions are removed from the data. From a theoretical standpoint, ASR_C is a preferred mode for short newborn EEG data such that the original length of data remains unchanged. On the other hand, ASR_R removes a certain number of artifactual samples depending on the quality of the data and the ASR cut-off parameter k. This might not be favourable, given that newborn EEG is already short and cannot undergo further removal of data. However, a systematic evaluation is required to study the differences in ASR modes in terms of the quality of ASR-cleaned data.

2.4 Experimental Datasets

This section describes the datasets considered for calibrating and validating the proposed methods.

2.4.1 Newborn EEG

Newborn training and test datasets belong to two independent studies performed at the Neonatal Neuroimaging Unit (CIMeC, University of Trento) installed in the maternity ward of Rovereto Hospital "Santa Maria del Carmine" (Rovereto, Italy). These studies were conducted at different periods and for different purposes; as such, no single newborn participated in both study. The local ethical committee approved both studies for clinical research (Comitato Etico per le Sperimentazioni Cliniche, Azienda Provinciale Servizi Sanitari, Province of Trento, Italy); parents were informed about the content and goal of the study and gave their written informed consent.

Both datasets were recorded by an EGI EEG system (GES400, Electrical Geodesic, Inc, Eugene, OR, USA) with 125 channels. Scalp voltages were referenced to the vertex, amplified, and

digitised at 250 Hz. Electrode impedances were kept below 100 kΩ. Newborns were tested in a calm, dimly illuminated space in the maternity ward, seated on the lap of a trained researcher in front of a 60 cm x 33.8 cm LCD screen (distance eyes-screen: about 30 cm) while wearing the EEG cap. Video recording from a hidden camera on the top of the screen provided online monitoring of the infant. The newborn's parents, when present, were off the sight of the infant (separated by a curtain) and instructed to keep silent during the recordings. For both datasets, visual stimuli were presented dynamically with sinusoidal contrast modulation (the visibility of each stimulus gradually rises with respect to the grey background from 0% at the beginning of the cycle to 100% at mid-cycle, then gradually decreases to 0% towards the end of the cycle (see Figure 1 in (Buiatti et al., 2019)), at a rate of 0.8 Hz (frequency-tagging paradigm). We used sinusoidal contrast modulation instead of squared on–off dynamics to minimise nonlinear effects in the brain frequency response (Norcia et al., 2015) and make the stimulation more pleasant for babies (de Heering and Rossion, 2015). The slow presentation rate (0.8 Hz) was chosen to ensure newborns fully perceived the stimulus at each cycle of the periodic, peekaboo-like presentation.

The training dataset for ASR parameter calibration is part of an ongoing study investigating the neural bases of number perception in newborns (Buiatti et al., in preparation). Visual stimuli consisted of a set of 4 or 12 coloured simple geometrical shapes, presented in blocks of 50 s or until the subject stopped attending to them; shape, number, and spatial arrangement were constant within each block and randomly changed between blocks. For the whole duration of the study, an auditory stimulation consisting of sequences of syllables was simultaneously presented (the response to the auditory stimulation will not be considered here). The training dataset includes all the subjects that attended at least 15 s of visual stimulation, independently whether they attended one or both number conditions (11 newborns, six males; mean age 40 ± 16 hours; all were healthy [APGAR(1 min) $\geq$ 8, APGAR(5 min) = 10 for all subjects] and born full-term (gestation age, 39.9 ± 0.9 weeks; weight, 3.26 ± 0.30 kg).

The test dataset belongs to a study investigating the cortical bases of facelike pattern processing (Buiatti et al., 2019). Visual stimuli consisted of a white head-shaped form containing three black squares. They differed only in the spatial configuration of the three squares to form the three stimuli (upright face, inverted face, and scrambled face). Stimuli were presented in blocks of 50 s or until the subject stopped attending them. Subjects were ten healthy newborns (six males; mean age 60 ± 22 hours). All were healthy [APGAR(1 min) $\geq$ 8, APGAR(5 min) = 10 for all subjects] and born full-term (gestation age, 39.7 ± 1.5 weeks; weight, 3.41 ± 0.28 kg). Further details in (Buiatti et al., 2019).

We applied a low-pass FIR filter for both datasets with a cut-off frequency of 40 Hz (using EEGLAB's default filter). Subsequently, we used a non-causal high pass filter between 0.15 and 0.3 Hz and a stop-band attenuation of 80 dB.

2.4.2 Infant EEG

Infant training and test datasets belong to a study investigating semantic understanding of common nouns in preverbal infants, performed at the Cognitive Development Center (CDC, Central European University) and whose results are published in (Parise and Csibra, 2012). The original study had 2 experimental conditions, namely, the Mother-Speech condition and the Experimenter-Speech condition. No single infant participated in both conditions simultaneously. Ethical approvals were obtained from the ethics committee of the Central European University, Budapest; parents were informed about the content and goal of the study and gave their written informed consent. All infants were born full term (gestational age: 37–41 weeks) in the normal weight range (> 2500 g).

The training and the test dataset included 14 healthy infants (Training: 6 females; mean age = 278 days, range = 266–285 days. Test: 5 females; mean age = 277 days, range = 269–286 days). Both datasets were acquired using an EGI amplifier (GES 300, Electrical Geodesic, Inc, Eugene, OR, USA) at a sampling rate of 500 Hz with a low-pass filter at 200 Hz. Continuous EEG was recorded by 125-channel Geodesic Sensor Nets referenced to the vertex. Infants were tested in a quiet, dimly illuminated room in the CDC BabyLab, sitting on a high chair 70 cm in front of a 9-inch, 800 × 600, 100 Hz CRT monitor. The infants were video-recorded from a hidden camera below the presentation monitor throughout the session. The infant's mother and an experimenter sat on chairs on either side of the infant.

For both datasets, each trial started with a live auditory stimulus delivered either by the experimenter (Training dataset) or the infant's mother (Test dataset). In contrast, a dynamic fixation stimulus (a colourful rectangle 343×363 pixels) was presented on top of an occluder. After the live auditory stimulus ended, the fixation stimulus stopped moving, and the display remained frozen for 600–800 ms. Then the fixation stimulus disappeared, and the occluder started to fall forward (a 90° rotation on the basis-hinge), revealing an object behind it (see Figure 1 in Parise and Csibra, 2012). The object, laying on a black background, was fully visible for 1000 ms before the occluder began to rise, hiding the object again. This was followed by an intertrial interval lasting 1100 to 1300 ms. The pictures of 15 different objects were used; their average size was 302.5 × 321.6 pixels. Trials were presented as long as the infants were attentive. The minimum inclusion criterion for the infants was at least 10 artifact-free trials in each of the two experimental conditions—further details in (Parise and Csibra, 2012).

Both datasets were band-pass filtered between 0.3 and 30 Hz using the default EEGLAB filter.

2.4.3 Adult EEG

We used the data from the study (Schneider, Pereira, Tonin, and del R. Millán, 2019) validating alpha-power lateralisation as feedback to enhance visual covert attention tasks. Fourteen subjects (seven male, seven female) with a mean age of 23 years (SD = 1.52 years) took part in the recordings on three different days, resulting in 130 EEG files (refer Schneider, Pereira,

Tonin, and del R. Millán, 2019 for more details related to the experimental setup). EEG was recorded with a 64-channel HIamp EEG system (g.tec, Austria) at a sampling rate of 512 Hz. The electrodes were positioned in the standard international 10-10 system. All datasets are available on the OpenNeuro platform (Schneider, Pereira, Tonin, and del R. Millan, 2019). Out of 130 files, only 113 were usable, and others were found corrupted due to import issues. Before applying the LOF algorithm, we filtered the data at 40 Hz to remove the high-frequency noise components and subsequently, a high-pass filter was used to remove DC drifts. The ground truth bad channels are labelled via visual inspection by authors of the original study and indicated as "bad" in the channel description for each EEG file on the OpenNeuro platform.

2.5 Experimental Evaluation

This section describes the experiments conducted to validate the efficiency of the abovementioned algorithms: LOF and ASR, on different datasets. To recap, LOF is evaluated for bad channel detection, and ASR is considered for bad segment rejection. For the bad channel detection part, a robust evaluation metric F1 Score (defined in the Section 2.5.1.1) is used to account for the class imbalance (number of bad channels >> number of good channels). As an evaluation metric, an SNR measure of the neural response (defined in Section 2.5.2.1) is used for the bad segment rejection.

As this thesis introduces LOF for EEG bad channel detection, it is validated in adult EEG, in addition to newborn and infant EEG. Instead, since ASR has been validated on adult EEG acquired using both standard and mobile EEG (Blum et al., 2019; Kumaravel, Kartsch, et al., 2021; Mullen et al., 2015), this thesis presents the first-time validation of ASR in developmental EEG. In particular, this section presents the preliminary results achieved using newborn EEG. As a later development, extensive validation of ASR with statistical analysis on both newborn and infant EEG will be presented in the next chapter.

2.5.1 Bad Channel Detection using LOF

2.5.1.1 LOF Threshold Calibration

Before detecting bad channels, it is essential to calibrate the rejection threshold as described in the Section 2.3.1.2. Therefore, we consider the LOF_{thr} as a hyperparameter to be optimised using the supervised approach (i.e., with bad channels annotated by the respective authors as the true labels). Precisely, we utilised the k-fold cross-validation technique (Pedregosa et al., 2011) to systematically identify the optimal LOF_{thr} (exhaustive search in the range between 1 and 5, in steps of 0.1) at which the F1 Score is maximised. F1 Score is defined as follows:

$$F_1 = \frac{2 * TP}{2 * TP + FP + FN} \tag{2.9}$$

where TP, TN, FP, and FN indicate the number of true positives, true negatives, false positives, and false negatives, respectively.

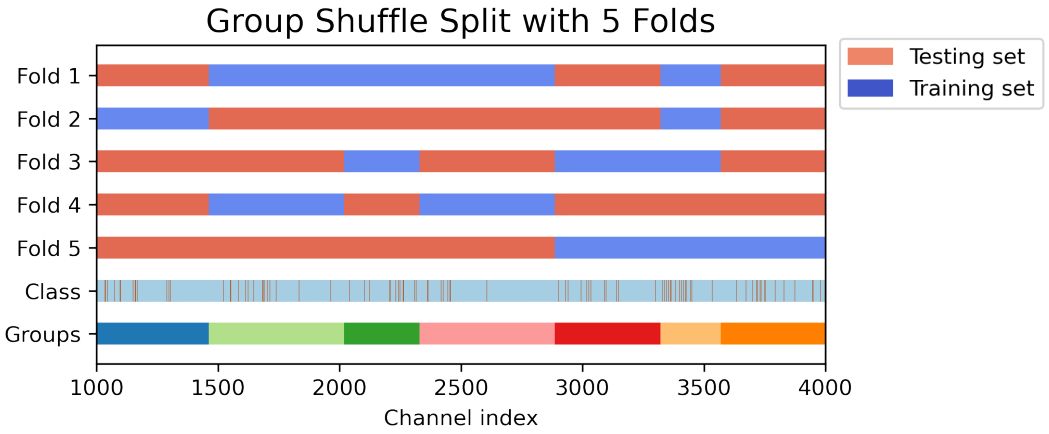


Figure 2.2: A working example of the group shuffle split cross-validation technique with test size = 50%. The x-axis represents the channel indices. The brown vertical lines in the *Class* row indicate bad channels, while the blue background represents good channels. *Groups* indicate the subjects, each containing a different number of sessions. For illustration purposes, we restricted the number of folds to five and the number of subjects (*Groups*) to seven. Note that no group simultaneously takes part in training and test sets in any given fold, thereby avoiding subject-specific data leakage.

We used the number of folds k = 10, a common choice in machine learning (James et al., 2013; Kuhn and Johnson, 2013), and for each fold, 50% of the data are used for testing on both newborns and infants datasets. As the considered adult EEG contains multiple sessions from the same subject (see Section 2.4.3), and to avoid subject-specific leakage in the training set, we used the group shuffling procedure (using the *GroupShuffleSplit* method from SciKit (Pedregosa et al., 2011)) rather than using the default random shuffling in each fold. An example is shown in Figure 2.2. For visualisation purposes, we show only five folds and seven groups (i.e., seven subjects with diverse sessions each). The 'Class' label indicates two classes: good and bad channels (indicated as vertical lines in orange). The 'Groups' label shows different colours for each subject, and the number of channels in each group varies (depending on the number of EEG recording sessions for each subject). It can be seen that the groups used as the training set for a particular fold are not used as the testing set (thereby avoiding data leakage). In each fold, different combinations of groups are used for training to effectively validate LOF on the limited EEG samples (113 files with 62 channels each leading to a total of 7006 EEG channels).

2.5.1.2 Experimental Results

We performed 10-fold cross-validation (Kuhn and Johnson, 2013) for each population dataset (with group shuffling (Pedregosa et al., 2011) for adult data and random shuffling for infants and newborns data). The average F1 Score across all folds is summarised in Figure 2.3. The numerical values are also reported in Table 2.1.

LOF unequivocally outperformed the other methods in all kinds of data, proving its robustness to different SNR ranges of real data obtained using distinct experimental paradigms. For newborns (Figure 2.3a) and infants (Figure 2.3b), we observed improved performance of up to 40% compared to other SoA methods. For adults (Figure 2.3c), an improvement in performance up to 87.5% was observed.

Data/Method	Mean F1 Score (s.e.m.)				
	Kurtosis	**FASTER**	**CRD**	**HAPPE**	**LOF**
Newborn	0.30 (0.022)	0.40 (0.014)	0.38 (0.019)	0.45 (0.016)	**0.63 (0.018)**
Infant	0.23 (0.012)	0.17 (0.011)	0.21 (0.006)	0.25 (0.008)	**0.35 (0.007)**
Adult	0.14 (0.008)	0.15 (0.006)	0.11 (0.008)	0.24 (0.006)	**0.45 (0.016)**

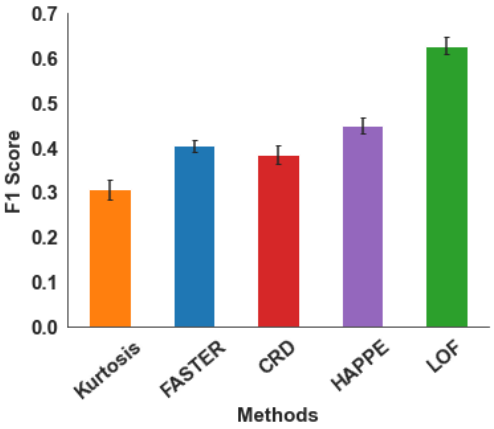
Table 2.1: a) Using F1 Score Performance of LOF across different datasets.

We then investigated how the optimal LOF_{thr} varies within and across populations using 10-fold cross-validation (see Figure 2.4). For newborns (noisy data, frequency-tagging paradigm), on average, the optimal threshold was identified as 2.6 ± 0.16. For infants (mildly noisy data, ERP paradigm), it was 1.6 ± 0.24. For adults (relatively clean data, event-related design with spectral power analysis), a further relaxed threshold of 1.4 ± 0.07 was identified to be optimal.

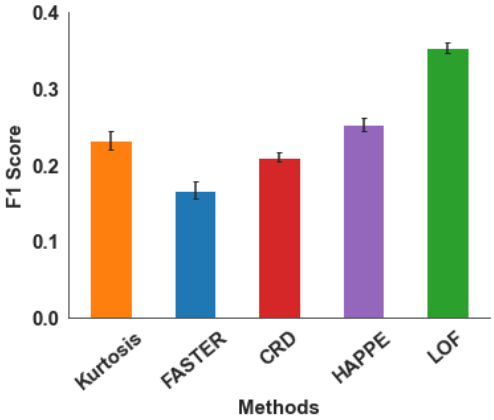
2.5.1.3 Discussion

Most current bad channel measures rely on distribution-based statistics (Mean, Variance, Kurtosis). The primary drawback of such measures is that the underlying EEG data distribution is not purely Normal/Gaussian. Therefore, fitting the data into standardised distributions might not produce satisfactory results. Further, these methods have been calibrated and validated on only one kind of EEG (i.e., either adult EEG or infant EEG). Given the differences in EEG signal quality according to the population and experimental design, these measures might not be reliable for other kinds of EEG than the ones they are intended for.

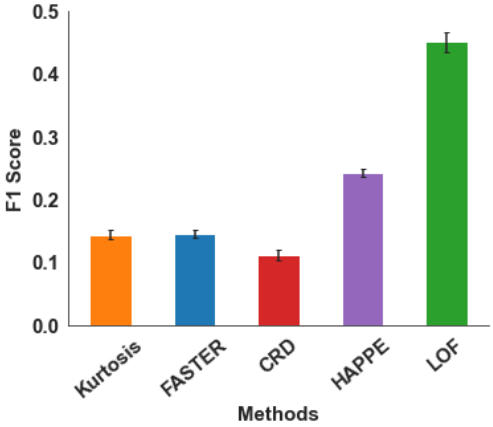
This work introduced a unique, robust measure (Local Outlier Factor) for detecting bad channels adapted to EEG acquired in any setting. LOF is an unsupervised outlier detection initially proposed for suspicious activity detection in Knowledge Discovery in Database (KDD) applications. However, there needs to be a clear indication of which should be the decision threshold to detect outliers. In theory, a data object (in our case, an EEG channel) is an outlier if it has a LOF score of more than 1.0. Our preliminary analysis suggested that this threshold is



(a) Newborn EEG.



(b) Infant EEG.



(c) Adult EEG.

Figure 2.3: Performance of bad channel detection methods using the 10-fold cross-validation technique. The error bars represent the s.e.m. across validation folds.

too strict (resulting in higher false alarms) for EEG data, which motivated the need to find the optimal LOF threshold. In this study, we showed how to find the optimal LOF threshold using a single dataset (employing a 10-fold cross-validation) to get the best results. Our analysis notably suggested that an optimal threshold for LOF lies around 2.5 for noisier data (newborns EEG) and approximately 1.5 for relatively cleaner data (infants and adults EEG). We strongly recommend the users follow a similar procedure to calibrate the LOF threshold for their own data. Precisely, we suggest the users take a portion of datasets to be analysed (or previously collected datasets using similar EEG settings) and visually inspect for the bad channels to calibrate the LOF threshold. In cases where it is impossible (due to the unavailability of labelled data), we suggest an initial threshold of 1.5 for infant and adult EEG and 2.5 for newborn EEG based on our study results. In the future, it is desirable to have variants of LOF or other local outlier detection algorithms without the subjectivity of the decision threshold. To this end, it is worth investigating the algorithms proposed in (Yuen and Mu, 2012; Yuen and Ortiz, 2017)

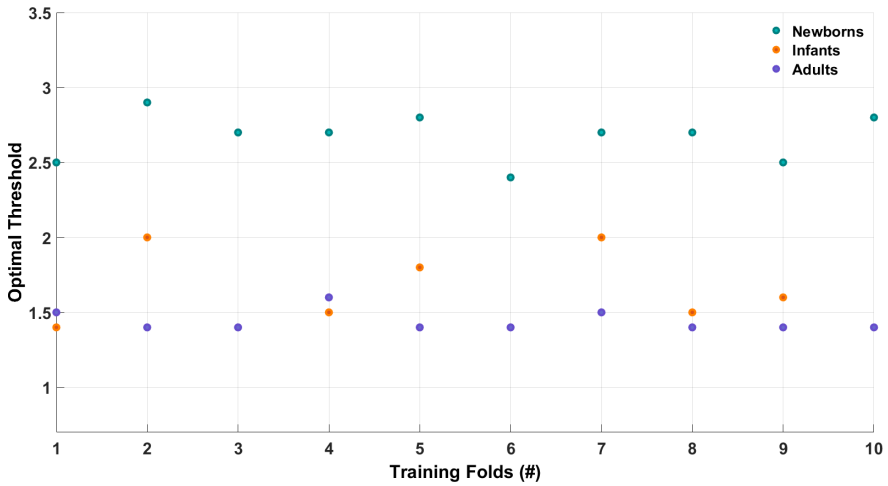


Figure 2.4: Summary of the optimal range of LOF_{thr} for different populations. For newborns (low SNR data), a relaxed threshold of 2.6 is optimal, whereas for infants (better SNR data), a value of 1.6 is found to be optimal. Finally, for adults (high SNR data), a conservative threshold of 1.4 is optimal.

for EEG bad channel detection.

Since LOF is a density-based approach, we also investigated the influence of the number of channels on the algorithm's performance. Our empirical results in both simulated and real adult EEG data suggest that LOF suits high-density EEG setups with at least 32 channels. Therefore, we recommend that users not use the proposed approach on low-density EEG (i.e., less than 32 channels). Further developments in the future are required to make LOF suitable even for low-density EEG.

A desirable property of LOF is that it does not assume any distribution for the raw EEG data. The LOF measure is loosely coupled to clustering algorithms (such as the *k*-nearest neighbour's algorithm (Fix and Hodges, 1989)) and is computed using the relative density of the identified clusters (Breunig et al., 2000). Thanks to this, LOF is adaptable to EEG acquired in different settings. With optimal parameters, LOF succeeded in detecting the annotated wrong channels compared to the traditional methods, such as Kurtosis, FASTER, and CRD, which assume a normal distribution for the EEG signal.

Remarkably, the second-best performing algorithm was HAPPE (Gabard-Durnam et al., 2018) in all datasets. This merit is likely because HAPPE is designed to deal with low SNR datasets (infants and children EEG). At the same time, other methods were validated on adult EEG (where the data quality is comparatively better). We also highlight that it is the only algorithm that uses normalised power values (i.e., frequency domain) to detect bad channels. All other methods use time-series measures (e.g., Hurst Exponent, Pearson Correlation, Channel Variance). This observation suggests that the frequency-specific measure is more efficient in

detecting artifacts at the channel level.

Given the outstanding performance of LOF, it is a promising bad channel detection in EEG acquired in any context from any population. Thanks to its versatility, LOF can be integrated into other existing EEG artifact removal pipelines, such as FASTER (for adult EEG) or HAPPE (for infant EEG), by replacing their respective bad channel techniques with LOF, which might lead to better overall artifact removal. We made the source code freely available as an EEGLAB plugin (Kumaravel, 2022). Even though we have not investigated the performance of LOF on Magneto-encephalography (MEG) data, LOF can also benefit MEG artifact removal.

2.5.2 Bad Segments Rejection using ASR

ASR requires a preliminary bad channel removal step executed using the LOF algorithm. Hence, the results presented in this section include the overall validation performance of LOF and ASR in newborn EEG.

2.5.2.1 ASR Parameters Calibration and Validation

Our considered newborn EEG data are collected using the frequency-tagging paradigm (see Section 2.4.1). To quantify the stimulus-related EEG response for both ASR parameter calibration and validation, we used the measure Frequency-Tagged Response (FTR) as defined in (Buiatti et al., 2019). FTR at the tag frequency (0.8 Hz) was calculated as the ratio between the power spectrum at the tagged frequency and the background power, i.e. the value at 0.8 Hz of the power-law fit of the power spectrum estimated from the six neighbouring frequency bins (±0.3 Hz), where the power-law fit was computed by fitting a line to the logarithm of the power at the six neighbouring frequency bins (*polyfit* function (MATLAB, 2018)).

EEG data were segmented in partially overlapping epochs of 10 s (overlap varied between one-half and three-fourths of epoch length to adjust to the variable length of clean data segments). For each electrode, each epoch's Fourier transform $F(f)$ was calculated using a fast Fourier transform algorithm (FFT function (MATLAB, 2018)). To avoid rejecting data segments shorter than 10 s but potentially containing relevant neural signals, zero-padding to 10 s was applied before FFT for data segments between 5 s and 10 s. Data segments shorter than 5 s were discarded. The power spectrum was calculated from these Fourier coefficients as the average over epochs of the single-epoch power spectrum:

$$PS(f) = \langle F(f) \times F^{*}(f) \rangle_{ep}$$

2.5.2.2 Experimental Results

To identify the optimal ASR parameter k and processing mode, we applied ASR on the newborn training dataset while systematically varying ASR parameter k between 1 and 100 for both

modes of processing (bad segment removal (ASR_R) and correction (ASR_C)). As a validation measure, after a preliminary bad segment removal by visual inspection, we identified a broad occipital cluster of electrodes showing a visual response (Figure 2.5, top inset); we then computed the average visual response FTR in this predefined occipital cluster for each k and processing mode. Results show that both processing modes achieve a similar maximum value of FTR by ($t(10) = -0.28$, $p = 0.78$), but for different k values: $k = 24$ for removal mode, $k = 13$ for correction mode. One possible explanation of this difference is that while for k between 20 and 30, the correction is not very effective, for $k < 15$, the removal mode rejects too many segments, providing too few samples for the computation of FFT. Since the two processing modes offer equivalent results to their optimal k, we will test both modes in the validation phase.

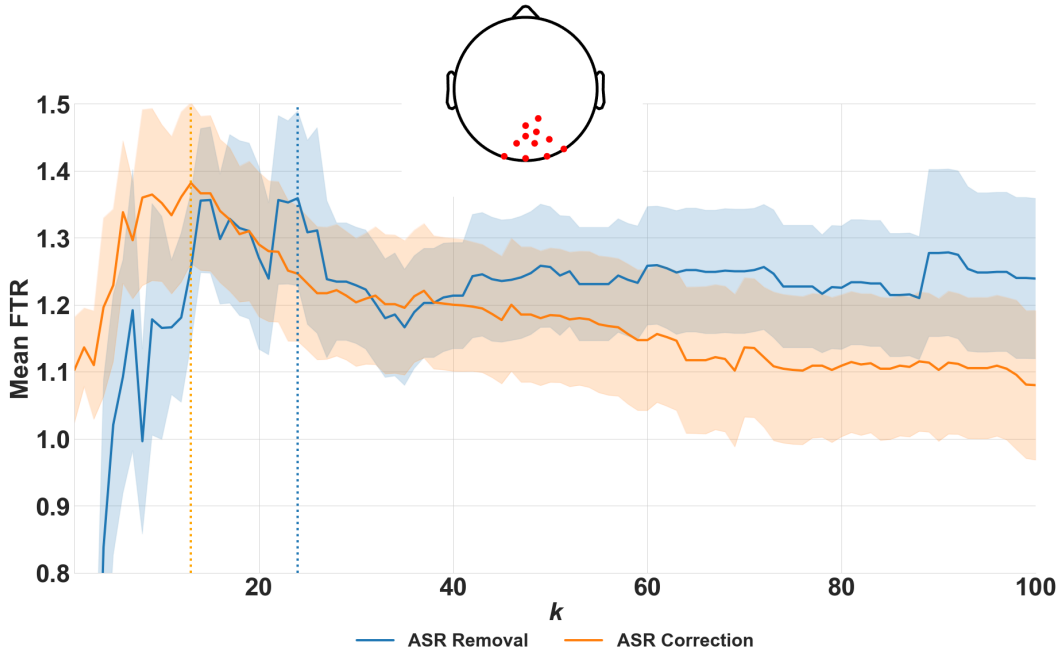
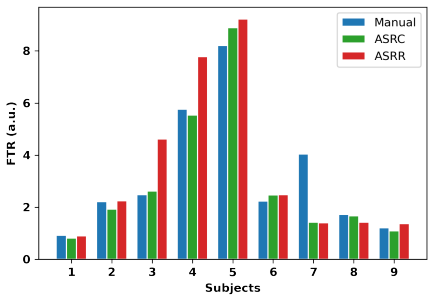


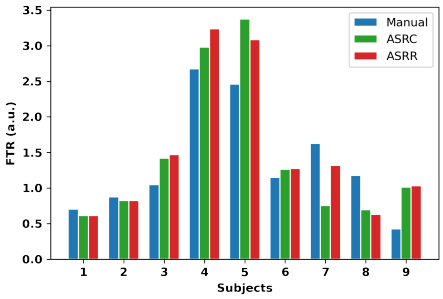
Figure 2.5: A grid-search analysis to find the best ASR parameter settings on the newborn Training Dataset: Average visual response (FTR) on a predefined occipital cluster of electrodes (topography in top inset) as a function of ASR Parameter k and Processing Mode, computed on the Training Dataset ($n = 11$). The mean FTR is maximum at $k = 13$ for ASR Correction (ASR_C) and $k = 24$ for ASR Removal (ASR_R). The shaded portion represents the standard error mean (s.e.m.) across subjects.

One out of ten subjects was discarded from further analysis as there were not enough samples to compute the FFT after ASR removal (ASR_R, $k = 24$). To have a fair comparison, this subject is also excluded from other methods, such as manual and ASR_C processing. The results are summarised in the Figure 2.6. Compared to the golden standard manual preprocessing, ASR improves the considered neural response FTR for some subjects (3, 4, 5, 9) and degrades for some subjects (7, 8). However, on average, the neural response FTR is comparable across all

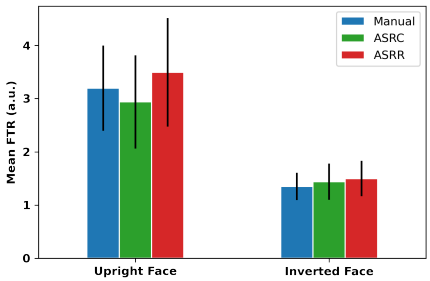
methods (Figure 2.6c) for both experimental conditions. Among the ASR processing modes, ASR_R slightly outperformed ASR_C. This shows that ASR preprocessing, despite being an automated method, works well for newborn data after a systematic parameter calibration.



(a) Upright Face Condition



(b) Inverted Face Condition



(c) Summary

Figure 2.6: Performance comparison of Manual, ASR_C, and ASR_R artifacts preprocessing using the neural measure: FTR. The error bars represent standard error mean (s.e.m.) across subjects.

2.5.2.3 Discussion

ASR is an efficient artifacts removal algorithm but strongly depends on two crucial user-defined parameters. The selection of these parameters is not univocal: the most systematic investigation on this issue (Chang et al., 2018) proposes that the optimal value of the ASR k parameter for adults lies "between 20 and 30", implicitly suggesting that it may be variable. Indeed, the preliminary results (presented in the Section 2.5.2.2) confirm that ASR performance significantly depends on the choice of both ASR Parameter k and processing mode. Therefore, a systematic calibration of these parameters is necessary on a subset of EEG data (a.k.a. training set) under study. In our case, we used the FTR measure, a Signal-to-Noise Ratio (SNR) metric for the frequency-tagged EEG datasets. In general, the evaluation metric for ASR calibration depends on the experimental paradigm and the application and must be carefully chosen to quantify the signal denoising efficiency. Given the comparable performance of ASR with respect to the golden standard manual processing, ASR can be considered

a potential replacement for subjective, time-consuming manual processing. As an automated approach, ASR ensures reproducible results, which is compromised in manual processing. Finally, among the ASR processing modes, ASR_R results in effective cleaning, even if it usually shortens the original data length. ASR_C preserves the data length while the correction of artifactual samples is not yielding the best results (evident in subjects 2, 3, 4, 5, and 9). This is most likely because ASR_C alters the neural response together with artifacts. Importantly, the statistical significance of experimental effects is preserved and similar to the manual rejection of artifacts in the original study (Buiatti et al., 2019). The overall evaluation results of the NEAR pipeline (integrating LOF and ASR algorithms) are presented in detail in the next chapter.

2.6 Conclusion

Typical EEG preprocessing comprises two main steps for artifact removal. They are 1) Bad channel rejection and 2) Bad segment rejection. In this chapter, we discussed the existing techniques in each category and why they might not be suitable for newborn/infant EEG. We have introduced a novel algorithm based on the LOF technique for EEG bad channel detection problems. Among the advantages, unlike other methods, LOF does not assume a normal distribution and assigns an outlier score for each channel based on its nearest neighbours. Since this is the first time that LOF has been applied to EEG for bad channel detection, we analysed its performance on developmental EEG and extended it to an adult EEG dataset. In all cases, LOF showed a robust performance proving its adaptability to different kinds of EEG data. The results are published in the MDPI Sensors Journal (Kumaravel, Buiatti, et al., 2022). Then, we successfully adapted the ASR algorithm to newborn EEG for the bad segment detection problem. Crucially, for the first time in the literature, we formally defined the standard ways to calibrate the ASR parameters (see Section 2.5.2.1) and showed that adapting the algorithm in different contexts with optimal parameters is possible. The preliminary results discussed in the Section 2.5.2.2 are partially published in the Journal of Developmental Cognitive Neuroscience Special Issue (Kumaravel, Farella, et al., 2022).

3 Preprocessing Pipeline for Offline Analysis

In this chapter, we propose NEAR (Newborn EEG Artifact Removal), a fully/semi-automated pipeline for efficient artifact removal from raw newborn/infant EEG data. NEAR consists mainly of two algorithms (LOF and ASR, discussed in the previous chapter) in addition to standard EEG preprocessing steps such as filtering, interpolating removed channels, and re-referencing. We start this chapter by providing an overview of two existing pipelines for developmental EEG (Section 3.1), and subsequently, we present the NEAR pipeline (Section 3.2) and its validation results (Section 3.5). Finally, we provide a general discussion of the proposed pipeline in Section 3.6. The materials presented in this chapter are published in (Kumaravel, Farella, et al., 2022).

3.1 Overview of Existing Pipeline for Developmental EEG

3.1.1 Harvard Automated Processing Pipeline for EEG (HAPPE)

The Harvard Automated Processing Pipeline for EEG (HAPPE) is a standardised, automated pipeline for developmental EEG with a high degree of artifact contamination and often short recording lengths (Gabard-Durnam et al., 2018). HAPPE pipeline consists of 9 steps, including bad channel rejection using pop_rejchan.m (Delorme and Makeig, 2004) and wavelet-integrated ICA decomposition to recover artifactual segments. Bad ICs are classified automatically using MARA (Winkler et al., 2014). HAPPE has been validated on resting-state developmental EEG data (age between 3 and 36 months). HAPPE is unsuitable for event-related designs (Gabard-Durnam et al., 2018), so we will only compare it with NEAR on continuous datasets.

3.1.2 Maryland Analysis of Developmental EEG (MADE)

The Maryland Analysis for Developmental EEG (MADE) is an automated, standardised preprocessing pipeline specifically developed for developmental populations (Debnath et al., 2020). MADE uses FASTER (Nolan et al., 2010) to remove bad channels and ICA to correct

data from artifacts. Bad ICs are classified automatically using Adjusted-ADJUST (Leach et al., 2020), an adapted version of ADJUST (Mognon et al., 2011) for infant data. Residual epochs contaminated by ocular artifacts are removed using a predefined amplitude threshold. MADE was validated on infants starting from 1 year of age to childhood (3–6 years old) and late adolescence (16 years old).

3.1.3 Drawbacks of HAPPE and MADE applied to Newborn EEG

First of all, neither of these existing preprocessing pipelines for developmental EEG was validated on newborns or infants less than 3 months old. To fill this gap, either these pipelines should be evaluated on newborn EEG or new pipelines should be developed. Secondly, despite the authors of these pipelines mentioning the drawbacks of ICA applied to short and heavily contaminated infant EEG, they eventually proposed modified versions of ICA (Wavelet-integrated ICA in HAPPE (Gabard-Durnam et al., 2018) and Adjusted-ADJUST ICA detection (Leach et al., 2020)). Further efforts are required to find alternatives to ICA as there might be better solutions for this data.

3.2 Proposed Pipeline: NEAR

NEAR preprocessing pipeline consists of a set of custom MATLAB scripts that can be executed as a fully automated EEG preprocessing within the EEGLAB (Delorme and Makeig, 2004) framework. The core, innovative parts of the pipeline integrating EEGLAB scripts with original custom scripts consist of the artifact removal processing block: preliminary calibration of the bad channel detection threshold (LOF) and of the ASR cut-off parameter, bad channel detection using LOF algorithm and correction/removal of bad segments using ASR, both endowed with original visualisation of the outcomes. In addition, we provided the scripts (based on EEGLAB functions) for a fully automated EEG processing from raw to clean data: importing and filtering raw data, interpolating removed channels and re-referencing. Depending on the application requirements, these auxiliary steps can easily be modified. Figure 3.1 shows the steps involved in the NEAR preprocessing pipeline. In the following sections, we describe each step in detail. A step-by-step tutorial, including figures illustrating the main steps of NEAR artifact removal, is presented in the Appendix B of this thesis.

3.2.1 Import Raw Data

NEAR supports import functionality of four main formats: .mff, .raw, .set and .edf. We considered these formats because most developmental EEG raw data fall into one of these categories. Users can import the data with EEGLAB importing tools for other formats and use NEAR with the resulting EEGLAB format .set.

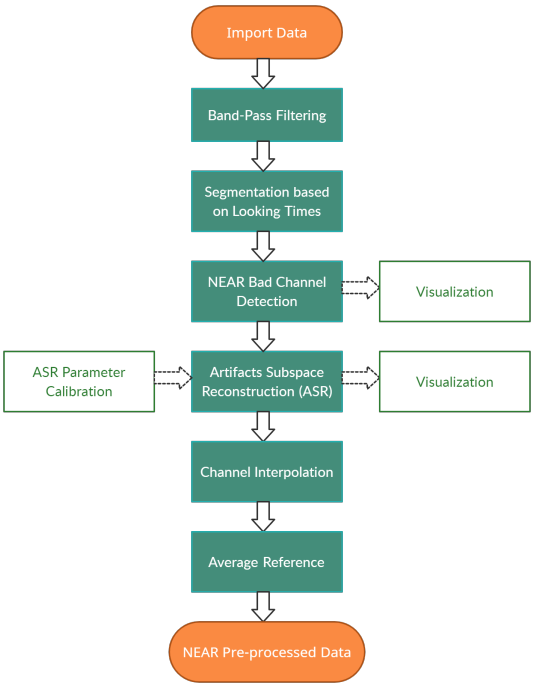


Figure 3.1: NEAR Outline

3.2.2 Band-Pass Filtering

The principle underlying band-pass filtering is that it is convenient for all the subsequent analyses to keep the frequency range of the signal we want to analyse and discard the higher and lower frequencies, especially those that likely contain artifacts. At the higher end, it is beneficial to use a low-pass filter with a cut-off frequency below the power line (50 Hz or 60 Hz) to avoid line noise. At the lower end, i.e. below 1 Hz, the EEG signal typically contains eye movement, respiration and heartbeat artifacts. We, therefore, recommend using the highest high-pass filter cut-off frequency that preserves the signal of interest, paying attention to the width of the filter roll-off and, in the case of ERP designs, to the risk of introducing spurious effects (Acunzo et al., 2012).

3.2.3 Data Segmentation

For studies involving a stimulation paradigm, an essential preprocessing step for identifying the relevant data in newborn/infant EEG recordings is to restrict data analysis to the intervals during which newborns/infants were effectively attending to the stimuli. Therefore, we recommend segmenting the data related to stimulation (i.e. segmenting periods for continuous stimulation or segmenting event-related epochs in case of event-related designs). Furthermore, for visual stimuli, we strongly recommend recording newborns/infants with a camera or an eye-tracker and light conditions guaranteeing precise monitoring of eye movements

and devoting careful attention to identifying the effective looking times. This preprocessing step is crucial because it minimises noise in the data and removes data segments associated with unattended intervals that are usually very artifacted, potentially causing biases in the subsequent artifact analysis. For this reason, this step is performed before detecting bad channels and segments.

Our scripts can also be adapted for resting-state EEG studies to retain good segments (or remove bad segments) of data, known as apriori. See Appendix B.7 for details.

3.2.4 Bad Channel Detection

Bad channel detection in newborn/infant EEG data is challenging because, due to the typically short preparation time devoted to lowering electrode impedance and frequent movement artifacts, electrode contact and stability are generally much lower than in adult data. In particular, after some preliminary tests on our data, we realised that the existing methods of bad channel detection are generally too strict for newborn EEG data. To overcome this issue, we implemented an algorithm in which the core step is a novel bad channel detection method based on LOF, a robust, data-driven outlier detector. The three steps of the algorithm are as follows:

1. **Flat Signals:** Because of defective contact with the scalp or disconnection from the recording device, sometimes electrodes record a flat signal. To remove these channels, we adopted the function *clean_flatlines* from the EEGLAB *clean_rawdata* plugin (https://github.com/sccn/clean_rawdata). A channel is marked as flat by default if it records a flat signal for more than 5 consecutive seconds.

2. **Local Outlier Factor (LOF):** LOF is extensively described in Section 2.3.1.

3. **Periodogram Analysis:** To detect channels that predominantly recorded motion-related artifacts that manifest as an increase in power in the beta range and a decrease in power in delta and alpha ranges (Georgieva et al., 2020), we implemented a bad channel detection method based on a spectral measure (*periodogram* function, MATLAB). For the datasets analysed in this work, we observed that while this method captures the most significant bad channels, they were already detected by at least one of the previous two steps. Therefore, we kept this method optional in NEAR's bad channel detection plugin (see Figure B.1).

3.2.5 Artifact removal using Artifact Subspace Reconstruction (ASR)

For the first time, in this work, we evaluate ASR on noisier developmental EEG data and propose it as one of the core blocks in our pipeline. In addition, we propose a calibration procedure for adapting the ASR algorithm to developmental data.

To calibrate ASR to newborn/infant EEG data, we analysed two crucial user-defined parameters of ASR, as discussed in the previous chapter (see Section 2.3.2).

3.2.6 Bad Channel Interpolation

The removed channels are interpolated from neighbouring channels using EEGLAB's function *pop_interp*. As suggested by EEGLAB developers, we recommend using "spherical" interpolation. However, using the NEAR pipeline, it is possible to use other supported techniques such as "v4".

3.2.7 Re-referencing

For re-referencing, NEAR provides options for both average re-referencing (recommended and most commonly used in developmental EEG studies) and re-referencing to a particular channel (e.g., Cz). For this task, NEAR uses EEGLAB's *pop_reref* function.

3.2.8 NEAR Parameters Calibration

A key feature of NEAR is the preliminary calibration of its artifact removal parameters (LOF bad channel threshold, ASR parameter k and ASR processing mode). We provide scripts for this calibration and highly recommend NEAR users form it based on previously analysed datasets from the same setup and experimental design, as these parameters impact the preprocessing quality.

3.2.9 Miscellaneous Functions of NEAR

NEAR supports both single-subject and batch processing (in the case of multiple subjects). The relevant scripts for these functionalities can be found in the repository. Finally, NEAR supports saving functionality and provides a comprehensive report summarising the preprocessing done on each input EEG file. This report might help review the effects of preprocessing done on the raw input EEG.

3.3 Validation Datasets

This section describes the datasets used for validation and comparative evaluation of NEAR with other methods and pipelines.

3.3.1 Simulated Data

Simulated data were generated with SEREEGA (Krol et al., 2018), a Matlab-based toolbox that simulates EEG datasets consisting of neurophysiologically realistic continuous and event-related brain activity. We generated two datasets simulating newborn EEG data with a frequency-tagging stimulation as in (Buiatti et al., 2019) and an event-related stimulation similar to the one in (Parise and Csibra, 2012). More specifically, we generated a 64-channel EEG dataset with the following components:

1. **Component 1** A stimulus-response, in the form of a sinusoidal Steady-State Visual Evoked Potential (SSVEP) (stimulation frequency = 0.8 Hz) for the frequency-tagging stimulation, and in the form of an event-related potential (latency=300 ms) for the event-related stimulation. Both responses were localised in two bilateral sources in the early visual cortex (MNI coordinates: [-8 -76 10] and [8 -76 10]).

2. **Component 2** Event-unrelated ongoing EEG activity originating in 62 randomly selected cortical sources, plus in the two sources of the first component located in the early visual cortex. Such activity is generated as Brown noise (power spectrum increasing as for), mimicking the one observed in newborns (Fransson et al., 2013). Notably, the signal-to-noise ratio between component 1 and component 2 was of the same order of magnitude as the one measured on real, artifact-free newborn EEG data. The first two components represent the ground truth.

3. **Component 3** Artifacts in (5 randomly chosen) single channels consisting of irregular potential shifts and flat signals mimicking electrical discontinuities and low-frequency fluctuations (0–10 Hz) mimicking local bad contacts and movement artifacts.

4. **Component 4** Transient high-amplitude artifacts involving all the channels in the form of intermittent abrupt potential shifts or smoother Gaussian-like fluctuations, where both the amplitude at each channel and the duration varies randomly for each transient artifact (mean duration=1.6 s). Time and amplitudes are of the same order of magnitude as those observed in real newborn data. This component mimics motion artifacts, which are very frequent in newborns.

   Figure 3.2 shows that this simulation represents the main features of newborn EEG ongoing activity and artifacts. The scripts generating the simulation datasets are available at https://github.com/vpKumaravel/NEAR, and the simulated datasets described in the Results are available here: https://osf.io/79mzg/.

3.3.2 Newborn Datasets

The readers may refer to Section 2.4.1 for details.

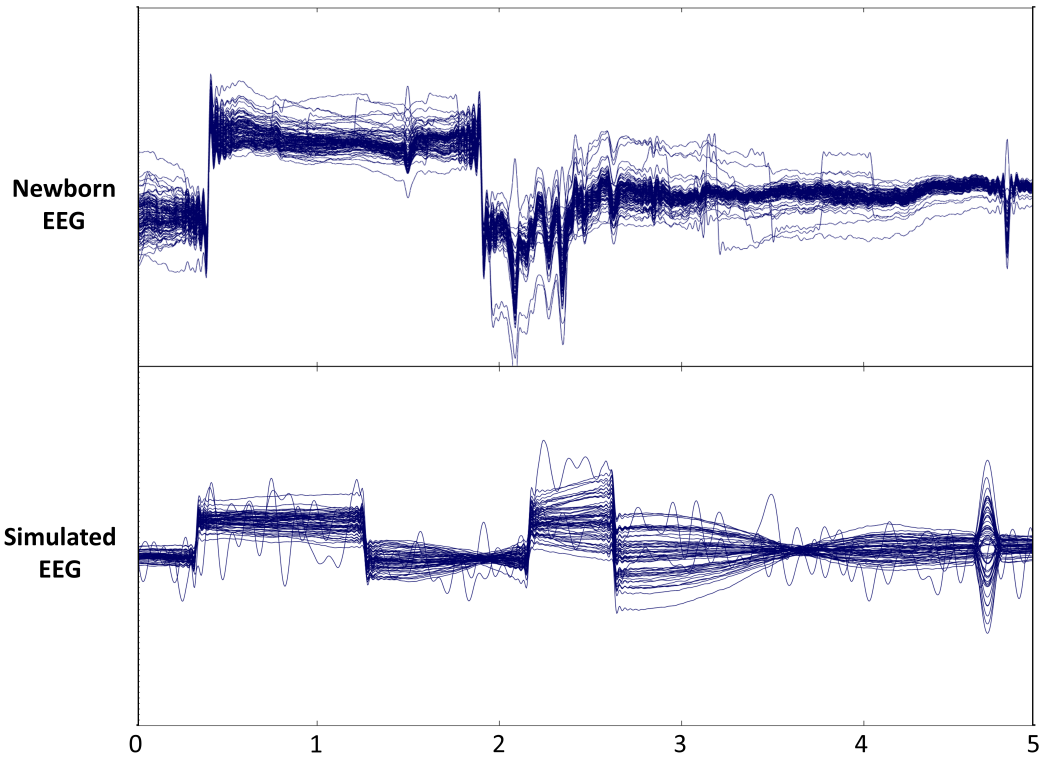


Figure 3.2: Simulated EEG. *Top Panel:* Newborn EEG data from (Buiatti et al., 2019). *Bottom Panel:* Simulated EEG data (ground truth plus artifacts). Data are shown in butterfly mode (all electrode signals overlapped). The y-axis represents the arbitrary voltage values.

3.3.3 Infant Datasets

The readers may refer to Section 2.4.2 for details.

3.4 Standard Preprocessing of Real EEG Datasets

3.4.1 Newborns

Bad channels were detected on both datasets after band-pass-filtering and segmentation. Channels were marked as bad if they 1) had a standard deviation (computed on the whole data length by using the TrimOutlier toolbox: https://sccn.ucsd.edu/wiki/TrimOutlier) higher than 150 μ V (to detect channels with high-amplitude artifacts) or lower than 1 μ V (to detect flat or weakly responsive channels); 2) showed artifactual patterns after accurate visual inspection of the time course and power spectrum plots of suspicious channels and comparison with their neighbours.

Once bad channels were removed, identification of bad data segments was based on 1) the

detection of amplitude jumps exceeding ± 200 μ V; 2) the presence of paroxysmal artifacts after accurate visual inspection of the time course and topography of the EEG data.

3.4.2 Infants

Both infant datasets were automatically and manually edited. Automatic data rejection for body and eyes movements was performed whenever the average amplitude of an 80 ms sliding window exceeded ± 200 μ V at any channel. A bad channel score was obtained by considering the channels marked as rejected as bad for at least 40% of the epochs. Bad channels were automatically interpolated in epochs in which ≥ 10% of the channels contained artifacts; epochs in which > 10% of the channels contained artifacts were automatically rejected. Data was then manually edited by visual inspection of each individual epoch.

3.5 Validation Results

3.5.1 Simulated Data

We first validated NEAR on two synthetic EEG datasets simulating EEG signals that contain an SSVEP at 0.8 Hz like in (Buiatti et al., 2019) (frequency-tagging dataset) and an ERP response similar to the one recorded in (Parise and Csibra, 2012) (ERP dataset), respectively. Both datasets also include three key components of newborn/infant EEG data: Brown-noise-like background EEG, artifacts in single channels mimicking bad or unstable electrode contacts, and transient high-amplitude fluctuations across most of the channels mimicking motion artifacts. Signal-to-noise ratios, data duration and proportion of artifacts are similar to the ones of real data (Buiatti et al., 2019; Parise and Csibra, 2012). Since it is difficult to incorporate enough variability to generate realistically different training and test datasets within the simulation framework, we set NEAR parameters to predefined values: LOF threshold = 2 and ASR parameter $k = 20$

3.5.1.1 Frequency-tagging Dataset

The ground truth data (SSVEP plus Brown-noise-like background EEG) shows a clear peak in the power spectrum at the stimulation frequency (0.8 Hz) that stands out of the background EEG power spectrum (blue line in Figure 3.3). The topography of the associated FTR at 0.8 Hz shows a neat posterior medial activation (Figure 3.3, bottom panel) fully compatible with the early visual cortex sources generated in the simulation (see Section 3.3.1). Artifacts cause a massive positive shift of the power spectrum at low frequencies, almost completely masking the SSVEP response peak (red line in Fig. 3.3). Consequently, the topography of the FTR at 0.8 Hz does not show any clear posterior activation (Figure 3.3, bottom panel).

NEAR bad channel detection algorithm efficiently captured the simulated 5 bad channels (and no additional channels). ASR_R was very efficient in removing all the short artifacted segments

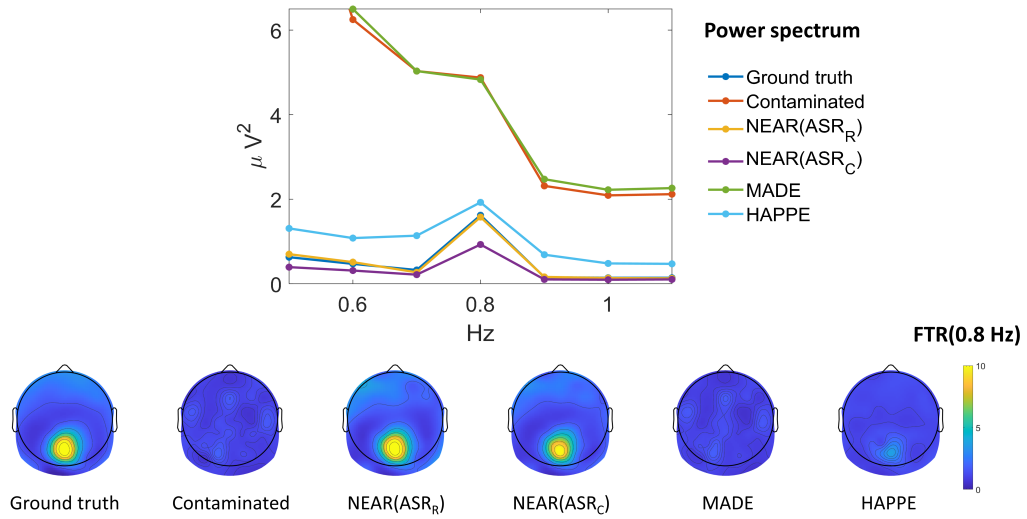


Figure 3.3: Simulated FT-EEG Processed. *Top panel:* Power spectrum of the simulated frequency-tagging dataset between 0.5 and 1.1 Hz, averaged over the electrodes showing the largest FTR amplitude in the ground truth data (PO3, POz, PO4). *Bottom panel:* Topography of the FTR (defined in Section 2.5.2.1) at 0.8 Hz (the stimulation frequency).

from the data: the resulting peak at the stimulation frequency in the power spectrum almost overlaps with the one of the ground truth data (yellow line in Figure 3.3), and the topography of the FTR at 0.8 Hz is very similar to the one of the ground truth (Figure 3.3, bottom panel). ASR_C performance was slightly inferior. While the power spectrum peak was recovered, its amplitude was lower than the ground truth. The overall power spectrum at low frequencies was shifted to lower values (magenta line in Figure 3.3). This could depend on the fact that while all transient artifacts were correctly detected and removed by ASR, the correction also suppressed part of the SSVEP and of the background EEG. Nevertheless, the FTR topography was similar to the ground truth one, even with a slightly lower amplitude (Figure 3.3, bottom panel).

For comparison with state-of-the-art methods for artifact removal, we also tested the ICA-based artifact removal pipelines of MADE (Debnath et al., 2020) and HAPPE (Gabard-Durnam et al., 2018). MADE was unable to correct or remove almost any of the transient artifacts, as shown by its power spectrum (green line in Figure 3.3) and its FTR topography at 0.8 Hz (Figure 3.3, bottom panel), which are both very similar to the ones of the contaminated data. HAPPE was more successful: it corrected most of the low-frequency artifacts (cyan line in Figure 3.3) and FTR topography shows a posterior activation similar to that of the ground truth, although with a much lower amplitude than the ground truth and NEAR processing with ASR in both modalities (Figure 3.3, bottom panel). The rationale behind this reduction in overall amplitude might be due to the wavelet-based ICA thresholding, as also highlighted by the authors (Gabard-Durnam et al., 2018).

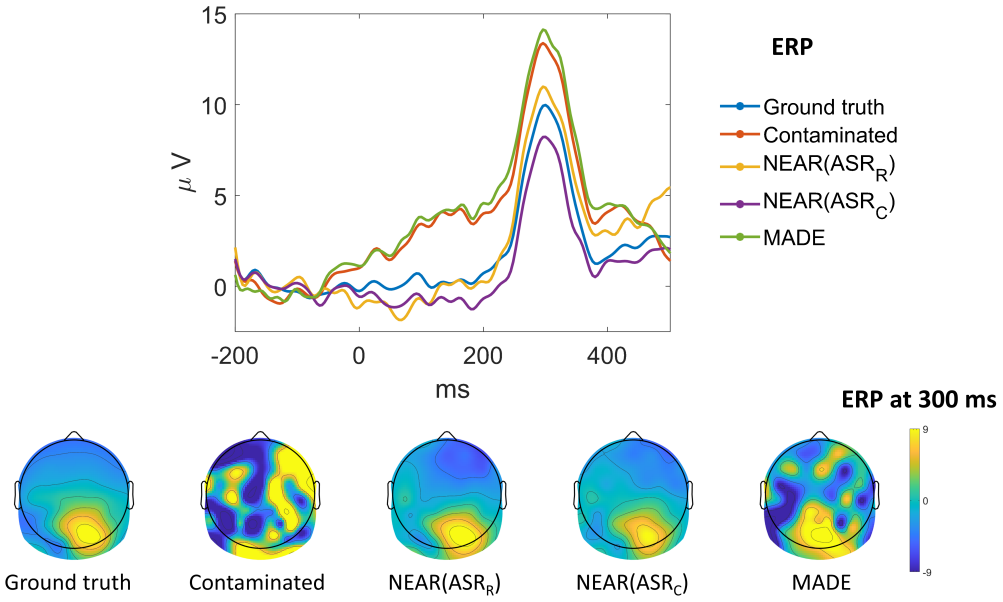


Figure 3.4: Simulated ERP-EEG Processed. *Top panel:* ERP of the simulated ERP dataset around the onset of the simulated stimulus, averaged over the electrodes showing the largest ERP amplitude in the ground truth data (PO3, POz, PO4). *Bottom panel:* Topography of the ERP averaged between 275 and 325 ms.

3.5.1.2 ERP dataset

Results on the ERP dataset were very similar to the ones of the frequency-tagging dataset. The ground truth ERP mildly fluctuates around zero until 200 ms, then rises at 300 ms (the peak latency) and decreases afterwards (blue line, top panel of Figure 3.4). Its topography at the peak latency is neatly posterior (bottom panel, Figure 3.4). Artifacts cause the ERP to rise even before the stimulus onset spuriously. Although the ERP peak is visible in the posterior electrodes (red line, top panel of Figure 3.4), the topography at the peak latency is very noisy (bottom panel, Figure 3.4).

NEAR bad channel detection algorithm efficiently captured the simulated 5 bad channels (and no additional channels). ASR_R was very efficient also in this case in removing all the transient artifacted segments from the data: the ERP peak at 300 ms almost overlaps with the one of the ground truth, even if the ERP profile is a bit noisier at higher latencies (yellow line, top panel of Figure 3.4), possibly an effect of the lower number of trials. The topography at 300 ms is very similar to the one of the ground truth (Figure 5, bottom panel). ASR_C ERP peak has a lower amplitude than the ground truth, but the ERP profile outside the peak is very clean with common fluctuations around zero (magenta line in Figure 3.4). The ERP topography at 300 ms is as neat as the one of the ground truth (Figure 3.4, bottom panel).

In comparison, also, in this case, MADE could not remove the artifacts on the electrodes showing the posterior activation (green line, top panel of Figure 3.4). However, its topography

at 300 ms shows moderate success in removing artifacts from other electrodes, though much less successfully than NEAR (Figure 3.4, bottom panel).

3.5.2 Newborn Data

For the visual response, standard processing resulted in a significant effect even with one less subject (t(8) = 3.03, P = 0.016) (Figure 3.5, first row, left-hand panel). Compared to standard processing, ASR_R resulted in a somewhat lower power peak at the tag frequency accompanied by a similar decrease in the background power (Figure 3.5, second row, left-hand panel), likely resulting from a more efficient noise reduction together with a slight power reduction. This minor difference impacted equivalently on the numerator and denominator of the FTR, obtaining a significant effect (t(8) = 3.04; P = 0.016) equivalent to the standard processing and a response that is statistically indistinguishable from the standard mode (paired t-test between standard and ASR_R of the difference between power and background at the tag frequency across subjects: t(8) = 0.034, P = 0.97). The power spectrum resulting from ASR_C is further reduced, particularly at the tag frequency (Figure 3.5, third row, left-hand panel), probably due to a slightly sub-optimal reconstruction of the steady-state response in bad segments. Nonetheless, the statistical effect is also significant (t(8) = 2.60, P = 0.032), and the response is only marginally lower than the standard mode (paired t-test as above: t(8) = 1.91, P = 0.093). On the contrary, the overall profile of the power spectrum resulting from MADE processing is notably higher than the one from the standard mode and with a much wider variance at low frequencies (<0.8 Hz), likely the effect of residual low-frequency artifacts (Figure 3.5, fourth row, left-hand panel). Still, the visual response is statistically significant also in this case (t(8) = 2.47, P = 0.039), though marginally lower than the one obtained with ASR_R (paired t-test t(8) = 1.95, P = 0.086) and with standard correction (paired t-test t(8) = 2.14, P = 0.065). HAPPE (Figure 3.5, fifth row, left-hand panel) also recovers a statistically significant peak of the visual response (t(8) = 2.58, P = 0.033). Still, it is significantly lower than for ASR_R (t(8) = 3.04, P = 0.016), ASR_C (t(8) = 2.59, P = 0.032) and standard processing (t(8) = 3.01, P = 0.016).

Validation of the face-like pattern response shows similar results. NEAR with ASR_R processing recovers a statistically significant effect (t(8) = 2.79, P = 0.023). The face-like response is statistically equivalent to that obtained with the standard processing (paired t-test between standard and ASR_R of the difference between FTR for face-like and inverted face-like patterns across subjects: t(8) = -0.38, P = 0.71) (Figure 3.5, first and second row, middle panel). Similar results are obtained with ASR_C: a significant face-like effect (t(8) = 2.69, P = 0.027), and no significant difference with standard mode (t(8) = 1.67, P = 0.13) (Figure 3.5, third row, middle panel). However, MADE processing resulted in a shallower spectral peak (Figure 3.5, fourth row, central panel) and recovered only a marginally significant face-like effect (t(8) = 2.02, P = 0.078), showing again a slightly significant difference compared to both ASR_R (t(8) = 1.96, P = 0.085) and standard processing (t(8) = 1.88, P = 0.097). HAPPE processing resulted in an even shallower peak (Figure 3.5, fifth row, middle panel), failing to report a significant face-like effect (t(8) = 1.21, P = 0.26). In this case, the difference with NEAR methods is not significant (vs

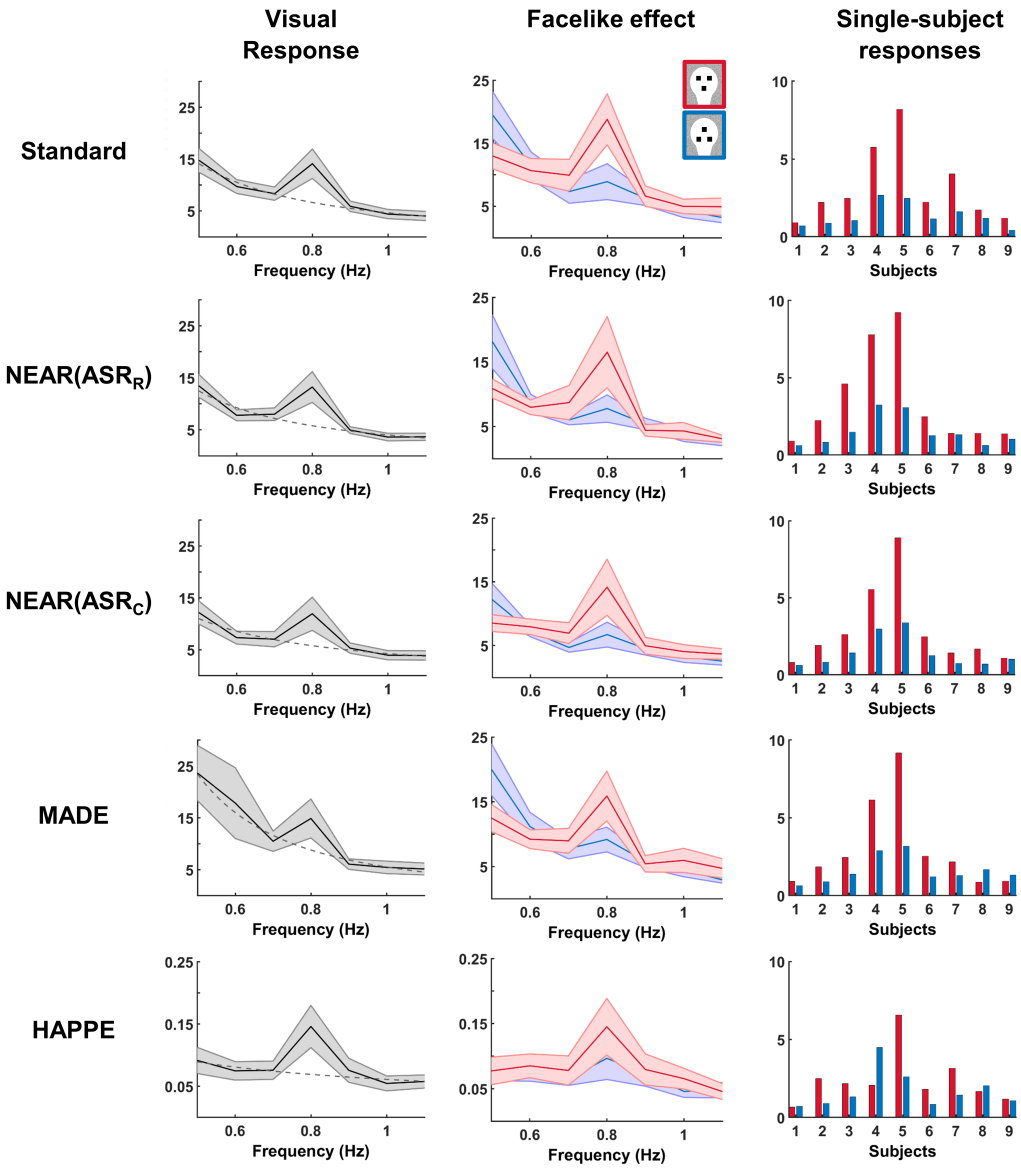


Figure 3.5: Performance of NEAR in obtaining statistically significant neural responses from the raw newborn Test Dataset. Each row corresponds to an artifact removal method: Standard processing (Buiatti et al., 2019), NEAR using ASR_R, NEAR using ASR_C, MADE and HAPPE, respectively. *Left-hand column:* Power spectrum elicited by the overall visual stimulation. Shaded contour indicates the s.e.m. across subjects. The spectral peak at the tag frequency is statistically significant for all the methods, but NEAR using ASR_R obtains the highest t-value. *Middle column:* Power spectrum associated with upright (red line) and inverted (blue line) face-like stimuli. While the face-like effect is statistically significant for both NEAR processing modes, it is only marginally significant after MADE processing and not significant after HAPPE processing. *Right-hand column:* Single-subject FTR for upright (red bars) and inverted (blue bars) face-like images. The face-like effect is present in all subjects for NEAR processing but not for MADE and HAPPE processing.

ASR_R: t(8) = 1.59, P = 0.15; vs ASR_C: t(8) = 1.35, P = 0.21), nor the one with standard processing (t(8) = 1.98, P = 0.08). These results are reflected in the single-subject responses (Figure 3.5, right-hand column panel): While NEAR with both ASR_R and ASR_C recovers a preference for face-like patterns for all the subjects as with standard processing, two subjects following MADE and HAPPE processing show an inverted effect.

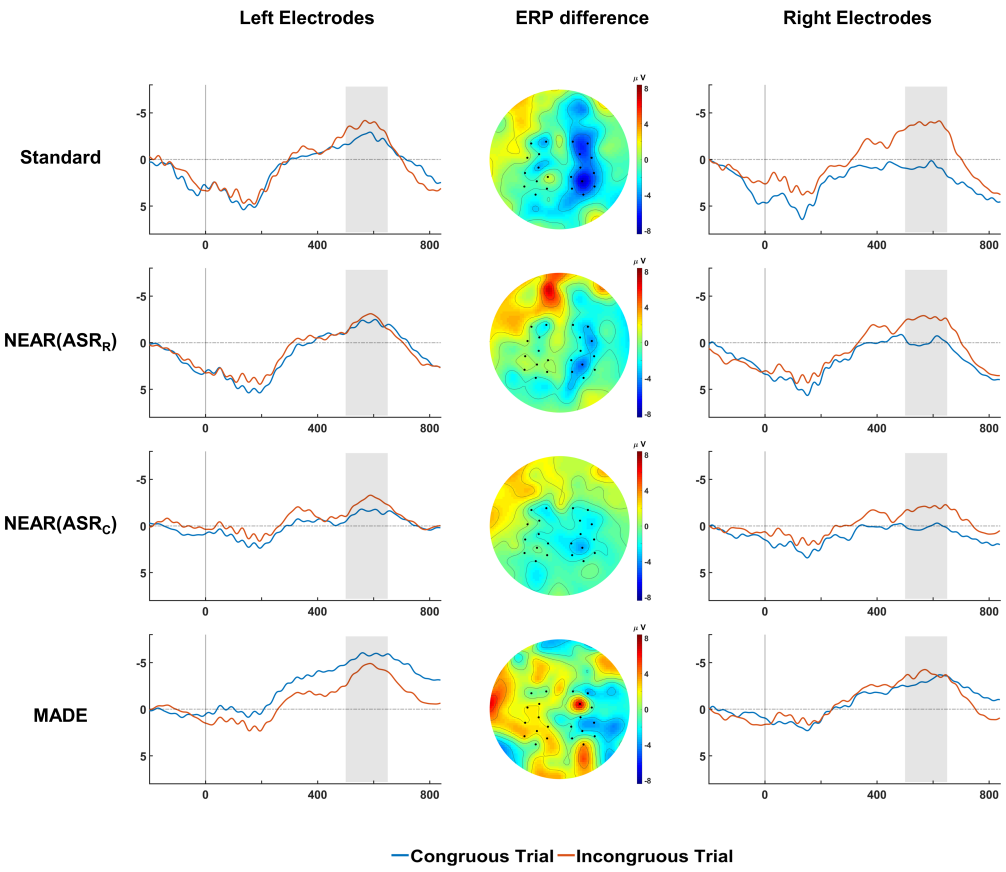


Figure 3.6: Real EEG Newborn. Event-related potential (ERP) results for each of the processing modes: Standard processing (Parise and Csibra, 2012), NEAR using ASR_R, NEAR using ASR_C and MADE, respectively. The figure shows grand-average waveforms on congruous and incongruous trials in left (*left-hand panels*) and right (*right-hand panels*) regions of interest (marked by black points on the scalp maps). The grey shading indicates the time window of the infant N400 (500–650 ms), and the vertical line marks the time at which the object in each trial appeared from behind an occluder. The scalp maps (*middle panels*) depict the spatial distribution of the difference in ERP amplitude between incongruous and congruous trials in the given time window.

3.5.3 Infant Data

As for newborns, we validated the overall performance of NEAR pre-processing by direct comparison of the statistical significance of the main effect obtained by manual preprocess-

ing in the original work (Parise and Csibra, 2012): an N400 differential response between incongruous and congruous conditions higher on the right region-of-interest than on the left one (where the regions of interest were identified by the electrodes between C3 and P3 and between C4 and P4, over the left and right hemisphere, respectively (Figure 3.6).

We also compared NEAR's performance with the state-of-the-art artifact removal pipeline for developmental data MADE (Debnath et al., 2020). Results show that the only method that recovered a significant ANOVA with factors condition and hemisphere is ASR_R ($F(1,13) = 5.13$, $P = 0.041$). In contrast, no significant effect is observed for ASR_C ($F(1,13) = 1.68$, $P = 0.22$), nor for MADE ($F(1,13) = 2.90$, $P = 0.11$). More specifically, NEAR using ASR_R yielded a clear congruency effect on the right hemisphere that was absent on the left hemisphere, similar to standard processing (Figure 3.6, first two rows). NEAR using ASR_C resulted in similar but shallower effects compared to ASR_R (Figure 3.6, third row). MADE also exhibited a congruency difference between the hemispheres in the same direction but with the congruous condition higher than the other methods (Figure 3.6, fourth row). However, no significant difference was found between the three methods on the size of the effect (paired t-test between the ERP difference between hemispheres of the difference between conditions of ASR_R vs MADE: $t(13) = 0.01$, $P = 0.99$; ASR_C vs MADE: $t(13) = -1.00$, $P = 0.33$; ASR_R vs ASR_C: $t(13) = 1.70$, $P = 0.11$).

3.6 Discussion

This chapter presented NEAR, a pipeline that transforms artifacted raw developmental EEG data into clean data ready for downstream analysis. We demonstrated that NEAR's novel artifact removal procedure efficiently removes artifacts both from newborn and infant EEG data (high sensitivity) while preserving the EEG signal of neural origin (high specificity). NEAR will hopefully contribute to establishing a more objective and reproducible preprocessing procedure within the developmental EEG community, a much-needed improvement considering the negative consequences of the variability of EEG data editing practices (Monroy et al., 2021). Hereafter we comment on some key aspects of NEAR in the general context of EEG artifact removal.

3.6.1 An Artifact Removal Method for Non-stereotyped Artifacts

The most problematic and predominant artifacts in newborn EEG data are non-stereotyped transient high-amplitude fluctuations involving variable sets of channels. By precisely simulating these artifacts, we showed that ASR processing included in NEAR is very efficient in detecting and removing them. On the other hand, ICA-based methods such as MADE (Debnath et al., 2020) and (to a lesser extent) HAPPE (Gabard-Durnam et al., 2018) failed in processing these artifacts, most probably because they are developed to detect mainly the stereotyped ones. Notably, both MADE and HAPPE are more successful on newborn EEG data than on simulated data, possibly because real newborn EEG artifacts are more stereotyped (i.e. their spatial distribution and temporal profile are partly correlated across occurrences)

than simulated ones, which are generated by random shuffling of the artifact topographical distribution. As discussed more extensively below, a combination of detection methods for non-stereotyped and stereotyped artifacts might solve the wide range of EEG artifacts, especially in developmental data.

3.6.2 ASR Parameter Calibration

One core tool NEAR uses is ASR (Mullen et al., 2015), an efficient algorithm that nevertheless depends on some user-defined parameters. The selection of these parameters is not univocal: the most systematic investigation on this issue (Chang et al., 2018) proposes that the optimal value of the ASR k parameter for adults lies "between 20 and 30", implicitly suggesting that it may be variable. Moreover, while ASR's default processing mode is to correct the data from artifacts (a choice driven by the original aim of providing an efficient algorithm for real-time applications), the main developers of the EEGLAB software suggest removing the artifacted segments identified by ASR because the effects of ASR correction on the data "are not clearly understood" (https://eeglab.org/tutorials/06_RejectArtifacts/cleanrawdata.html, Retrieved June 7, 2021). Our study confirms that ASR performance significantly depends on the choice of both ASR Parameter k and processing mode (Figure 2.5). The quality of developmental EEG data may vary substantially between different EEG setups, and data analyses may require different thresholds. Therefore, we propose an adaptive approach to ASR: run ASR on a dataset previously collected with the same EEG setup and analysed with the same analysis chosen for the current data and find the k and processing mode that best recover the EEG effects observed on that dataset. We provide a script for this calibration procedure and recommend that NEAR users perform it before applying NEAR to newly recorded data. Once these parameters are identified, validation shown in this work suggests that NEAR might be safely used in automatic mode. If a training dataset is unavailable, we recommend tuning NEAR parameters on at least a few subjects by measuring a well-known sensory response for both processing modes and k between 10 and 30. In any case, we strongly recommend that users keep monitoring the efficiency of NEAR (an easy task provided by the visualisation tools along NEAR's pipeline and the report file), as unpredicted single-subject variations are always possible.

3.6.3 Artifact Removal (ASR_R) vs Correction (ASR_C)

Testing ASR on simulated data showed that removal mode is slightly more efficient than correction mode in cleaning the data from the effect of artifacts; results from the simulation data suggest that while correction efficiently suppresses the artifacts, it also severely attenuates the underlying neural activity. This effect was consistently observed in the application of NEAR on both the newborn and infant data. A possible explanation for this lies in the core part of the ASR algorithm: Principal Component Analysis, which blindly separates data into different Principal Components (PCs) based on their explained variance. Then, the algorithm assumes the PCs with higher variance than the ones obtained in the calibrated data are "bad" and removed. However, there is a reasonable probability that these removed components may

also contain the neural response. Therefore, removing those components eliminates the noise along with the neural response (although marginal). This observation led us to adopt EEGLAB's recommendation to set the removal mode by default for offline analysis unless the performance of the correction mode on some training datasets shows significantly better results. In both processing modes, we recommend users notice the amount of data being rejected (in case of ASR_R) or modified (in case of ASR_C) and the mean reduction of RMS variance in the processed signal. These values can be found in our report files. In particular, for ASR_C, we recommend users customise these values to set inclusion criteria for the subjects into the group-level analysis to avoid the risk of primarily relying on the reconstruction of heavily artefacted data.

3.6.4 Using NEAR on other experimental designs

NEAR has been trained and validated on a frequency-tagging paradigm by using a measure of the SSVEP and on an event-related design by using an ERP measure. The adaptive approach of NEAR provides a straightforward strategy to tune NEAR parameters to data recorded from other experimental designs that include event-related measures like time-frequency analysis or resting-state measures like (de) synchronisation in specific frequency ranges or connectivity measures.

3.6.5 Combining NEAR with ICA for developmental EEG artifact removal

In comparison with NEAR, the pipeline for artifact removal of developmental data MADE performed moderately worse on newborn and infant data, mainly because it was not equally efficient in removing low-frequency artifacts. We see two possible reasons for this difference: 1) MADE's bad channel and bad segment identification tools were not calibrated to newborn EEG data; 2) As mentioned above, the benefit of removing artifacts by ICA is limited by the fact that most artifacts in newborn EEG data are non-stereotyped, therefore not easily captured by ICA. Nonetheless, ICA (and in particular Adjusted-ADJUST (Leach et al., 2020), the IC classifier developed for infant data and included in MADE) may be beneficial as a further processing step after NEAR because it could correct the data from residual stereotyped artifacts without any further data rejection. However, one issue that might be problematic for an efficient ICA decomposition of developmental EEG data with high-density systems is the very limited duration of the clean data segments. Rather than reducing the number of electrodes as in (Leach et al., 2020) (which would drastically decrease EEG spatial resolution, preventing a potential source reconstruction (Odabaee et al., 2013)), a possible solution would be to use PCA for dimensionality reduction. However, the application of PCA on EEG data has significant limitations (Artoni et al., 2018); therefore, investigations on alternative methods to run ICA on high-density EEG data of short duration would be very useful.

4 Toward a Wearable EEG Hardware/-Software Solution for Newborns

Wearable/Mobile EEG systems enhance the ease and flexibility of brain data acquisition in developmental populations. Further, such systems support longitudinal, large-scale studies that can help understand the developmental neural signatures of typical and atypical brains. However, the current commercial mobile EEG systems (BrainProducts, AntNeuro, Cognionics, mBrainTrain) do not support low weight, minimal volume, maximal comfort of the contact of the electrodes with the scalp, which are crucial requirements for recording newborns. In this thesis, we explore a current state-of-the-art ultra-low power custom EEG device called BioWolf (a research prototype developed by the University of Bologna and ETH Zurich). In particular, we investigate the quality of frequency-tagged EEG acquired by this system using 8 soft, dry electrodes with and without any artifacts removal algorithm. As an initial step, we validated our algorithms on data collected from adult subjects using a similar experimental paradigm (e.g. low-frequency SSVEP stimulation) that we intend to use in developmental subjects. The materials presented in this chapter are published in (Kartsch et al., 2022; Kumaravel, Kartsch, et al., 2021; Kumaravel, Paissan, and Farella, 2021; Paissan et al., 2022).

4.1 BioWolf Platform

BioWolf system is based on Mr.Wolf, a multi-core Parallel Ultra-Low Power (PULP) System-on-Chip (SOC). Mr.Wolf is implemented in TSMC 40 nm CMOS technology, combining a 12 Kgates RISC-V processor that manages power and peripherals, along with a parallel cluster of eight RISC-V processors supporting flexible DSP extensions including floating-point units (Schiavone et al., 2017).

BioWolf system comprises three SoC: 1) Mr. Wolf for handling the signal processing operations, 2) Nordic ARM-SoC (nRF52832) for Bluetooth Low Energy (BLE) communications, and 3) an Analog Front End (AFE) for bio-signal acquisition from Texas Instruments (ADS1298). BioWolf has a built-in nano-power buck-boost regulator for energy harvesting, a fuel gauge to measure the battery energy level (BQ25570), an electrode impedance tracker that provides feedback on the quality of the electrode contact and an Inertial Measurement Unit (IMU) device (IIS2DH)

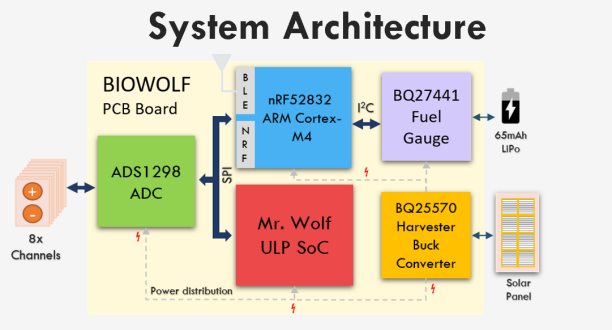


Figure 4.1: The system architecture of BioWolf

to measure acceleration and temperature. These additional subsystems provide enhanced support for high-quality EEG recordings in portable and wireless settings. Figure 4.1 shows the block diagram of the complete system.

Since BioWolf is built with a combination of processors (Mr. Wolf and Nordic), BioWolf can be configured to work in three different modes: Stream, Processing and Deep-Sleep. In *stream* mode, the Mr. Wolf SoC goes to sleep as no computation is required while collecting data. The Nordic SoC is the master of the Serial Peripheral Interface (SPI) bus to read data directly from the Analog-to-Digital Converter (ADC). The acquired samples can be streamed using the BLE link embedded in the Nordic SoC. A typical application scenario for this mode is when BioWolf is configured to collect data for offline analysis. In *processing* mode, Mr. Wolf acts as the Master on the SPI bus and directly samples data from the ADC. The processing required by the BCI application is then executed on this energy-efficient ULP, while the other devices act as slaves. The processed signals can be transmitted using the BLE link through the Nordic SoC to bench-top devices such as PCs or smartphones. This mode is helpful in online BCI applications. The *deep-sleep* mode of the BioWolf system aims to minimise power consumption while the system is on standby. Near-field Communication (NFC) field is used as a wake-up strategy (Kartsch et al., 2019).

For the acquisition of EEG signals, BioWolf is coupled with a custom elastic EEG cap featuring eight dry soft electrodes from Idun Technologies (DRYODE^{TM} - IDUN Technologies, 2016). Each electrode includes an active circuitry (Guermandi et al., 2015) based on the AD8603 (Analog Devices), which improves the overall Common Mode Rejection Ratio (CMRR) and minimises the artifacts due to cable movement. Figure 4.2 shows the complete EEG cap implementing all the components introduced above.

4.2 Artifacts Preprocessing for Low-Density EEG based on BioWolf using ASR

Light-weight, minimally-obtrusive mobile EEG systems with a small number of electrodes (i.e., low-density) pose a higher risk for signal contamination with non-stereotypical artifacts due to hardware limitations and the challenging environment where signals are collected. A

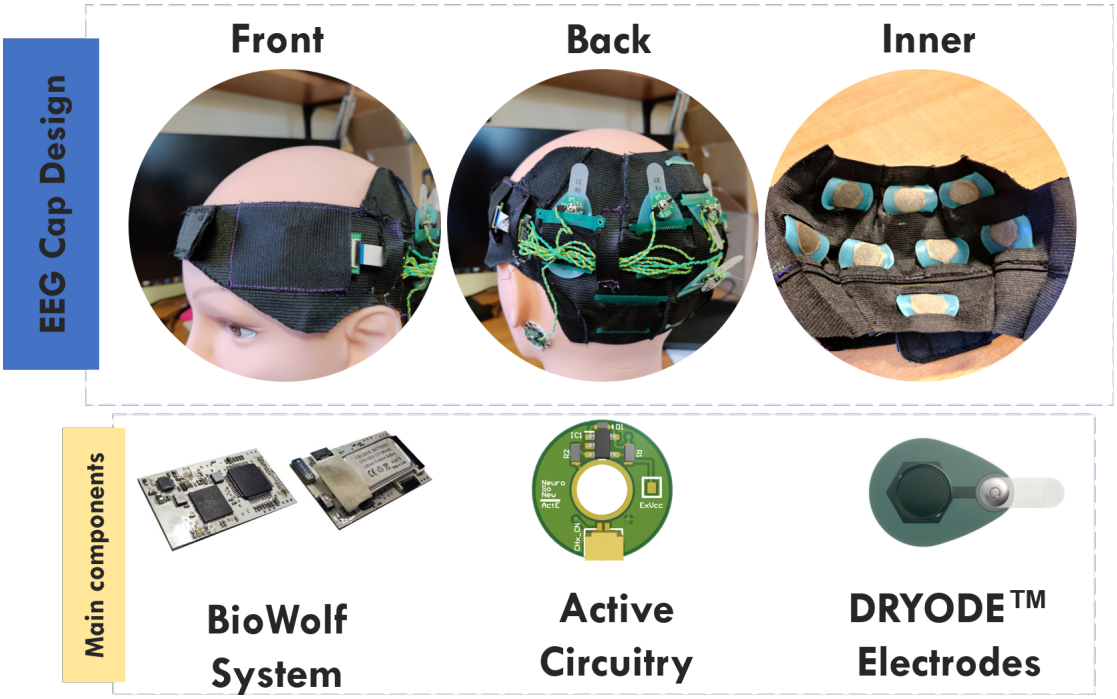


Figure 4.2: Portable EEG system embedding BioWolf and eight Dryode^{TM} electrodes featuring active signal buffering.

promising solution is Artifacts Subspace Reconstruction (ASR), a component-based approach that can automatically remove non-stationary transient-like artifacts in EEG data (Mullen et al., 2015). Since ASR has only been validated with high-density systems (Chang et al., 2018; Mullen et al., 2015), it is unclear whether it is equally efficient on low-density portable EEG. In this section, we evaluate ASR's cleaning efficiency on data collected using the BioWolf platform.

4.2.1 Experimental Setup

Six healthy subjects (aged 25-35 years) participated in data collection, which was performed in a controlled environment (dimmed light, silent, and apart from electrical interference sources). All participants reported no history of neurological or psychiatric disorders and provided written consent to participate in the experiments. Stimuli consisted of a black and white 10x10 square checkerboard presented on a uniform grey background at a distance of 80 cm from the subject's eyes, subtending a visual angle of approximately 15x15° and based on a sinusoidal on-off 100% contrast temporal modulation. Compared to abrupt on/off dynamics, such sinusoidal modulation reduces the number of harmonics (Norcia et al., 2015) while producing pleasant and less fatiguing visual stimulation. Additionally, a thin black diagonal cross overlapped with the checkerboard to help with visual fixation. All textures have been generated with Psychtoolbox 3.0.12 based on Matlab 2020b (MATLAB, 2020). Stimuli were

presented in a series of three trials of 25 seconds each at three different temporal frequencies (2Hz, 4Hz, 8Hz) in randomised order. The complete testing session contained six series alternating between *Artifact-Free* and *Artifact* conditions. During an *Artifact-Free* session, subjects were asked to relax (hence, minimising artifacts) while focusing on the stimuli. During an *Artifact* session, subjects actively induced artifacts by changing the seating position, moving their heads and repeatedly blinking while keeping their attention on the stimuli. They were asked to keep their visual fixation at the centre of the checkerboard for all trials. Before starting the experiments, the acquisition system (battery and triggers) and the electrode contact quality were checked. As the dry electrodes' flat surface only allows for proper skin contact when no hair is present, an electrode gel was used to adjust the contact quality of haired subjects when required.

4.2.2 EEG Data Analysis

Data analysis was performed with EEGLAB toolbox (Delorme and Makeig, 2004) and custom-made software based on MATLAB (MATLAB, 2018).

4.2.2.1 Preprocessing

EEG data were first low-pass filtered with a cut-off frequency of 40 Hz to remove the line and high-frequency noise. Subsequently, a high-pass filter was applied to remove low-frequency artefacts based on *clean_drifts* EEGLAB function (Delorme, 2018) (transition band: 0.15-0.3 Hz). Bad segments were processed by ASR using the dedicated EEGLAB plug-in (Delorme, 2018). To evaluate the aforementioned user parameters, we processed the data with k values ranging from 1 to 100 for both *ASR Correction* and *ASR Removal* processing modes.

4.2.2.2 Fourier Analysis

For each of the three stimulation frequencies, stimulus-specific trials were extracted, segmented in overlapping epochs of 10s (overlap factor varied between 1/2 and 3/4 of epoch length to adjust to the variable length of clean data segments), and Fourier transform F(f) of each epoch was computed using MATLAB's FFT function. The power spectrum was calculated from these Fourier coefficients as the average over epochs of the single-epoch power spectrum: $PS(f) = \langle F(f) \times F^{*}(f) \rangle_{ep}$

4.2.2.3 Stimulus-specific Response: FTR

To extract a stimulus-specific normalised measure, we computed the Frequency-tagged Response (FTR) as the ratio between the power spectrum at the stimulation frequency and the average of the power spectrum over the six neighbouring frequency bins (±0.3 Hz) (Buiatti et al., 2019). To test the performance of ASR in recovering a reliable stimulus-specific re-

sponse from the *Artifact* sessions, for each subject and stimulation frequency, we selected the electrode with the highest FTR in *Artifact-Free* sessions.

4.2.3 Experimental Results

4.2.3.1 Artifacts impact stimulus-specific EEG responses

As expected, artifacts severely deteriorate data quality and hamper stimulus-specific detection. We computed FTR on the band-pass filtered data to quantify this without ASR processing. As shown in Fig 4.3 (blue bars), the *Artifact* sessions produced a significantly lower FTR compared to *Artifact-Free* sessions. This effect appears at all three stimulus frequencies. It is worth mentioning that the average FTR for 2 Hz is around 1, which is equivalent to no stimulus-specific response.

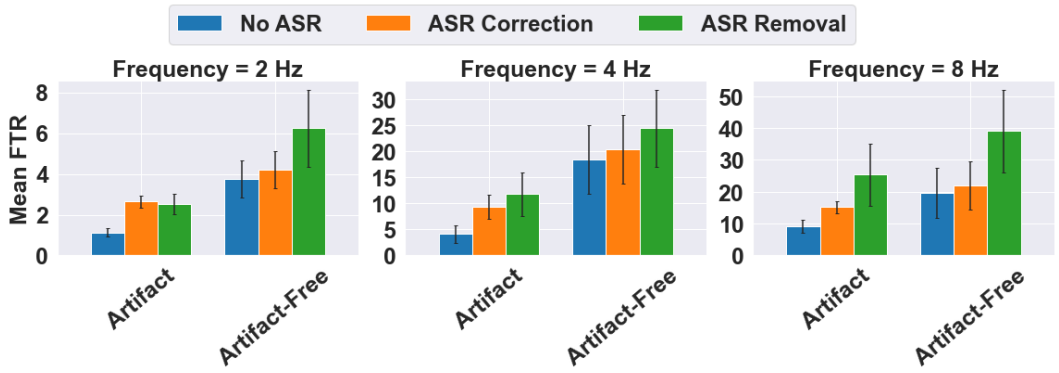


Figure 4.3: Comparison of average FTR (see section 4.2.2.3) across all subjects without ASR (blue), with ASR and best k for each subject for both ASR Correction (orange) and ASR Removal (green). Error bars show the standard error mean (s.e.m.) across subjects.

4.2.3.2 Overall Performance of ASR

For each stimulus frequency, we identified a range of ASR parameter k and processing modes that remarkably enhanced the FTR for all subjects in *Artifact* condition. Specifically, ASR performance was more successful in *Correction* mode for 2 Hz (18.7% FTR increase) and in *Removal* mode for 4 Hz (67.5% FTR increase) and 8 Hz (49.5% FTR increase).

One possible explanation for this frequency-dependent behaviour is that a 2 Hz stimulus-specific signal requires more samples to perform FFT than a 4 Hz or an 8 Hz signal. With *Removal*, some subjects do not have enough samples to compute FFT, leading to an average poor FTR score. However, with higher k-values (> 25) leading to a relaxed threshold, *Removal* performs similarly to *Correction* as evident in Figure 4.4 (blue vs orange).

The best performing k values ranged between 10 and 20 for 2 Hz and 4 Hz and between 15 and 40 for 8 Hz in their respective best processing mode. It is worth noting that ASR processing also improved the stimulus-specific response in *Artifact-Free* sessions, especially in *Removal*

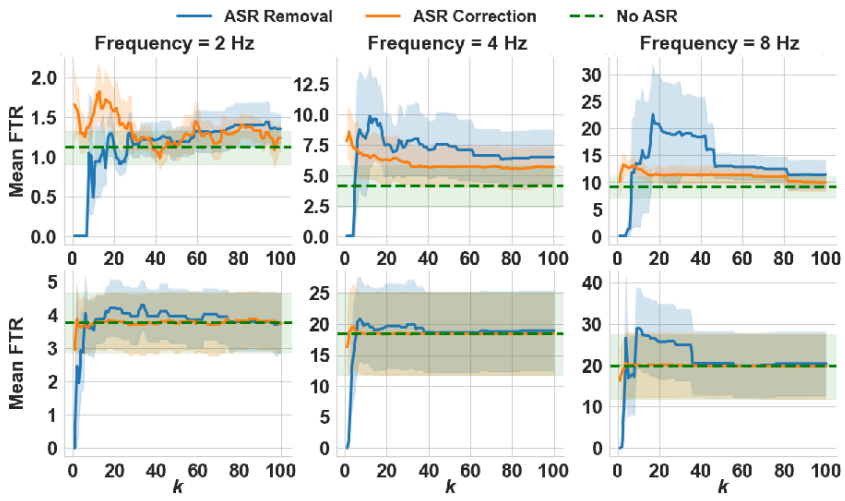


Figure 4.4: ASR Performance: FTR metric as a function of ASR parameter k and processing mode for *Artifact* (top panel) and *Artifact-Free* (bottom panel) Conditions.

mode and for 8 Hz [Figure 4.4 (bottom panel)].

4.2.3.3 Subject-specific ASR Parameter Optimization

To further investigate the variability of optimal k at the single-subject level, we studied the distribution of k-values that maximised the FTR for each subject and frequency in both processing modes. In line with the state-of-the-art guidelines (Chang et al., 2018), the distribution of k is concentrated between 1 and 40 in both *Artifact-Free* and *Artifact* conditions (Figure 4.5). We also observed that for *Correction* mode, optimal k values range between 1 and 20 for most subjects, while for *Removal* mode, the distribution is shifted to higher values starting from $k = 8$. This suggests that when opting for *Removal* instead of *Correction*, a slightly higher k is recommended to avoid too much data removal hence preserving sufficient data for post-processing.

For each processing mode, we then computed the average FTR of all subjects with their respective optimal k-values. As can be clearly seen in Figure 4.3, ASR significantly improved FTR (up to 45%) for all three stimulation frequencies in *Artifact* sessions.

4.2.3.4 Conclusions

In this study, we demonstrated that, by choosing appropriate processing parameters, ASR is efficient in removing artifacts and recovering significant stimulus-specific responses at multiple stimulation frequencies from low-density EEG data recorded by a portable ultra-low-power system. Additionally, we found that ASR performance is highly dependent on its k parameter and processing mode (correction/removal), so a poor choice might lead to a potential degradation in signal quality. Therefore, it is recommended first to establish the optimal k and processing mode on similar EEG data from the same EEG system (i.e., training

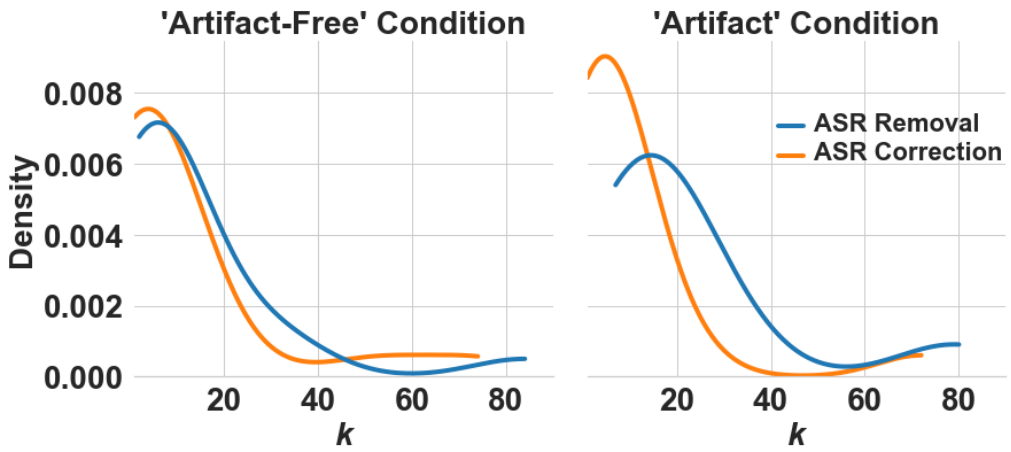


Figure 4.5: Distribution of single-subject optimal k values for *Artifact-Free* (left) and *Artifact* (right) conditions across all stimulation frequencies.

data).

We conclude that ASR is a reliable artifact processing technique in low-density, potentially real-time BCI systems that can measure neural response in diverse real-world applications and/or among clinical and pediatric populations. In the future, it is worth investigating ASR cleaning in newborn data collected using the BioWolf platform.

4.3 Neural Network Solutions for EEG Bad Channel Detection

Deep learning (DL) techniques have been successfully used in recent years to decode brain activities using EEG data (Lawhern et al., 2018). Most proposed methods use a hand-made preprocessing strategy to remove EEG noise before feeding the data into Neural Networks (NNs) (Lawhern et al., 2018; Mayor-Torres et al., 2021). In contrast, others do not explicitly mention the preprocessing performed. A dedicated NN-based preprocessing pipeline integrated with the downstream analysis pipeline might significantly improve system performance. Importantly, as there is an emerging trend in utilising NN-based solutions in resource-constrained hardware such as BioWolf (Kartsch et al., 2019), it is worth developing such solutions for EEG artifact removal in low-density setups. In this section, we present solutions for EEG bad channel detection. The first part of this section presents a simple Convolutional Neural Network (CNN) to classify good and bad channels. We started with a CNN-based approach because CNN's are known to capture spatiotemporal features with precision by using small size kernels. The second part presents a comparative analysis of four state-of-the-art NN architectures in terms of both classification accuracy and interpretability of learned features. The materials presented in this section are published in (Kumaravel, Paissan, and Farella, 2021) and (Paissan et al., 2022).

Layer	# filters	Shape	Activation
Input	-	(1, 62, 330000)	
Padding	-	(0, 0, 1, 1)	
Conv2D	8	(3, 256)	ReLU
Depth-wise Conv2D	32	(3, 3)	ReLU
Depth-wise Conv2D	64	(3, 3)	ReLU
Global Average Pooling	-	-	-
Dense		64 to 62	

Table 4.1: cleanEEGNet architecture overview. The table does not show batch normalisation performed between every convolutional layer and its activation.

4.3.1 CleanEEGNet

This section reports our preliminary results obtained with a 2D Convolutional Neural Network - cleanEEGNet for EEG bad channel detection. The model hyperparameters (such as kernel size and stride) are carefully chosen to mimic the conventional detection of bad channels performed via visual inspection.

4.3.1.1 Architecture

The proposed architecture comprises three convolutional blocks and one linear layer, as shown in Table 4.1. This architecture resembles the well-known EEGNet (Lawhern et al., 2018). As mentioned above, cleanEEGNet is tailored to mimic visual inspection. The temporal kernel of the first layer is set to 256 points, thus having a 0.5 s window with an overlap of 0.25 s with the previous one (stride at 128). The choices on these parameters are exploited based on: 1) knowledge from a domain expert who identifies bad channels via visual inspection and 2) temporal and spatial properties of the EEG signal. For example, the Pearson correlation of channels spatially located close to each other is crucial in detecting bad channels (Delorme, 2018; Delorme et al., 2007; Nolan et al., 2010). The rationale is that noise components in EEG are usually uncorrelated to components of neural origin; hence correlation is a reliable metric. For example, FASTER (described in Section 2.1.2) and CRD (described in Section 2.1.3) bad channel detection methods use inter-channel correlation features in their respective algorithms. Henceforth, we hypothesised that our network would learn to distinguish good and bad channels with the above hyperparameters.

After the first layer, the other convolutional blocks are composed of depth-wise separable convolutions to extract higher-order correlations without drastically increasing the computational burden (i.e., number of parameters). Subsequently, we perform an average on the time axis to avoid learning the time distribution of bad segments in the signal, followed by a linear classifier. Finally, we made the source code freely available on GitHub (Kumaravel and Paissan, 2021).

4.3.1.2 Experimental Dataset

Fourteen subjects (seven females) with a mean age of 23 years took part in the study that involved recordings on three different days, resulting in 130 EEG files (See (Schneider, Pereira, Tonin, and del R. Millán, 2019) for more details related to the experimental protocol). All datasets are made freely available by the authors in the OpenNeuro database (Schneider, Pereira, Tonin, and del R. Millan, 2019). However, for this work, we could use 113 files mainly due to technical issues importing European Data Format (EDF) files. We used 11 subjects (90 files which are 80% of the total files) for training and validation and 3 subjects (23 files, i.e., 20% of the total files) for testing.

EEG was recorded with a 64-channel HIamp EEG system (g.tec, Shiedlberg, Austria) at a sampling rate of 512 Hz referenced using the linked ears configuration. The electrodes were positioned in the international 10-10 system. EEG data were filtered at 40 Hz to remove the 50 Hz line noise interference using the default EEGLAB filter. Subsequently, high-pass filtering with the transition edge [0.25, 0.75] Hz was employed to remove DC drifts using the EEGLAB *clean_drifts* function (Delorme, 2018). Moreover, before feeding data to the neural network, we normalised it by using the average max amplitude among datasets, such that the majority of the values stay in the [-1, 1] range.

It is noteworthy that experts labelled 415 out of 7006 channels as bad, resulting in a 6:94 positive class vs negative class ratio. Given such an imbalanced proportion, we used the following unbiased metrics - Precision, Recall, Balanced Accuracy (bACC), and F1 score, which are defined as

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.2}$$

$$\text{F1 score} = \frac{2 * TP}{2 * TP + FP + FN} \tag{4.3}$$

$$\text{bACC} = \frac{1}{2}\left(\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right) \tag{4.4}$$

where TP, TN, FP, and FN indicate the number of true positives, true negatives, false positives,

Method	Precision	Recall	F1 Score	bACC
FASTER	0.1667	0.1699	0.1683	57.9
CRD	0.1959	0.2876	0.2331	61.69
HAPPE	0.1892	0.4268	0.2622	66.78
HAPPILEE	0.2968	0.3433	0.3184	65.02
cleanEEGNet	0.52	0.59	0.55	72

Table 4.2: Performance Comparison of cleanEEGNet Architecture

and false negatives, respectively.

As the proposed approach is the first work based on NN for this task, we chose four state-of-the-art automated conventional methods for comparison: FASTER (Nolan et al., 2010), Clean Raw Data (CRD) (Delorme, 2018), HAPPE (Gabard-Durnam et al., 2018), and HAPPILEE (Lopez et al., 2022).

4.3.1.3 Results and Discussion

We compared the classification performance of cleanEEGNet with the automated traditional approaches, and we found an improvement of 17% in balanced accuracy and 72% in F1 score, as summarised in the table 4.2.

Among the conventional methods, the most recent algorithm HAPPILEE performs the best with a maximum F1 Score of 0.32. A possible explanation could be that HAPPILEE integrates both HAPPE and CRD pipelines, thereby better estimating bad channels. Moreover, HAPPILEE uses optimal user-defined parameters calibrated using a set of 19 EEG files, whereas the other methods have not employed any calibration procedure.

4.3.2 Interpretable CNN-based Approach

To ensure the reliability of any NN model, it is important to understand how the network makes a prediction by visualizing the learned features or processing steps. Therefore, in this work, in addition to identifying the best performing NN architecture, we aimed to provide an interpretation of our neural network for bad channel detection by analysing the latent space features in the frequency domain. Unlike the previous section, here we deal with classifying single-channel artifacts aimed for wearable EEG devices. For this purpose, we artificially contaminated randomly chosen clean EEG channels with EOG or EMG artifacts in a controlled manner, as done in the EEGDenoiseNet evaluation (H. Zhang et al., 2021). This way, we generate the dataset to train the models in a supervised approach. Precisely, we have three classes: EEG (clean), EEG+EOG (EEG contaminated with ocular artifacts), and EEG+EMG (EEG contaminated with muscular artifacts). Since each kind of artifact (EOG or EMG) has its own spectral characteristics, this dataset provides a unique way to interpret the learned features by comparing the power spectrum of raw data and NN's learned filters. Further, we

changed the signal-to-noise ratio (SNR) of the training and testing samples between −7 dB and 4 dB to adapt to different contamination levels, sampled in physiologically plausible range (X. Chen et al., 2017; De Clercq et al., 2006; Wang et al., 2016).

4.3.2.1 Architectures

This study benchmarks four CNNs (one proposed, three existing architectures) for single-channel artifact identification. These architectures work on single-channel EEG time series; thus, it is possible to interpret their filters in the frequency domain, as explained later in this section. Moreover, working in one dimension reduces the computational cost of the architecture, which can be suitable for real-time online applications and resource-constrained wearable devices.

In this study, we propose a one-dimensional convolution to detect artifacts in single-channel EEG data. Afterwards, we tune the network topology to improve the filter's shape using a sinc layer (Bria et al., 2021). Sinc layers are convolutional layers in which the kernels are constrained to represent band-pass filters, where the lower and upper frequencies are learned for every feature map. In our case, the choice of the sinc layer was driven by (i) its good performance in the low-data regime, (ii) the easier interpretability of the filter power spectral density and (iii) low parameter count, which counts as an asset in the real-time and wearable-oriented domains. Moreover, a similar architecture was previously explored in the EEG domain and yielded promising results (Bria et al., 2021; Mayor-Torres et al., 2021).

The architectures referred to as 1D-CNN and SincNet in this work have the following layer sequence:

1. convolutional layer (sinc or standard, depending on the architecture);
2. batch normalisation and ReLU activation function;
3. global average pooling;
4. linear layer (2x);
5. softmax function;

To assess the performance of these two architectures in a more general way, we compared them with two baselines, a standard one for EEG processing with deep learning, namely EEGNet (Lawhern et al., 2018) and the network proposed in the paper presenting the dataset, EEGDenoiseNet (H. Zhang et al., 2021). Both baselines were developed and designed for different purposes with respect to artifact classification. Therefore, we adapted the approaches without compromising their advantages and drawbacks.

In particular, for EEGNet, we reduced the network input to single-channel EEG to comply with the dataset presented above. This guarantees that we can interpret the temporal filters

presented in the original EEGNet paper (Lawhern et al., 2018). EEGDenoiseNet (H. Zhang et al., 2021) does not present an encoder-decoder structure; thus, there are several ways to adapt the network for the task. Therefore, we tested the network's performance by adding a classifier at the end and removing one convolutional block at a time. The optimal network is the one that maximises the validation accuracy: a single convolutional block, namely two one-dimensional convolutions with a kernel width of 3, and the classifier. However, due to its low-kernel width (=3), this network is not considered for model interpretability analysis as there are not enough samples to compute FFT. The overall parameter count of this architecture is ~ 4000.

The hyperparameters of the neural networks, such as the kernel size and the number of filters, were optimised to maximise the validation accuracy. We randomly sampled SNR values during this optimisation stage to help the architectures generalise on different SNR ranges. Moreover, to fairly compare the architectures, we performed hyperparameter optimisation (HPO) with three techniques, namely random search, ASHA (Li et al., 2020) and Bayesian Optimization using the Orion toolkit (Bouthillier et al., 2021). Every HPO technique was run for a maximum of 200 iterations. Among the three optimisation strategies, the one that yielded the best validation accuracy is ASHA.

The neural networks were trained with a double-optimizer strategy. In particular, two Adam optimisers were used to (i) learn the weights of the convolutional backbone and (ii) fine-tune the fully connected layer weights. All architectures were trained for 100 epochs with early stopping to avoid overfitting.

With these models, our goal is to solve a 3-class classification problem (class 0: Clean EEG, class 1: EEG + EOG, class 2: EEG + EMG). Given that our dataset is balanced, we used the classification accuracy metric.

We performed Fourier analyis on the kernels to interpret the learned filters. Afterwards, we computed their Power Spectral Density (PSD) to interpret the most important frequency bands that every model uses for classification. As a ground truth reference, we performed PSD analysis on test data constructed using an SNR of -7 dB. As shown in Figure 4.6, EOG-contaminated samples result in a higher mean power in Delta [1-4] Hz and Theta [4-8] Hz bands with respect to the clean EEG samples. Instead, EMG-contaminated samples result in a higher average power in Beta [13-30] Hz and Gamma [30-120] Hz bands.

4.3.2.2 Experimental Dataset

We used a semi-synthetic EEG dataset to train, validate and test our proposed solutions: EEGDenoiseNet (H. Zhang et al., 2021). EEGDenoiseNet contains 4514 clean EEG channels, 3400 EOG (eye blink and eye movement artifacts), and 5598 EMG (muscular artifacts) channels. All of these are sampled at 512 Hz. These signals come from different subjects and are collected using different EEG acquisition setups. Since the considered dataset ensures universality in

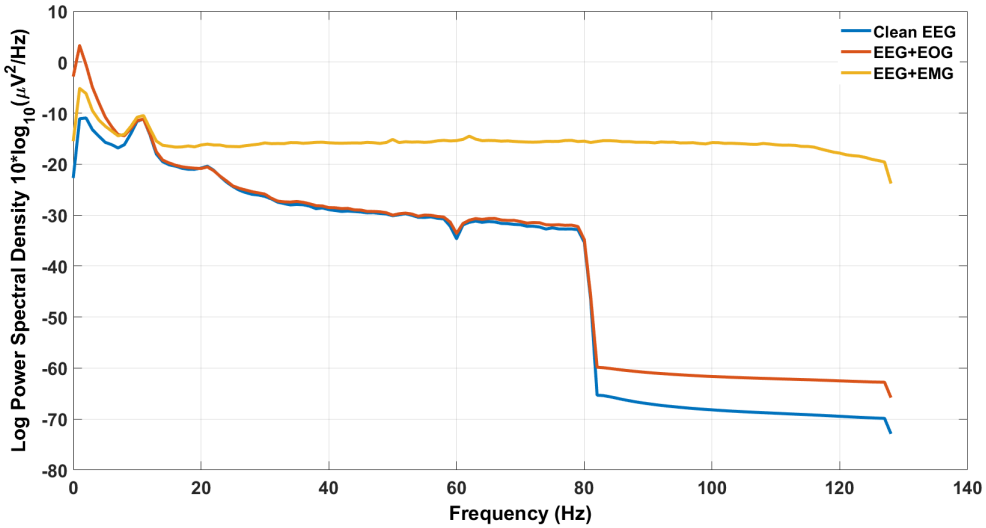


Figure 4.6: Mean log power spectral density computed on clean EEG channels (blue), EEG channels contaminated with EOG artifacts (red) and EEG channels contaminated with EMG artifacts (yellow). EOG artifacts demonstrate higher power in Delta and Theta bands, whereas EMG artifacts demonstrate higher power in Beta and Gamma bands.

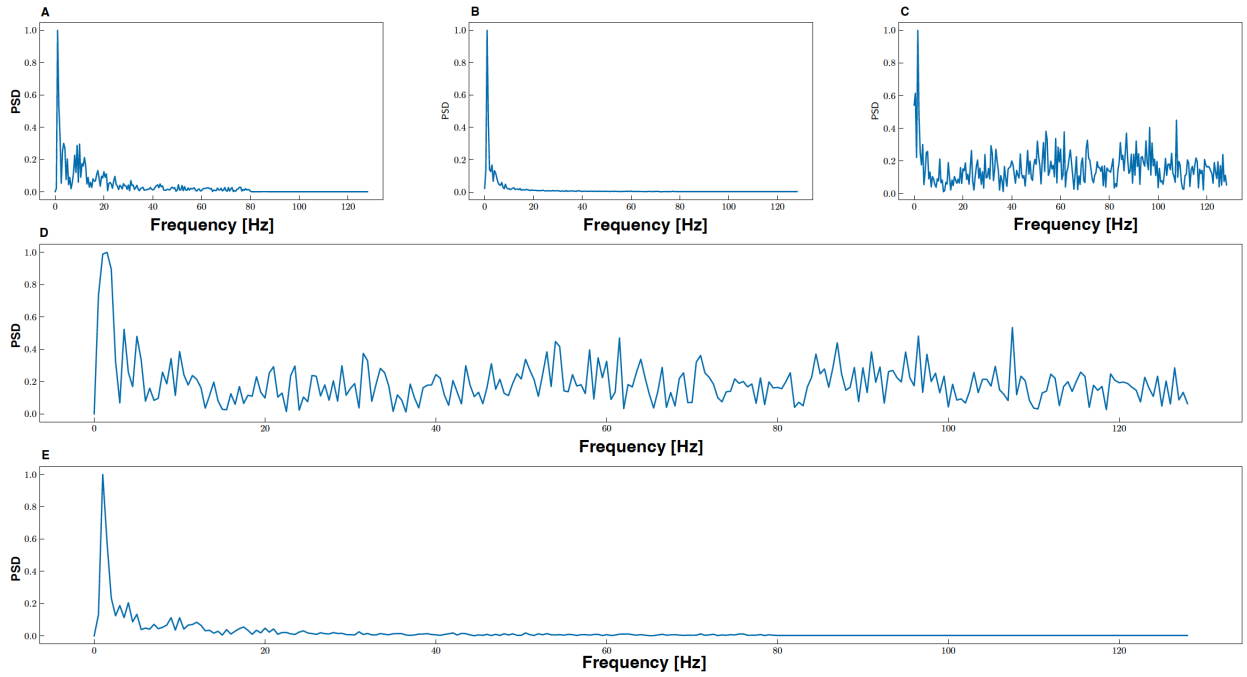


Figure 4.7: An Example of: (a) a clean EEG sample (b) an EOG sample (c) an EMG sample (d) a clean EEG sample contaminated with EMG artifacts and (e) a clean EEG sample contaminated with EOG artifacts

the acquisition setup, models trained on this dataset should perform well also on unseen data. A significant challenge in EEG artifacts processing is that most freely available EEG datasets lack the labelling of channels for "good" and "bad". In this work, this problem is overcome by artificially contaminating the clean EEG channels with either EOG or EMG artifacts using a deterministic approach.

Specifically, we generated a balanced dataset of clean and noisy channels by systematically changing the SNR values in the physiological range (from −7 dB to 4 dB) (X. Chen et al., 2017; De Clercq et al., 2006; Wang et al., 2016). Noisy (EOG or EMG, labelled as n) and clean (EEG labelled as x) channels are mixed using a linear combination in which the coefficients are tuned to respect a specific SNR by inverting Equation (4.5).

$$SNR[dB] = 20 \log \frac{RMS(x)}{RMS(\lambda n)} \tag{4.5}$$

Samples of the generated dataset are presented in Figure 4.7. The dataset was used for 50% for training, 20% for validation and 30% for testing. During training, noisy samples are generated with randomly sampled SNR values within the specified range. For testing, the samples are generated with specific SNR values to study the sensitivity of the neural architectures with respect to the artifacts contamination level. Before feeding the data to the neural network, we performed Z-score normalisation.

4.3.2.3 Results and Discussion

We present the results in two main categories: i) model performance and ii) interpretation of the learned filters (except for EEGDenoiseNet for the reasons mentioned in Section 4.3.2.1).

First, looking at the classification accuracy, we observed that three models, namely 1D-CNN, EEGNet, and SincNet, achieve a comparable performance (~ 99%) for the low SNR ranges between −7 dB to 4 dB. As the SNR increases, meaning that the data is less contaminated with artifacts, a steeper performance drop was observed for EEGNet than for 1D-CNN and SincNet. EEGDenoiseNet, on the other hand, reported the lowest classification accuracy of around 64%, and the performance drop with the increase in SNR was not significant (see Figure 4.8).

By performing the power spectral density analysis on the kernel features, we observed that 1D-CNN and EEGNet consider the lowest frequency bands (Delta [1-4] Hz and Theta [4-8] Hz) and the Gamma band ([80-120] Hz) to be necessary to classify EOG and EMG artifacts respectively. Instead, the SincNet considers the frequency band between 5 and 40 Hz significant for the classification. Considering the similar performance (in terms of accuracy) of SincNet with the other two models, one might interpret that SincNet focuses more on learning the neural features (EEG Alpha and Beta bands) rather than the artifact-related features (Delta, Theta and Gamma bands) as the other models (see Figure 4.9). We could not perform a similar analysis

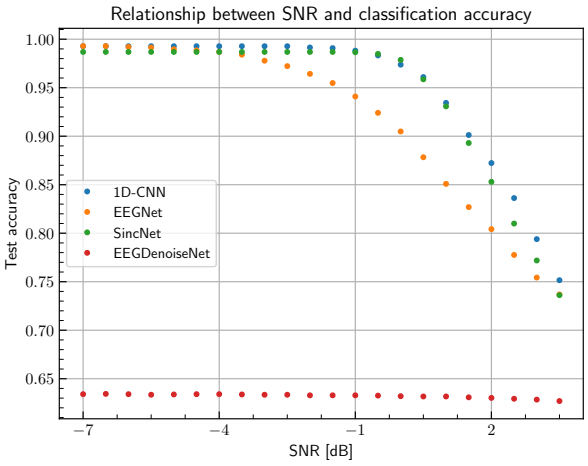


Figure 4.8: Comparison of bad channel classification accuracy with respect to signal-to-noise ratio (SNR) for 1D-CNN and SincNet models.

for EEGDenoiseNet because of the low kernel length (n=3), which is insufficient to perform PSD.

Further, we evaluated the number of training parameters required for each model to explore the possibility of porting these models on a resource-constrained wearable device (such as BioWolf (Kartsch et al., 2019)). While SincNet and 1D-CNN are the most parameter-efficient models, with only 227 and 1.6K parameters, the other networks have a relatively higher parameter count. Precisely, EEGNet has 3.1K parameters while EEGDenoiseNet has 3.5K.

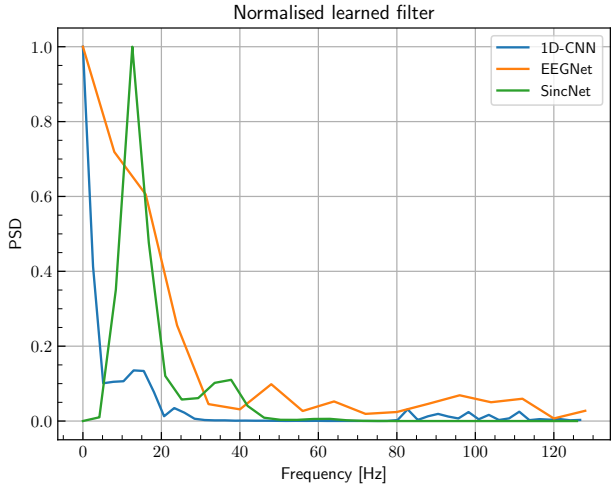


Figure 4.9: Frequency response of the learned filters of the considered models.

Artifacts are one of the common hindrances to interpreting brain functionalities using non-

invasive EEG technology. The most common artifacts in EEG are eye-movement and muscular movement-related artifacts. While there have been NN models used to detect the artifacts, they either were not evaluated for interpretability or were not validated on single-channel EEG signals. This work compares the performance of four neural network architectures for artifact classification in single-channel EEG with three classes: EEG, EEG contaminated with EOG, and EEG contaminated with EMG. In addition, we also evaluated and compared the models for interpretability, ensuring the models' performance reliability with unseen future data. While the models 1D-CNN, EEGNet, and SincNet performed comparably in classification accuracy, EEGDenoiseNet performed the worst. By analysing the learned features in the frequency domain, we found that both 1D-CNN and EEGNet separate EOG and EMG artifacts from EEG. SincNet, on the other hand, learns more of the EEG-specific information rather than the artifacts-related information, nonetheless achieving comparable performance. Due to its low kernel width, EEGDenoiseNet was not considered for model interpretability analysis. Among the considered models, both 1D-CNN and SincNet are parameter-efficient to be incorporated in a wearable EEG device for real-time artifact detection.

4.4 Frequency-Tagged Response Detection using Normalized Canonical Correlation Analysis (NCCA)

Frequency-tagging (FT) is widely used to test the integrity of sensory processing, especially in the visual domain (Steady-State Visually Evoked Potentials - SSVEP (Norcia et al., 2015)) and in the auditory domain (auditory steady-state responses - ASSR (Picton et al., 2003)). Typical presentation rates depend on the frequency ranges in which these sensory systems are most responsive: 8-20 Hz for SSVEP and around 40 Hz for ASSR. In these frequency ranges, it is possible to obtain very high SNR because the amplitude of both ongoing EEG activity and biological EEG artifacts is low.

Recent studies have extended the use of frequency-tagging to lower frequencies (0.5-7 Hz, encompassing the classical delta and theta frequency ranges) either because they focused on higher-level neural processing characterised by longer temporal scales (e.g. syllables and words in speech perception in adults (Buiatti et al., 2009) and infants (Kabdebon et al., 2014)), or because the experimental design was based on the infrequent presentation of critical stimuli among control ones (e.g. selectivity for faces among other visual objects in adults (Rossion et al., 2015) and infants (de Heering and Rossion, 2015)), or because neural processing is slow due to the immaturity of the visual system (e.g. face perception in newborns (Buiatti et al., 2019)). Obtaining reliable brain responses within this low-frequency range is more challenging because since frequencies are lower, longer presentations are needed to capture the related oscillatory responses. The interference of both EEG ongoing activity and artifacts increases for decreasing frequencies. Still, even at low frequencies, the spectral specificity of the stimulus-related response and the robustness to artifacts of frequency tagging makes it very well-suited for neuro-cognitive testing of special populations with limited attentional span like infants and patients (Kabdebon et al., 2022) and/or for out-of-the-lab applications.

However, it is currently unclear which is the best analysis tool for fast and reliable detection of SSVEP from low-density, wearable, wireless EEG recordings and stimulation in the low-frequency range used for higher-level brain responses and/or with infants and newborns (0.5-7 Hz), and what is the minimum data length necessary to obtain a such reliable response.

For this purpose, we recorded SSVEP responses of fifteen adult subjects presented with sinusoidal on-off checkerboard visual stimulation at four different frequencies, ranging from values typically used in BCI applications (7-10 Hz) (Kartsch et al., 2019; Y. Zhang et al., 2012) to values typical of high-level cognition and/or neural processing in newborns and infants (1-3 Hz) (Buiatti et al., 2009; Buiatti et al., 2019; Kabdebon et al., 2014). To better quantify the "peak" of the brain response at the stimulation frequency with respect to the wide fluctuations of the background EEG at low frequencies, we tested a frequency-normalized version of CCA (NCCA, Y. Chen et al., 2022; Zheng et al., 2019). We compared it with the more traditional normalised power spectrum (Norcia et al., 2015) in detecting the SSVEP response. To tackle the issue of the limited amount of data in special populations such as newborns, infants and patients, we systematically investigated for each stimulation frequency the performance of the two methods for progressively shorter time windows to estimate the minimum recording length necessary to extract a reliable SSVEP.

4.4.1 Experimental Protocol

Fifteen healthy subjects (aged 25-35 years; 14 males) participated in data collection, performed in a quiet office environment with dimmed light and away from well-known sources of electrical interference. All participants reported no neurological or psychiatric disorders record and provided written consent to participate in the experiments.

Stimuli consisted of a sinusoidal on-off 100% contrast temporal modulation black and white 10×10 square checkerboard, presented on a uniform grey background at a distance of 80 cm from the subject's eyes and subtending a visual angle of approximately $15 \times 15°$ (Montagna et al., 2017). A sinusoidal contrast modulation was used because it generates fewer harmonics (Norcia et al., 2015). Since the on–off dynamics is smooth, it is a more pleasant and less fatiguing visual stimulation than a squarewave stimulation mode for the subjects. Stimuli also included a thin black diagonal cross to help visual fixation. To extract frequency response information from each stimulation frequency's harmonics without interference with the other stimulation frequencies, We selected four incommensurable stimulation rates: 1 Hz, 3.125 Hz, 7.8125 Hz and 10.6125 Hz. Three trials per frequency (single trial duration = 25 seconds) were presented in a randomised order. A minimum of 10 s inter-trial rest was also included to reduce visual fatigue. Textures were generated with Psychtoolbox 3.0.12 based on Matlab (MATLAB, 2020).

Before each experimental session, hardware (battery and triggers) and electrode contact quality checks were performed. Electrode quality was adjusted using conductive gel when required.

4.4.2 Normalized CCA Index (NCCA)

CCA computes the linear dependency between two multidimensional variables by finding a couple of linear combinations, maximising their correlation, one for each multidimensional variable. Thanks to its robustness, CCA recently gained significant popularity in FT-based BCI applications (Lin et al., 2006).

Specifically, CCA retrieves a set of maximised correlations (canonical coefficients), each resulting from a couple of linear combinations belonging to mutually orthogonal subspaces. In the context of FT, the two multidimensional variables are the n EEG input channels and a set of m reference signals that identify the frequency of one single stimulus, usually sine and cosine of the target frequency and one or more harmonics.

Our CCA implementation relies on the Golub algorithm (Golub, 1969), which performs a QR decomposition step over the input and the reference signals, followed by an SVD factorisation of the product between the two orthogonal matrices. Euclidean Norm is then applied over the resulting canonical coefficients (size $d = min(n,m)$) to provide a single correlation value.

Given the intrinsic noisy nature of EEG signals collected with portable devices such as BioWolf (Kumaravel, Kartsch, et al., 2021), signals fed to the CCA core are first filtered. In this work, we applied a 10-tap IIR high-pass filter (Cut: 0.2 Hz, Pass: 0.4 Hz). No other preprocessing signal step (such as artifact correction/rejection) was performed. Notably, preliminary tests showed that the artifact removal preprocessing used for FTA (see Section 4.4.3) does not improve CCA performance. Figure 4.10 provides a block diagram of the CCA algorithm, including preprocessing steps.

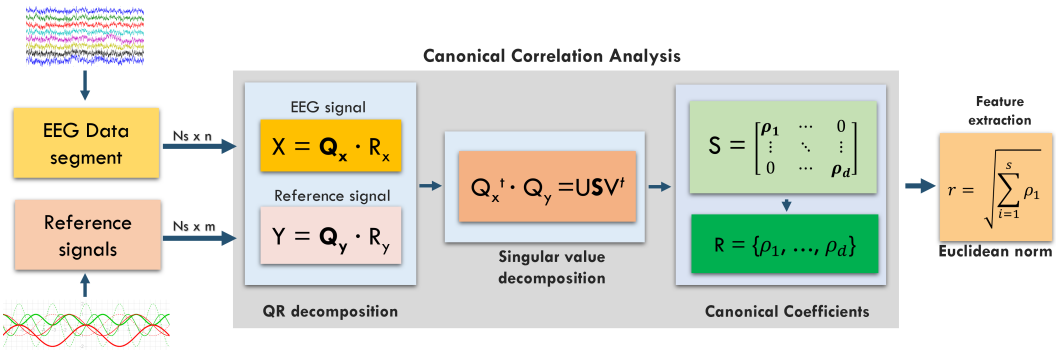


Figure 4.10: Block diagram for the CCA computation

NCCA is built on top of traditional CCA response aiming to provide an index of the presence of a "peak" at a given target frequency. NCCA is obtained by computing the ratio between the CCA response for a central (tag) frequency $Corr_{cf}$ and the mean value (often called background) between two adjacent frequencies $Corr_{cf+\Delta}$ and $Corr_{cf-\Delta}$(background), as denoted in the following:

$$NCCA_{tf} = \frac{Corr_{cf}}{mean([Corr_{cf+\Delta} Corr_{cf-\Delta}])} \tag{4.6}$$

In this work, NCCA is computed on a single segment length for each trial (typically three trials per stimulation frequency). Trial results are then averaged to obtain a single CCA value per subject and frequency. The background width of 0.4 Hz (Δ 0.2Hz from $Corr_{cf}$) was used for NCCA values reported in this work, empirically found to provide the best performance.

For illustrative and comparison purposes (as it will be introduced in the Figure 4.11), CCA was also employed to perform a frequency spectrum estimation (commonly achieved through the Fourier Transform). For this, CCA is computed for a range of frequencies of interest. The start of the frequency range and the ΔHz increments have been selected to precisely hit a tag frequency (typically at the centre of the frequency range).

4.4.3 Frequency Tagging Analysis (FTA)

Frequency Tagging Analysis (FTA) is a simple and effective technique based on Fourier analysis to estimate the frequency-tagged EEG response. FTA computes the frequency-specific neural response for each tag frequency by first performing a Fast Fourier Transform (*FFT* function, MATLAB) that precisely includes a frequency bin coinciding with the tag frequency by using a window length corresponding to a finite number of stimulation cycles. In this work, as a trade-off between the frequency resolution and the shortest window used, we set the window length to 4 seconds for 1 Hz stimulation (i.e., 4 cycles), 2.56 seconds for 3.125 Hz (i.e., 8 cycles), 1.536 seconds for 7.8125 Hz (i.e., 12 cycles) and 1.5 seconds for 10.6125 Hz (i.e., 16 cycles). For each stimulation frequency, EEG data from each trial were segmented in half-overlapping epochs of the corresponding window length, and the power spectrum of each electrode was computed as follows:

$$PS(f) = \left\langle F(f) \times F^*(f) \right\rangle_{ep} \tag{4.7}$$

F(f) is the Fast Fourier transform, and the average is computed across all the epochs belonging to a specific stimulation frequency. Since preliminary tests showed no improvement in channel selection, the power spectrum was averaged across all electrodes for subsequent analyses.

As for CCA processing, FTA requires a preliminary filtering step: the raw data is first low-pass filtered at a cut-off frequency of 40Hz (default EEGLAB filter (Delorme and Makeig, 2004; Widmann et al., 2015)) followed by a non-causal high-pass filter between 0.15 and 0.3Hz with a stop-band attenuation of 80dB.

Additionally, before FTA processing, data is preprocessed using the Artifacts Subspace Reconstruction (ASR) algorithm (Mullen et al., 2015) (as we observed poor performance in stimulus-specific response detection using the raw filtered data). For this, the default settings (k = 20, ASR Removal) are used, as they empirically yield better results (Chang et al., 2018; Kumaravel, Buiatti, and Farella, 2021; Kumaravel, Kartsch, et al., 2021). Trials with a duration of fewer than 20 s were discarded from further analyses.

4.4.4 Frequency-Tagged Response (FTR)

Readers may refer to Section 4.2.2.3 for further details.

4.4.5 Statistical Analysis and Sample Size

Both for NCCA and FTA, we computed the statistical significance of the peak (the frequency-tagged response compared to the background EEG) by performing a one-tailed Wilcoxon signed rank test (Hollander et al., 2013) between the CCA/power spectrum at the tag frequency and the average CCA/power spectrum at the neighbouring frequencies used to estimate the background EEG in the NCCA/FTR computations. We used the Wilcoxon signed rank test because it is a non-parametric test (Hollander et al., 2013) that does not require the assumption of normality, which is usually not guaranteed in the case of EEG data. To determine the minimal sample size required to observe the hypothesised effects reliably, we used an independent dataset (11 adult subjects) recorded with another EEG system (64 channels, Brain Products, Munich, Germany) with visual stimulation (black and white reversing checkerboards) and stimulation frequency range very similar to the ones used in this work. For more details about the EEG system, data pre-processing, and visual stimulation, see (Montagna et al., 2017), where part of this dataset has been analysed. We computed the effect size of the FTA (the standard measure used in this work) from the difference between the power spectrum at the tag frequency and the estimated background power at the tag frequency for a window of 20 s of data and a stimulation frequency of 4 Hz (i.e. in the lower, noisier part of the frequency range used in this work) on electrode POz of the independent dataset, which is located in the middle of the set of electrodes used in this work. The effect size was dz = 1.30. The minimal sample size, calculated with the software G*Power (Faul et al., 2007) (a-priori power analysis, Wilcoxon signed-rank test, matched pairs, one tail, α = 0.05), was 9. Therefore, the number of subjects (n = 15) used in this work is abundantly sufficient to observe the expected effect size.

4.4.6 Experimental Results

4.4.6.1 Both CCA and Power Spectrum detect stimulus-specific responses with "long" data segments

For long recording segments (20 s), both CCA and power spectrum show a clear stimulus-specific response at each visual stimulation in the form of a peak at the stimulation frequency (Figure 4.11 showing CCA (top row) and power spectrum (bottom row) at the four stimulation frequencies and, for a control comparison, at rest, for a representative subject). However, the two measures differ in the peak's sharpness and fluctuations of the resting state EEG. Especially for the two lower stimulation frequencies (1 Hz and 3.125 Hz), the power spectrum of the rest condition shows wide fluctuations; this effect combined with the low-frequency resolution of the power spectrum (0.25 Hz for 1 Hz tag frequency and 0.4 Hz for 3.125 Hz tag frequency) due to the trade-off between the number of cycles and window length (see section 4.4.3) and the 1/f neural noise leads to the potential spurious detection of peaks in the rest condition. On the contrary, CCA, which has the advantage of a virtually infinite frequency resolution, shows a sharp peak at all frequencies, even the lowest one, and a much smoother and flatter profile for the rest condition.

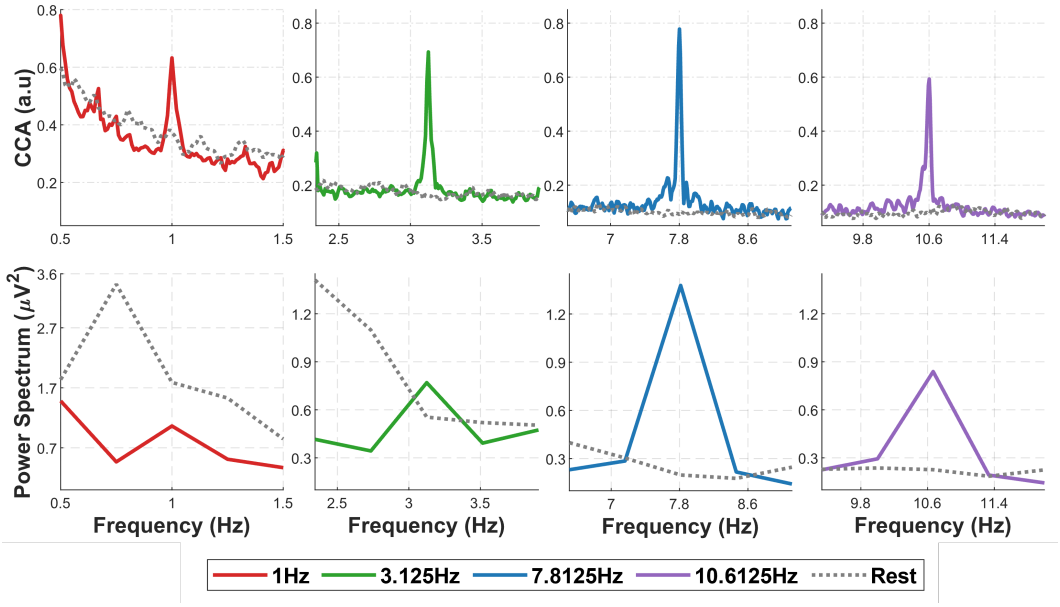


Figure 4.11: CCA (top row) and Power Spectrum (bottom row) of 20 s long EEG data segments relative to visual stimulation at (from left to right) 1 Hz, 3.125 Hz, 7.8125 Hz, 10.6125 Hz and on resting state data adjacent to the same segments, from a representative subject.

4.4.6.2 NCCA vs FTR detection performance in function of window length

We then tested the performance of both NCCA and FTR in detecting the stimulus-specific response at the stimulation frequency for progressively shorter recording time windows. The shortest window is 2 s for NCCA and 4 s for FTR (as the minimal window for detecting the peak at 1 Hz is 4 s, see section 4.4.3). Figure 4.12 shows the values reported by NCCA and FTR computed at each one of the stimulation frequencies on the segments corresponding to the four stimulation frequencies and on rest segments, averaged over all subjects. For increasing window length, NCCA rapidly departs from 1 (the value corresponding to no stimulus-specific response) specifically for the detector of the stimulation frequency, while its measures on the segments relative to stimulation at the other 3 frequencies, as well as on the rest segments, never increase and keep fluctuating around 1. This effect is clear for the NCCA at the three higher frequencies even from the shortest time window (2 s), while for the most challenging frequency (1 Hz), the effect starts to be evident from the 4 s time window.

FTR detectors are also successful in detecting the correct stimulus-specific response. Yet, they show a lower specificity, particularly for the lowest frequency (1 Hz). While FTR (1Hz) is higher than 1 for 1 Hz stimulation frequency even at the lowest time window (4 s), it is higher than one also for the other stimulation frequencies and rest segments. FTR detectors for higher frequencies better discriminate the stimulation frequency. Still, the 10 Hz one also slightly rises on the resting segments, probably due to difficulty distinguishing the stimulus-specific response from the spontaneous alpha peak.

These results are confirmed by testing the statistical significance of the peaks at each stimulation frequency (one-tailed Wilcoxon signed rank test between the CCA/power spectrum at the stimulation frequency and the CCA/power spectrum of the background EEG estimated from the neighbouring frequency bins, see Methods): for 1 Hz, the statistical value of the CCA peak crosses the threshold of $p = 0.01$ already from 4 s windows and rapidly reaches its maximum at 8 s, while one of the power spectrum peaks crosses the same threshold only at 6 s and increases very slowly for longer time windows. For 3.125 Hz, while both measures show high statistical significance from the shortest window, the CCA peak is already highly significant ($p \ll 0.01$) with 2 s of data. At the same time, FTR statistics improve slowly with the window length. This suggests that for a reliable stimulus-specific neural response at low stimulation frequencies, NCCA requires shorter data than the conventional FTA. Statistical significance for the detectors at the two higher frequencies is similar for the two measures, but NCCA is always at a plateau while FTR is occasionally lower.

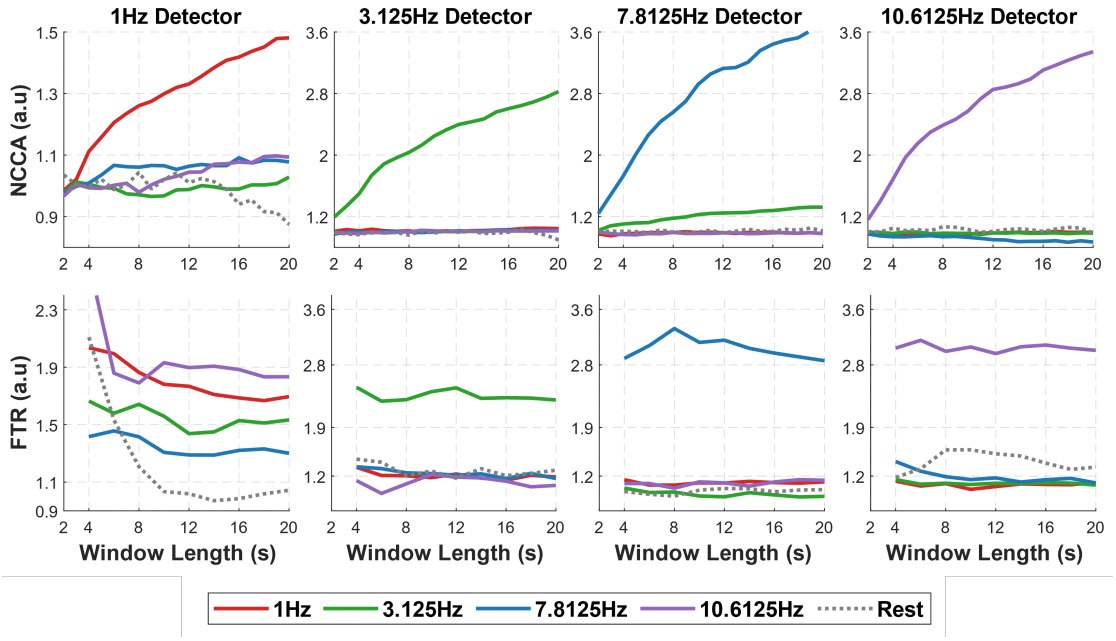


Figure 4.12: NCCA (top row) and FTR (bottom row) detectors at the four stimulation frequencies (from left to right: 1 Hz, 3.125 Hz, 7.8125 Hz, 10.6125 Hz) applied to data segments relative to the four stimulation frequencies and adjacent rest data, in the function of data window length.

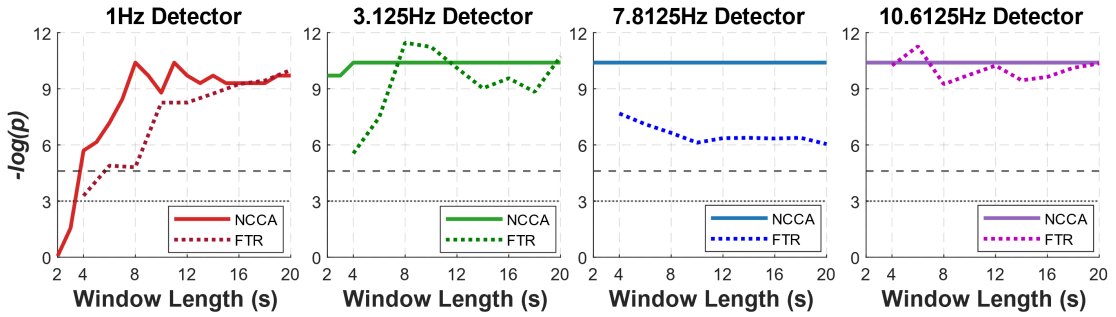


Figure 4.13: Statistical significance of NCCA and FTR (where p is the p-value of the one-tailed Wilcoxon signed rank test between the CCA/power spectrum at the stimulation frequency and the CCA/power spectrum of the background EEG estimated from the neighbouring frequency bins) at the four stimulation frequencies (from left to right: 1 Hz, 3.125 Hz, 7.8125 Hz, 10.6125 Hz) applied to data segments relative to the corresponding stimulation frequencies, in the function of data window length.

4.4.7 Discussion

Due to its reliability and high SNR, Frequency-Tagging design is widely used in a variety of experimental protocols, both in neuroscience research (Norcia et al., 2015) and clinical

applications like BCI (Chaudhary et al., 2016). The overwhelming majority of these studies use stimulation frequencies higher than 5-6 Hz (Norcia et al., 2015; D. Zhu et al., 2010) because they are based on the evoked potentials generated by the early visual system, which responds much more poorly at lower frequencies (Herrmann, 2001). Another line of research showed that lower stimulation frequencies (0.5-6 Hz) might be used to track the neural responses associated with higher-order processing like face perception (Buiatti et al., 2019; de Heering and Rossion, 2015; Rossion et al., 2015), spatial relations between social entities (Adibpour et al., 2021), biological movement (Cracco et al., 2022) and reading (Lochy et al., 2018), both in adults and along development (Kabdebon et al., 2022). However, all these latter studies recorded the data with non-portable high-density EEG systems in controlled laboratory settings.

Here, we demonstrate for the first time that it is possible to reliably detect frequency-tagged EEG response with a wearable, fully portable, wireless EEG with stimulation frequencies in the delta and theta frequency range down to 1 Hz in as rapidly as 4 seconds. To obtain this result, we tested NCCA, a frequency-normalized version of CCA, on data collected with a wearable, wireless EEG system based on BioWolf, a compact, ultra-light, ultra-low power recording platform (Kartsch et al., 2019), with four stimulation frequencies ranging from 1 to 10 Hz, and compared its performance with the traditional Fourier-based power spectral analysis (FTA).

We first observed that with a reasonable amount of data (20 s), CCA shows a clear, sharp peak over a shallow background for all the stimulation frequencies; the peak of the power spectrum is equally sharp for the two highest frequencies only, while it is increasingly lower and comparable to the background power fluctuations for the two lowest frequencies (Figure 4.11). These observations anticipate the performance of NCCA, which shows high specificity for the stimulation frequency (Figure 4.12) and high statistical significance of peak detection (Figure 4.13) at all stimulation frequencies, with data windows as short as 2 s for the three highest frequencies, and as short as 4 s for the lowest frequency. While FTR displays similar results for the two highest frequencies, and to some extent for 3.125 Hz, at 1 Hz, it fails to discriminate between different stimulation frequencies and has far lower statistical power than NCCA.

The reasons for the higher performance of NCCA compared to FTA likely rely on some key differences between the two methods that are particularly relevant for the low-frequency range analysed in this work.

1. Frequency resolution: Since FTA's frequency resolution is inversely proportional to the window length used to compute the FFT; it drastically decreases for decreasing frequencies, which is detrimental for the detection of a frequency peak because the sampling of the background EEG becomes poor; on the other hand, CCA frequency resolution is unbounded, as it only depends on reference signals, which can be generated at any specific frequency. This constraint in the choice of the frequency bins, together with the fact that the background EEG fluctuations increase with decreasing frequencies (1/f phenomenon), cause the FTR detector for 1 Hz to be much less efficient than the

NCCA one, as it is evident from the left-hand panels of the figure 4.11.

2. Channel selection: Since FTA is computed for each channel, merging the frequency response information across channels is not a trivial task, as less responsive channels may heavily smear the average response; on the other hand, CCA is a multivariable statistical method that optimally selects the relevant information from all the channels and provides a single response without requiring any preliminary channel selection.

3. Response at harmonic frequencies: CCA effortlessly incorporates the information in the tag frequency and its harmonics by using multiple reference signals, while FTA needs to be computed for each specific frequency, and averaging across harmonics may underestimate the overall response.

4. Artifact removal: Especially in the low-frequency range where artifacts are more relevant, FTA requires preliminary artifact removal, as previously shown on similar data (Kumaravel, Kartsch, et al., 2021) and confirmed by directly applying FTA on the raw data of this work (see Section 4.4.3 for details on pre-processing), which causes a reduction in data statistics; in contrast, CCA is computed on all the available data as it does not require any artifact pre-processing (preliminary artifact removal does not improve CCA performance, see Section 4.4.2).

The estimates of the minimal data length required for tagged-frequency response detection identified in this work have far-reaching practical consequences on the application of frequency tagging paradigms to special populations like newborns, infants and patients that typically have a limited attentional span (Kabdebon et al., 2022). For example, this means that we may be able to record a neural response associated to face perception in newborns (requiring the contrast between two different stimulations (Buiatti et al., 2019)) from as short as two EEG data segments of 7-8 s each during which newborns are attentive, a perspective suggesting potential implications for the development of related biomarkers to be recorded in out-of-the-lab scenarios.

These results demonstrate that low-density wearable EEG devices based on platforms such as BioWolf (Kartsch et al., 2019; Kumaravel, Kartsch, et al., 2021), typically used in BCI context with high stimulation frequencies (8-20Hz) (Arpaia et al., 2021; Belwafi et al., 2021; Salvaro et al., 2019), can also provide reliable signal quality even with very low stimulation frequencies (1 Hz). This denotes an excellent prospect for research and clinical studies on various neurocognitive functions in out-of-the-lab environments and/or with special populations such as newborns, infants or patients.

Further research is needed to test the efficiency of NCCA in developing or aging populations, which might challenge its robustness to artifacts. Also, exploring other stimulation modalities would test the possibility of extending it to other fields like speech processing (Buiatti et al., 2009; Kabdebon et al., 2014) or pain processing (Guo et al., 2020).

5 Conclusion and Future Work

5.1 Conclusion

In this thesis, I presented solutions for a crucial challenge in developmental EEG: Artifact processing. This includes introducing a novel algorithm for EEG bad channel detection (not restricted to developmental EEG) and adapting an existing, successful algorithm to neonatal and infant EEG for bad segments detection.

In chapter 2, we started with reviewing the existing algorithms (most of which are developed for adult EEG) and introduced the proposed algorithms: Local Outlier Factor (LOF) and Adapted Artifacts Subspace Reconstruction (ASR). LOF is designed as an EEGLAB plugin toolbox (see Figure B.1). We validated the performance of LOF using three real datasets: a newborn, an infant and an adult EEG. Considering the imbalanced proportion of good and bad channels in each dataset, we evaluated LOF and other existing approaches using a robust metric: F1 Score. We observed an unequivocal improvement in the performance of LOF with respect to other techniques in all considered datasets.

Then, as a preliminary analysis, we validated ASR in the newborn EEG dataset used in a previous study. In the literature, it was evident that ASR's performance depends on user-defined parameters, and a poor choice would degrade the quality of the signal. In this work, we proposed a standardised technique to identify the optimal ASR parameters for any population, including newborns and young infants. We noticed the success of ASR in recovering the stimulus-specific response (Frequency-tagging Response, FTR), which is comparable to the time-consuming manual cleaning. These preliminary results confirm that ASR can be adaptable and equally efficient in developmental datasets. The materials presented in this chapter are published in (Kumaravel, Buiatti, et al., 2022; Kumaravel, Farella, et al., 2022).

In chapter 3, we proposed NEAR, a novel artifact removal pipeline for human newborn EEG data. NEAR pipeline integrates LOF and Adapted ASR methods, along with default preprocessing steps such as filtering, interpolation of missing channels, etc. NEAR uses EEGLAB's functions to execute these additional preprocessing steps; as such, NEAR is compatible with

EEGLAB software. We validated the performance of NEAR using a simulated EEG dataset and two real datasets: a newborn and an infant EEG.

We used the SEREEGA toolbox to create neurophysiologically realistic continuous and event-related brain activity. We artificially contaminated a random set of channels and data segments, where the added noise is representative of newborn EEG (acquired from real newborn EEG). The ground truth for contaminated channels and segments is known. Therefore, NEAR can be validated faithfully. We generated two datasets simulating newborn EEG data with a frequency-tagging stimulation as in (Buiatti et al., 2019) and an event-related stimulus similar to the one in (Parise and Csibra, 2012), respectively. NEAR successfully recovered both datasets' neural responses (FTR and ERP) and performed better than the existing pipelines.

For the real datasets, the NEAR pipeline successfully reproduces established EEG responses from noisy datasets, with a higher statistical significance than the one obtained by existing artifact removal pipelines. The EEGLAB-based NEAR pipeline is freely available at https://github.com/vpKumaravel/NEAR. The materials presented in this chapter are published in (Kumaravel, Farella, et al., 2022).

Chapter 4 discussed the prospects of having a wearable EEG system to study the developmental populations and presented the current state-of-the-art ultra-low power research platform, BioWolf. This platform is developed in collaboration with the University of Bologna, Italy, and ETHZ, Switzerland. First, we evaluated the artifacts removal algorithm (ASR) in 8-channel data acquired using BioWolf from adult subjects. This study revealed that ASR could be equally efficient in removing the noisy portions of data (high sensitivity) and recovering the stimulus-specific neural response (high specificity), even in low-density setups. However, ASR is a computationally intensive algorithm unsuitable for resource-constrained platforms such as BioWolf. A separate section is dedicated to presenting our initial attempts at artifact-cleaning solutions using Neural Networks (NNs).

Then, we presented a neural feature extraction technique which is robust to the presence of artifacts in the FT-EEG data. We identified Canonical Correlation Analysis (CCA) as a promising technique. However, CCA has never been validated on low-frequency FT paradigms, commonly used in studies involving developmental populations. We proposed a novel index, frequency-Normalized Canonical Correlation Analysis (NCCA). As a normalised measure, NCCA is robust to noise compared to the state-of-the-art FTA. Moreover, our results demonstrated that NCCA is an efficient metric for detecting the neural response using a short amount of data (4s for a stimulus frequency at 1Hz). This promises future efficient out-of-the-lab neurocognitive testing in special populations such as newborns and infants. The materials presented in this chapter are published in (Kartsch et al., 2022; Kumaravel, Kartsch, et al., 2021; Kumaravel, Paissan, and Farella, 2021; Paissan et al., 2022).

In sum, this thesis work presented a set of methods for artifact removal and extraction of stimulus-related neural responses specifically adapted to newborn and infant EEG data that will hopefully contribute to strengthen the reliability and reproducibility of developmental

cognitive neuroscience studies, both in research laboratories and in clinical applications.

5.2 Future Work

The preprocessing pipeline (NEAR) proposed in this thesis is primarily designed for removing artifacts in newborn EEG (1-4 days old). Yet, after a systematic parameter calibration, we successfully validated its efficiency in a slightly older population (9 months old). To prove the adaptability of NEAR, further attempts are required to test its performance in artifact removal in young children and adolescents. Further, the proposed artifact removal methods presented in this thesis are general-purpose algorithms. They might help achieving an effective artifact cleaning irrespective of the experimental paradigm and the neural response. However, they often cannot meet portability, compactness, or computational-cost requirements. This section will discuss the possible future extensions of the work done, including techniques to support wearable EEG design for studying cognitive functions in newborns.

5.2.1 Application of NEAR for MEG Artifact Removal

Magnetoencephalogram (MEG) is a very similar method to EEG since the main sources of both kinds of signals are essentially the same, i.e. ionic currents generated by biochemical processes at the cellular level (Lopes da Silva, 2013). Further, MEG supports Frequency-tagging (FT) paradigm (Kalenkovich et al., 2022; Kamphuisen et al., 2008), a robust design for studying cognition in human newborns (Buiatti et al., 2019). Therefore, it is worth investigating if the proposed methods (LOF and ASR) and the NEAR pipeline are equally efficient in newborns/infants' MEG studies.

5.2.2 DEAR: An Extension to NEAR

One of the main reasons for the success of NEAR applied to newborn EEG is that it has advanced algorithms to deal with non-stereotyped artifacts (elicited due to head/facial/cable movements). The widely used ICA algorithm is very robust in capturing stereotypical artifacts such as eye blinks. However, given the lower eye blinking rates exhibited by human infants ($<$ 4 per minute), ICA may not be efficient for this population, as evident from the results obtained using MADE and HAPPE preprocessing pipelines. This low rate of blinking gradually increases to about 15-30 per minute in adulthood (Bacher, 2010). Therefore, it might be useful to integrate ICA with the NEAR pipeline (right after the ASR processing step in the NEAR outline (see Figure 3.1)) for much older pediatric populations, such as children, to explicitly deal with residual eye blink and other stereotypical artifacts. With this integration, NEAR can be extended to Developmental EEG for Artifact Removal (DEAR), covering all age ranges starting from newborns to children.

5.2.3 Neural Network Solutions for Artifact Removal

In this thesis, we discussed the preliminary NN-based solutions for the EEG bad channel detection problem, which is the preliminary step of artifact removal. While the first work (using cleanEEGNet architecture) focused only on classification performance, in the second work, we provided an interpretation of the learned features to gain credibility for the networks' prediction. However, much work is required for reliable overall artifact removal, and I foresee the required steps in two directions: First, optimising existing architectures to further improve the classification performance. Second, developing new architectures to automatically detect noisy portions of EEG and reconstruct them to reduce the noise. For the latter, it is worth investigating Encoder-Decoder Neural Networks (Cho et al., 2014), which are commonly employed for signal denoising (Fotiadou et al., 2020). Another noteworthy architecture is the U-net convolutional neural network, which has been successfully employed in image and speech signal denoising (Choi et al., 2021; Heinrich et al., 2018; Moliner and Valimaki, 2022; Tripathi, 2021). However, the neural networks require a significant amount of data (with annotation for artifacts), which is rare to find in the EEG community. It gets further difficult with developmental populations, given that acquiring datasets from newborns/infants is comparatively less frequent. Therefore, for solid progress in this direction, the researchers are encouraged to share their data as much as possible along with labelled artifacts.

5.2.4 Hardware/Software Co-design

Traditionally, hardware and software units are designed and developed as independent sections and later integrated as a single system to reach a specific application goal. Due to the increasing demand for achieving high performance while maintaining low energy consumption, an optimised concurrent design of hardware and software sections was proposed over a decade. This design approach is termed hardware/software co-design. Knowing the hardware architectures, their limitations and advantages make it possible to develop algorithms that can achieve high parallelism and even better performance. Likewise, knowledge of algorithms and their complexity can be helpful to accelerate the processing capabilities of processors, if required. Such unified systems are ideal solutions to meet portability requirements.

In this thesis, we carried out the former approach, focusing on the software part (artifacts removal and neural feature extraction) and attempting to fit the developed software into an embedded wearable EEG (BioWolf platform). This is because well-established algorithms did not exist specifically for newborn/infant EEG. As such, it was essential to understand the nature of the signal and the noise beforehand. As a result, there were challenges in adapting the computationally intensive algorithms (LOF and ASR) to fit into the portable EEG device (BioWolf). Nevertheless, the outcomes of this thesis provide unique knowledge about the kind of algorithms that one might consider for handling short and heavily contaminated newborn EEG, both to remove noise components and extract neural features of interest. For example, to study visual cognition/perception in human newborns, one might require a subset

of electrodes (primarily placed in Visual Cortex, see Figure 2.5 (top onset)) rather than using high-density commercial wearable devices. With the reduction in the number of electrodes, we could achieve a low-weight wearable design and reduced computational load without necessarily compromising performance.

5.2.5 Paradigm-specific Artifact Removal

Any unwanted signal spuriously affects the signal of interest is called noise/artifact. As such, the definition of noise depends on the kind of experimental paradigm and the expected neural feature. For example, for a typical ERP analysis, the background noise (irrelevant to the presented stimuli) can be cancelled by simply averaging some stimulus-specific experimental trial data. Instead, since the neural feature manifests as a sharp narrowband peak in the spectral analysis for the FT analysis, the broadband artifactual components can easily be isolated. Therefore, designing noise removal algorithms specific to a given experimental paradigm is beneficial. This requires extensive characterisation of predominant artifacts within a given experimental design. Indeed, the drawback of such an approach is the lack of generalisation of the developed algorithms. However, such highly stimulus-specific artifact removal techniques often offer reduced complexity, thereby accommodating high-quality wearable EEG systems. Since FT-based EEG acquisitions are increasingly used in newborn/infant cognitive studies, it might be worth developing artifact removal strategies specific to this paradigm.

5.2.6 Dual-layer EEG Approach

A recently emerging hardware and software design for wearable applications is the dual-layer EEG approach. In this design, mechanically coupled and electrically isolated inverted noise channels are used in addition to scalp EEG channels. Usually, the data acquired by these noise channels exhibit a higher correlation with artifacts (due to cable/electrode movements or motion noise) than its correlation with neural information. This approach was initially validated in a phantom head study (Nordin et al., 2018), where noise cancellation with dual electrodes demonstrated a successful reconstruction of simulated EEG signals. Subsequently, other studies utilised the same approach on real human EEG data while participants walked on a treadmill, navigated obstacles or played table tennis. Since the movement artifacts (or motion noise due to sucking, frowning or crying) are the predominant artifacts present in newborn /infant EEG, a dual-electrode (scalp EEG and noise) design and the associated artifact removal strategies might drastically improve the overall signal quality.

A Appendix: NEAR User Manual

This appendix contains the technical details of each of the user-defined parameters to be defined for NEAR preprocessing.

A.1 List of Parameters

A.1.1 Basic Parameters

The basic ones are just to enable and disable different steps in processing pipelines. For example, **isLPF** (see Table A.1 for the list of basic parameters).

A.1.2 Advanced Parameters

The advanced ones are detailed parameters for each of the enabled preprocessing steps. For example, if the user has enabled lowpass filtering by setting **isLPF** to 1, then it is important to define the relevant parameter **lpc** (low-pass cut-off) with a value (e.g., 40 Hz). See Table A.2 for the list of basic parameters.

A.2 Single Subject Processing

For performing NEAR preprocessing for a single subject, the users can open the file *NEAR_singlesubject_processing.m*. The parameters should be set as per the application requirements, and the function run_NEAR.m will process the given data using the parameters. For the downstream analysis, such as ERP analysis, the users can write their own code at the end of the script.

ID	Parameter Name	Parameter Description	Possible Values
1	isLPF	is Low Pass Filter (to be applied)?	0 to apply a Low Pass Filter; 1 otherwise.
2	isHPF	is High Pass Filter (to be applied)?	0 to apply a High Pass Filter; 1 otherwise.
3	isSegt	is Segmentation (to be applied)?	0 to segment the data; 1 otherwise.
4	isBadCh	is Bad Channel (to be detected)?	0 to detect bad channels; 1 otherwise.
5	isVisIns	is Visual Inspection? (for bad channels)	0 to see channel-wise statistics; 1 otherwise.
6	isBadSeg	is Bad Segmentation (to be detected)?	0 to detect bad segments; 1 otherwise.
7	isERP	is your data an ERP?	1 if you want to epoch the data; 0 otherwise.
8	isInterp	is Interpolation (to be performed)?	0 to perform interpolation; 1 otherwise.
9	isAvg	is average re-referencing (to be performed)?	0 to perform re-referencing; 1 otherwise.
10	isReport	is Report function enabled?	1 to save a report file; 0 otherwise.
11	isSave	is Save function enabled?	1 to save the processed data; 0 otherwise.

Table A.1: List of Basic Parameters

A.3 Batch Processing

If the users would like to apply NEAR preprocessing for a batch of EEG files, it can be done so by using the file *NEAR_batch_processing.m*. Just like the single subject processing, the parameters can be set quickly, and for each file if a downstream analysis is preferred, appropriate scripting can be done within the for loop.

A.4 Report and Saving Functionalities of NEAR

By setting the parameters isReport and isSave to 1, the users can have their processed datasets along with a comprehensive summary of the preprocessing steps performed. The files can be found in the folders "NEAR_Processed" and "NEAR_Reports" for processed .set files and report files respectively. In addition, the Local Outlier Factor (LOF) values for each of the datasets will be stored in the folder "NEAR_LOF" for future reference.

A.5 Hyperparameters Tuning

There are 2 main hyperparameters to be calibrated on the user's data (by conventional 70-30 training-testing split or a k-fold cross-validation scheme). They are:

ID	Parameter Name	Parameter Description	Possible Example Values
1	lpc	Low Pass Cut-off	40 Hz
2a	hptf	High Pass Transition Frequency	[0.25 0.75] Hz
2b	hpc	High Pass Cut-off Frequency	0.1 Hz
3a	sname	Segmentation File Name (in .xlsx)	segt_visual_attention.xlsx
3b	sloc	Segmentation File Location	Absolute Path of .xlsx file
3c	look_thr	Looking Times Threshold	5000 ms or if no threshold preferred
4a	isFlat	is Flat-line channels (to be detected)?	0 or 1
4b	flatWin	Tolerance window length	5 s
4c	isLOF	is LOF algorithm (to be applied)?	0 or 1
4d	dist_metric	Distance Metric for LOF algorithm	'euclidean' or 'seuclidean'
4e	thresh_lof	Threshold for LOF algorithm	1.5
4f	isAdapt	thresh_lof is incremented by 1 if the total % of bad channels exceeds this limit	10%
4g	isPeriodogram	is Periodogram analysis applied?	0 or 1
4h	frange	Frequency Range to be considered	[1 10] Hz (motion artifacts)
4i	winsize	window length for periodogram	1 s
4j	winov	window overlap length	0.66 s
4k	pthresh	Threshold for Periodogram analysis (in terms of SD)	2
6a	rej_cutoff	ASR Parameter (k)	20
6b	rej_mode	ASR Mode: Correction or Removal	'off' for correction or 'on' for removal
6c	add_reject	Additional Rejection by ASR?	'off' or 'on'
7a	erp_event_markers	Event Markers	'eyes open', 'eyes close'
7b	erp_epoch_duration	Epoch Duration in seconds	[0 1.2] s
7c	erp_remove_baseline	Remove baseline	1 if you want to; 0 otherwise
7d	baseline_window	Baseline Window in ms	[0 200] ms
8	interp_type	Type of Interpolation	'Spherical','v4', 'spacetime'
9	reref	Re-reference electrode (in case isAvg = 1)	[10] or 'E129'

Table A.2: List of Advanced Parameters. Note: The number in the ID is linked to the ID in A.1 to infer the relationship.

1. LOF Threshold (to be used in NEAR Bad Channel Detection)

2. ASR Parameter (to be used in NEAR Bad Segments Correction)

In addition to NEAR custom scripts, we provide basic scripts found in "TuneLOF" and "TuneASR" folders in this repository. While TuneLOF uses the most common metric F1 Score, TuneASR requires users to define a measure of interest namely ERP-SNR, or FTR relevant to the datasets considered in this work.

Note: For TuneLOF, we also provide a sample EEG file along with labelled bad channels found in the openNeuro website: https://openneuro.org/datasets/ds002034/versions/1.0.1

B Appendix: NEAR Step-by-step Tutorial

In this appendix, we will see how to execute the NEAR pipeline (described in Chapter 3) using a sample data. Throughout this tutorial, we will see the GUI interfaces only for the core blocks of NEAR (such as bad channels detection using LOF and bad segments rejection using ASR). For the rest of the blocks, the EEGLAB-based functions are provided. Depending on the application requirements, the users can add or remove functionalities or modify the existing ones. To get familiarized with this flexible pipeline, the readers may use the executable MATLAB script NEAR_PIPELINE_TUTORIAL_V1_0.m (available on the GitHub repository (https://github.com/vpKumaravel/NEAR), referred as "tutorial script" throughout this appendix) for a hands-on experience. Before running the tutorial, please be sure to clean the existing variables and open a fresh EEGLAB GUI. By executing the following commands:

```
clc; clear all; eeglab;
```

B.1 Defining the Dataset Name and Location

First, we define the dataset to be preprocessed using NEAR as follows:

```
dname = 'EEG1.set';% dataset name with extension
dloc = 'D:\\where\\is\\the\\data';% corresponding file location
chanlocation_file = ...
  '   \\eeglab2021.0\\sample_locs\\GSN-HydroCel-129.sfp';% channel ...
  location file location, if applicable
```

B.2 Setting the User-defined Parameters

A user manual is available on the GitHub repository (presented in Appendix A or downloadable by clicking on this link), in addition to the in-code comments provided in the MATLAB tutorial script will be useful in setting the user-defined parameters. Once these parameters are set, the

pipeline can be executed as an automatic version.

B.3 Importing the Data into EEGLAB

The current version of the NEAR pipeline supports importing files of four frequently used formats in the context of developmental EEG namely, .mff, .raw, .set, and .edf. For other formats, the users may replace the relevant commands in step 1 (in the tutorial script) accordingly.

1. **To import .mff file:** Please ensure to have the plugin "MFFMatlabIO3.7" in your EEGLAB environment. Click on File -> Import Data -> Using EEGLAB functions and plugins -> Import EGI .mff file. Alternatively, the following command would do:

   ```
   EEG=mff_import([dloc filesep dname]);
   ```

2. **To import .raw file:** Please be sure to have the latest EEGLAB version (2021). Click on File -> Import Data -> Using the File-Io Interface -> File (button in the pop-up). Otherwise, you may simply execute the following command:

   ```
   EEG = pop_fileio([dloc filesep dname], 'dataformat','auto');
   ```

3. **To import .set file:** Click on File and then Load existing dataset. Alternatively, the following command can be executed:

   ```
   EEG = pop_loadset('filename',dname,'filepath',[dloc filesep]);
   ```

4. **To import .edf file:** Please be sure to have the latest EEGLAB version (2021). Click on File -> Import Data -> Using the BioSig Interface. Alternatively, you may simply execute the following command:
   ```
   EEG = pop_biosig([dloc filesep dname]);
   ```

B.4 Importing the Channel Locations

To import channel locations, click on Edit and then Channel Locations in the EEGLAB GUI interface. In the pop-up window, click on Read Locations and choose the relevant channel location file (usually, .sfp files) for importing. Alternatively, the following command will work:

```
EEG=pop_chanedit(EEG, 'load',{chanlocation_file 'filetype''autodetect'});
```

N.B: This step shall be ignored if the imported file has the channel locations already.

B.5 Reshaping the Epoched into Continuous Data

This step is not required if the input data is continuous. Even though NEAR supports epoched data (such as ERPs) for pre-processing, if the input data is not continuous, ASR might not produce optimal cleaning results. Therefore, the users are advised to convert the epoched into continuous data by running the following command-line code:

```
EEG = eeg_epoch2continuous(EEG);
```

Once the data is converted to continuous, please ensure to set the parameter isERP to 1. By doing so, the event related epoching will be done after ASR processing. `isERP = 1;`

Tip: Once the data is imported, make a copy of the EEG Struct which can later be used for interpolation of removed channels or other statistical comparisons.

B.6 Band-Pass Filtering

Based on the application requirements, users can set the high-pass and the low-pass cut-off frequencies. Additionally, a notch filter to reject the 50 Hz/60 Hz can be implemented.

To low-pass filter the data, click on Tools and then Filter the data > Basic FIR filter (new, default) and insert the cut-off frequency in the text box “Higher edge of the frequency pass band”. Alternatively, execute the following command:

```
EEG = pop_eegfiltnew(EEG, [], lpc, [], 0, [], 0);
(Or)
EEG = pop_eegfiltnew(EEG, 'hicutoff', lpc);
```

where *lpc* is the low pass cut-off frequency in Hz. To high-pass filter the data, either use the GUI as mentioned before and insert the cut-off frequency in the text box “Lower edge of the frequency pass band”. Alternatively, run one of the following commands:

```
EEG=clean_drifts(EEG,hptf, []);
(Or)
EEG = pop_eegfiltnew(EEG, 'locutoff',hpc);
```

where *hpc* and *hptf* are the high pass cut-off frequency and high pass transition frequency respectively (both in Hz).

B.7 Events Segmentation

For any kind of experimental paradigm apart from resting-state EEG (for which no segmentation is typically required), users may need to select the portions of the data during which stimulation was present and during which newborns/infants attended to the stimulation before applying NEAR. This step (to be run from the command line only) is particularly recommended for paradigms using visual stimulation.

Here we describe instructions for the newborn datasets used in this work. Other cases can be easily dealt with similar methods.

Segmentation based on looking times: There are several ways to import the looking times. Here we propose to create an excel file with one sheet per subject, where the name of the sheet indicates the subject name. A sample file can be found in the GitHub repository. Once the file is prepared, the following commands can be executed to extract the looking times.

```
lookFile=importdata([segt_loc filesep segt_file]);
```

where *seg_loc* is the location and *segt_file* is the name of the excel file.

```
lookTimes=NEAR_getLookTimes(lookFile,sub_name, look_thr);
```

where *sub_name* is the sheet name (and the subject name) in the excel file lookFile, *look_thr* is the minimum duration (in ms, example 5000 ms) to accept as a valid segment.

Note: If you do not wish to impose a threshold, you may do so by setting look_thr = [].

Then to retain only the valid segments, the following command can be executed:

```
EEG = pop_select(EEG,'time',lookTimes);
```

The above command can be adapted to resting-state EEG analysis as well, by assigning the variable lookTimes with intervals of good segments known apriori. Instead, if the user has a list of bad segment intervals (which is more practical), the same command can be modified to:

```
EEG = pop_select(EEG,'notime',badSegments);
```

B.8 Detecting Bad Channels using LOF

Once the NEAR plugin is installed in EEGLAB (see instructions in GitHub ReadMe file), NEAR bad channel rejection tool can be found in the "Tools" Menu with the name "NEAR – Channel

Rejection".

Figure B.1: NEAR – Channel Rejection Tool Input GUI

By default, the two methods “Remove Flatlines” and “Remove Outliers (LOF Method)” are selected. As an optional method, a spectral analysis method is available (built with the MATLAB Periodogram method) denoted as “Detect Motion Noise (Spectral Analysis)”. The parameters displayed in the GUI are calibrated using our Training Dataset (see “NEAR parameter calibration” in chapter xx). We suggest the users to perform a preliminary analysis to get an intuitive understanding on the role of parameters. On click of the "Ok" button, the tool first detects the flat-line channels and flags them (if any) as bad channels; then LOF (and Periodogram, if applicable) is executed on the residual channels.

Note 1: For calibrating LOF cut-off, we have provided a dedicated script along with a sample dataset with ground truth bad channels. The script and the dataset can be found in our GitHub repository under the folder TuneLOF.

Note 2: LOF uses the Parallel Toolbox feature of MATLAB for a faster computation. MATLAB might take a few seconds to initialize the process when running for the first time. Once computations are performed, the following messages are displayed in the command window, summarizing the set of actions and the relevant outcomes. It is important to note that the displayed channels are not removed from the data but merely marked as “bad” when running by GUI.

Additionally, the following plots are displayed to help the users to take a decision on the removal of detected bad electrodes:

1. A histogram showing the computed LOF scores for each “non-flat-line” channel.
2. A histogram showing the measured power in the selected frequency range for each channel (if “Detect Motion Noise” is selected).

```
Flat channel(s) 18,76 have been found and marked for rejection

Performing LOF Algorithm on non-flat 122 electrode(s)
Starting parallel pool (parpool) using the 'local' profile ...
connected to 4 workers.
Elapsed time is 89.005734 seconds.
LOF scores are computed successfully
Adapted threshold  for this dataset is 2.5
Channel(s) 2,15,20,21,30,44,71,92,114 have been marked for rejection

Skipping Periodogram Analysis...

Summary:
       Flat: [18 76]
    Outlier: [2 15 20 21 30 44 71 92 114]
     Muscle: []
```

Figure B.2: NEAR Bad Channel Plugin Execution - MATLAB Command Window

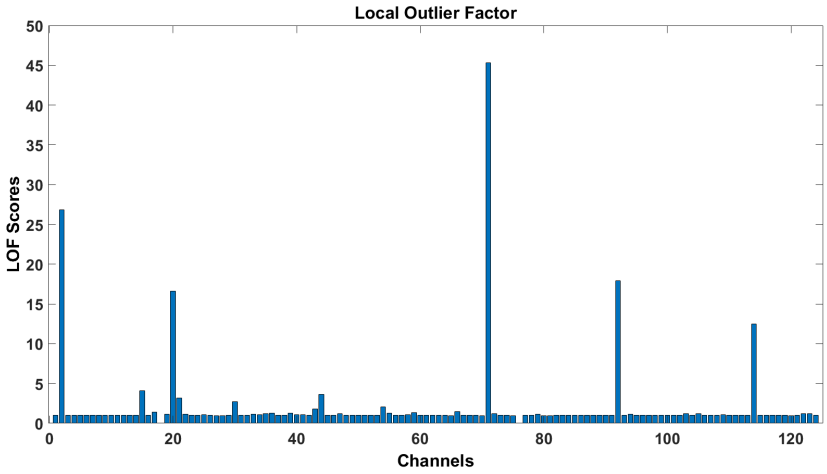


Figure B.3: Channel-wise LOF Scoring (from one of the sample newborn EEG)

3. An EEGLAB “scroll” window with the “Reject” button. The bad channels detected by flat-line and LOF are colored in red while the channels detected by the Periodogram analysis are marked in yellow.

To remove the detected bad channels, click on the “Reject” button. The following pop-up window will appear with the editable list of detected channels.

In this way, users can decide whether to remove the marked channels or modify the channel selection. Clicking on “Ok” creates a new dataset with the selected channels removed.

NOTE 1: Once NEAR bad channel detection is executed, the selected bad channels are saved in EEG.BadCh for future references.

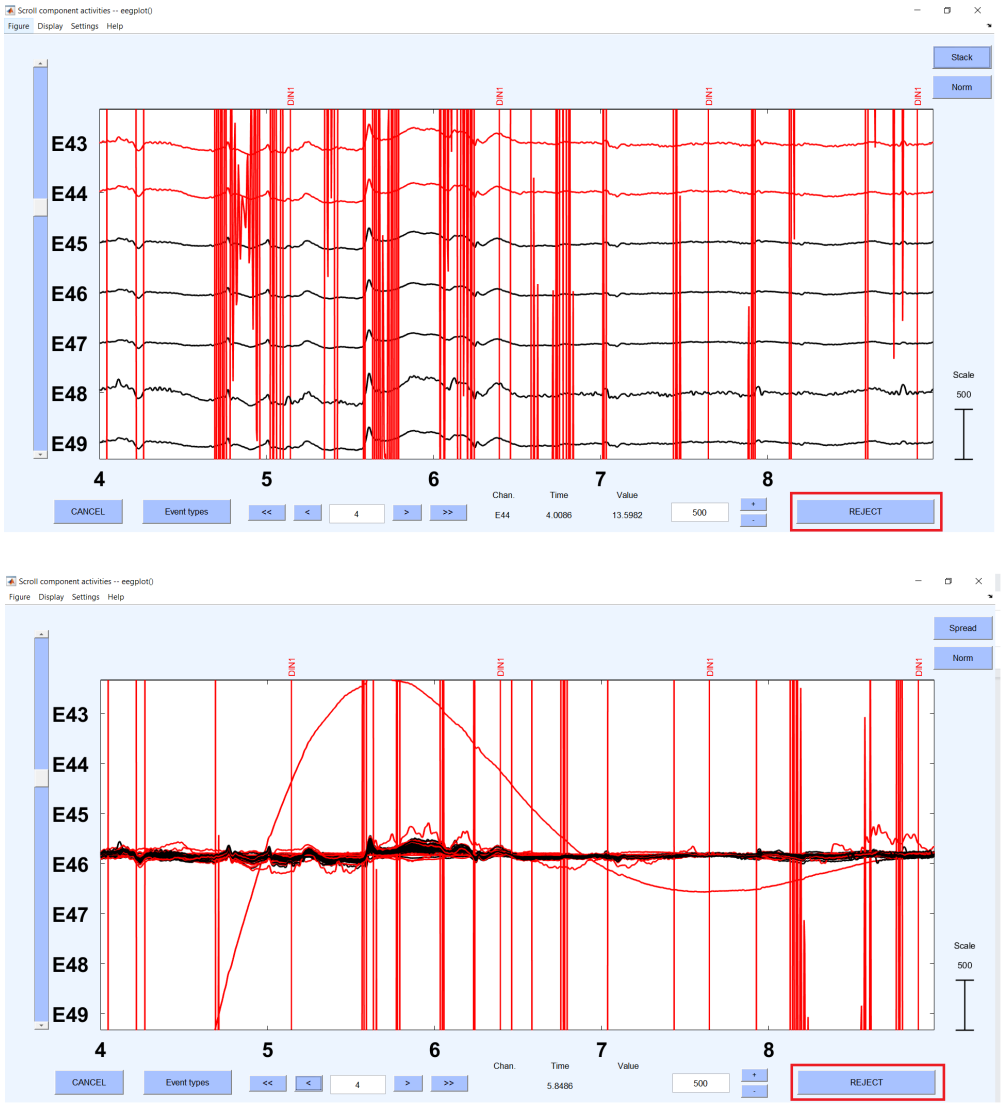


Figure B.4: EEGLAB "Scroll" Window with color-coded bad channels. Top panel: Spread view; Bottom panel: Stacked view

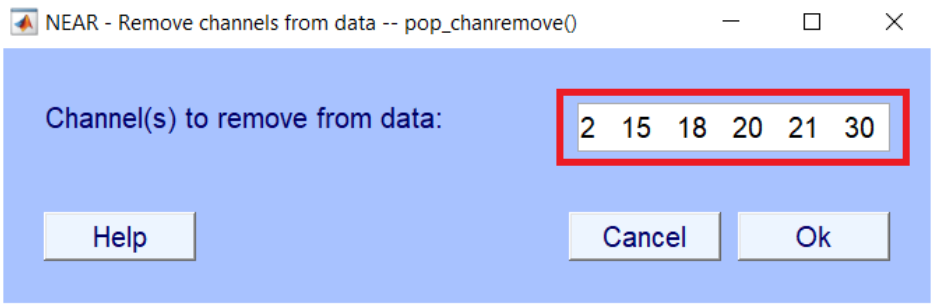


Figure B.5: List of bad channels detected by NEAR

NOTE 2: However, if NEAR is executed as an automated pipeline (e.g., in batch processing using command lines), the bad channels are removed without any user intervention. Instead, when executed using GUI, the tool always performs as a semi-automated toolbox. The equivalent command lines are as follows.

1) To run NEAR Bad Channel Algorithm:

```
{[EEG, flat_ch, lof_ch, periodo_ch, LOF_vec, thresh_lof_update] =
NEAR_getBadChannels(EEG, isFlat, flatWin, isLOF, thresh_lof,
dist_metric, isAdapt, isPeriodogram,frange, winsize,
winov, pthresh, isVisIns);}
```

For the details of these parameters, please refer to the user manual pdf document on GitHub.

2) To remove the detected bad channels:

```
badChans = sort(unique(union(union(flat_ch, lof_ch),periodo_ch))); ...
EEG = pop_select(EEG, 'nochannel', badChans);
```

B.9 Detecting Bad Segments using ASR

To run ASR, click on "Tools" Menu and select "Reject data using Clean Rawdata and ASR". The following pop-up window will appear. Since the data is already high-pass filtered and bad channels have already been removed, the user should deselect the associated options.

Enter the ASR parameter (k) obtained after calibration in the textbox provided under “Perform Artifact Subspace Reconstruction...”. For ASR correction mode, deselect “Remove bad data periods ...” option; for ASR removal mode, leave it selected as it is.

Alternatively, the following command can be used to run ASR:

```
EEG = pop_clean_rawdata(EEG, 'FlatlineCriterion',
'off','ChannelCriterion','off','LineNoiseCriterion','off,
'Highpass','off','BurstCriterion',rej_cutoff, ...
  'WindowCriterion',add_reject,
'BurstRejection',rej_mode,'Distance','Euclidian');
```

where *rej_cutoff* = ASR Parameter (24, or 13 in the screenshots above), *rej_mode* is either ‘on’ (for ASR Removal) and ‘off’ (ASR Correction), *add_reject* is either ‘on’ for additional rejection (not recommended) and ‘off’ for no additional rejection.

The original data and the ASR processed data for the two modes in their respective *k* values

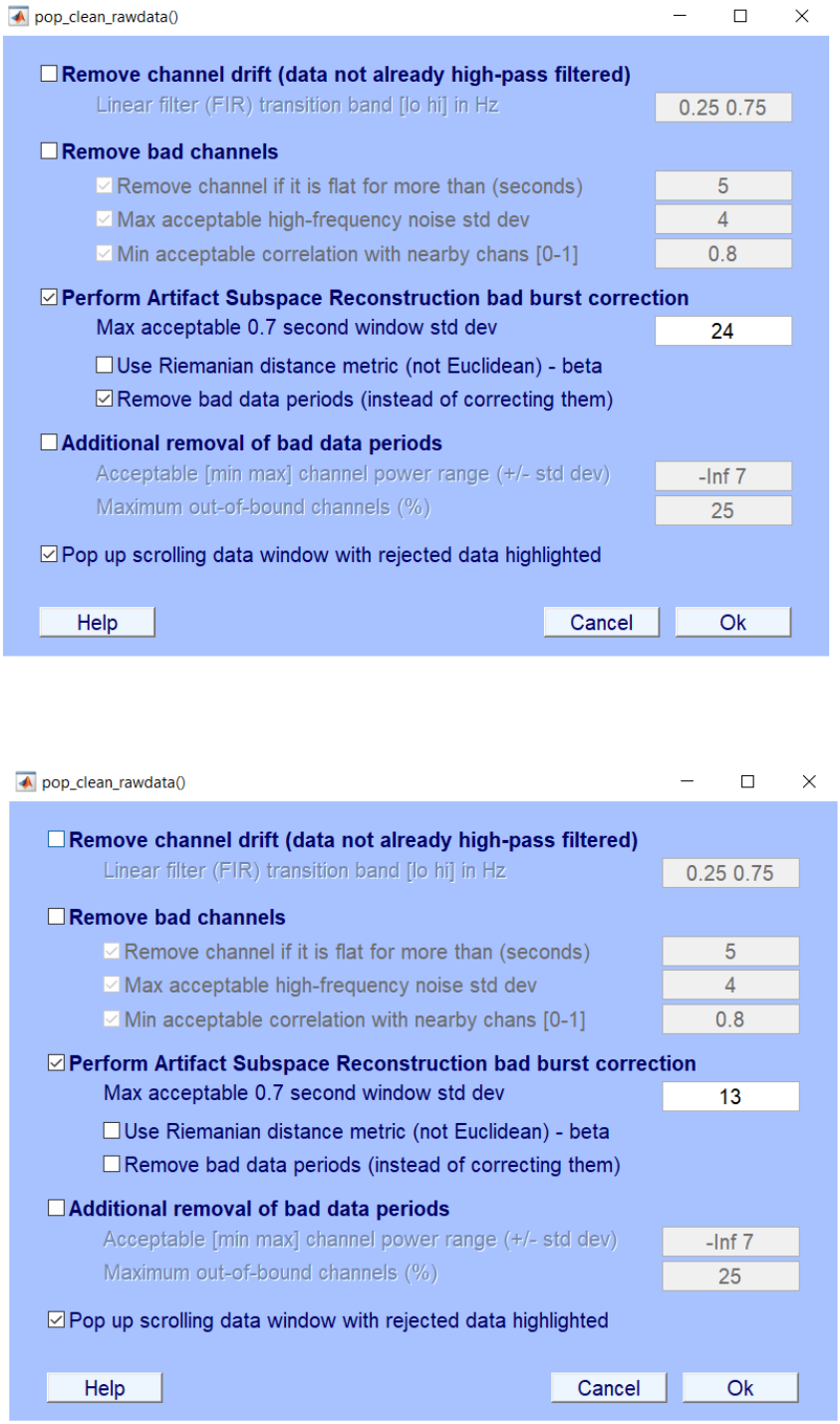


Figure B.6: Example of the use of ASR in NEAR. Top panel: ASR Removal Mode with k = 24. Bottom panel: ASR Correction Mode with k = 13

for a sample newborn data are shown below:

B.10 ERP Epochs Segmentation

This step can be necessary in two scenarios:

1. The imported data was already epoched and later had to be made continuous to enable ASR (as done in section B.5).
2. The imported data is continuous but has to be epoched for ERP analysis.

The following command can be executed to first epoch the data:

```
EEG = pop_epoch( EEG, erp_event_markers, erp_epoch_duration, 'epochinfo', ...
'yes');
```

where *erp_event_markers* is a cell array that contains the list of event markers. E.g., 'Eyes Close', 'Eyes Open', and *erp_epoch_duration* is the duration of each segment. E.g., [0 1000] ms. For baseline removal,

```
EEG = pop_rmbase( EEG, baseline_window ,[]);
```

where *baseline_window* is the baseline period. E.g., [0 200] ms.

B.11 Interpolating Missing Channels

To interpolate the removed bad channels, click "Interpolate Electrodes" on the "Tools" menu. Alternatively, the following command can be executed:

```
EEG = pop_interp(EEG, origEEG.chanlocs, interp_type);
```

where *origEEG* contains the copy of the imported raw EEG, and *interp_type* denotes the kind of interpolation (recommended: 'spherical').

B.12 Average Referencing

To average reference the data, click on the "Tools" menu and select "Re-reference the data" option. Alternatively, the command given below can be executed:

```
EEG = pop_reref( EEG, []);
```

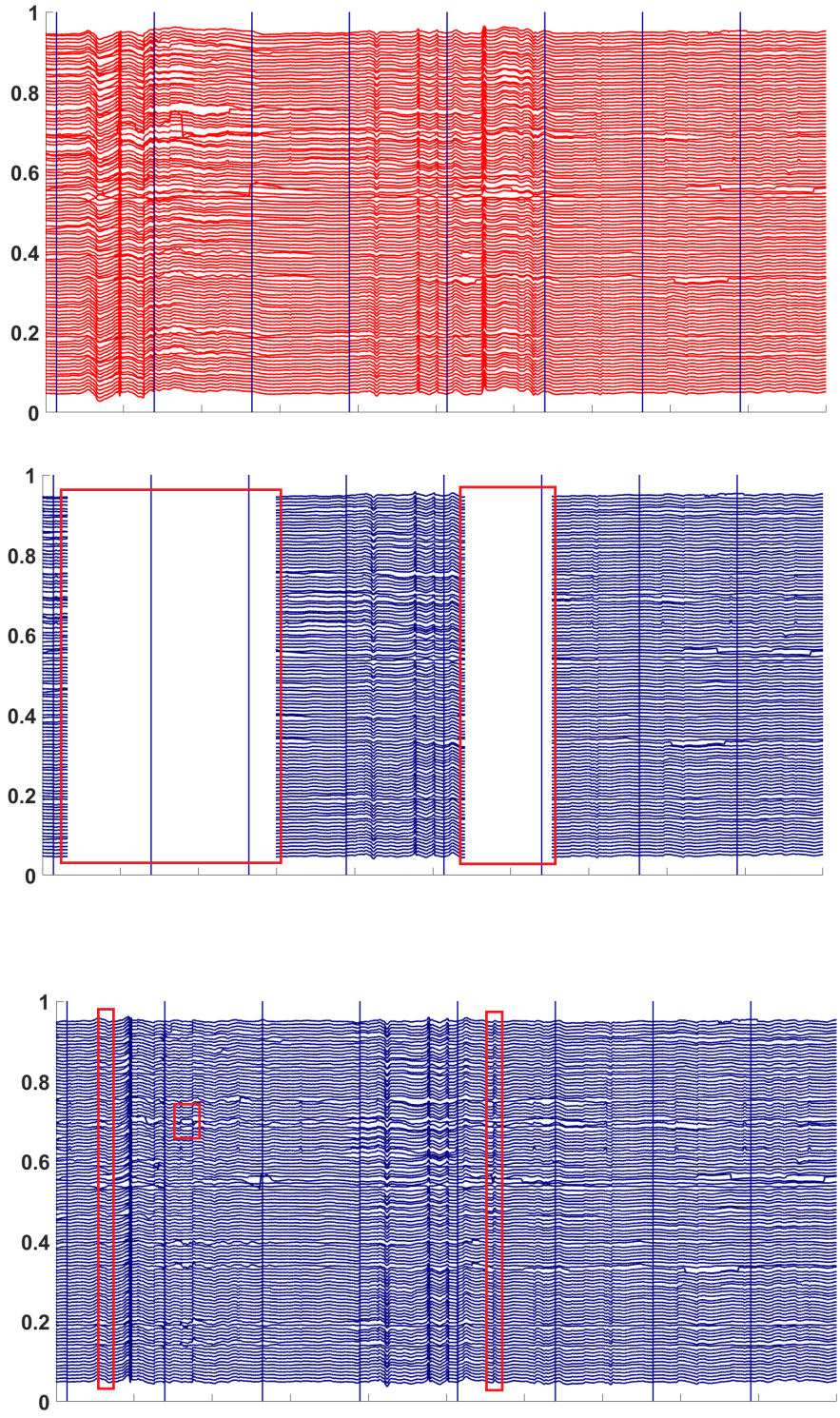


Figure B.7: ASR Processing. Top panel: Band-pass filtered data with bad channels removed by NEAR. Middle panel: Bad segments (highlighted in red) removed by ASR with k = 24. Bottom panel: Bad segments corrected by ASR with k = 13 (the most obvious ones, highlighted in red).

Instead, if you wish to perform re-referencing to a specific channel, it can be done so using the commands given in step 8 of the tutorial script.

B.13 Save and Report Functionality

To save the processed data, and to create a comprehensive summary on the analysis performed a report functionality is added to our pipeline. The processed data would be saved in .set format and can be found within the import data location in a separate folder: NEAR_Processed. Instead, the report file (in .csv format) can be found in a folder: NEAR_Reports. The relevant scripts can be found in step 9 of the tutorial script.

Bibliography

Acunzo, D. J., MacKenzie, G., & van Rossum, M. C. (2012). Systematic biases in early ERP and ERF components as a result of high-pass filtering. *Journal of Neuroscience Methods*, *209*(1), 212–218. https://doi.org/10.1016/j.jneumeth.2012.06.011

Adibpour, P., Hochmann, J.-R., & Papeo, L. (2021). Spatial Relations Trigger Visual Binding of People. *Journal of Cognitive Neuroscience*, *33*(7), 1343–1353. https://doi.org/10.1162/jocn_a_01724

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. https://doi.org/10.1176/appi.books.9780890425596

Arpaia, P., De Benedetto, E., Donato, N., Duraccio, L., & Moccaldi, N. (2021). A wearable ssvep bci for ar-based, real-time monitoring applications. *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 1–6. https://doi.org/10.1109/MeMeA52024.2021.9478593

Artoni, F., Delorme, A., & Makeig, S. (2018). Applying dimension reduction to EEG data by principal component analysis reduces the quality of its subsequent independent component decomposition. *NeuroImage*, *175*, 176–187. https://doi.org/10.1016/j.neuroimage.2018.03.016

Atkinson, J., & Braddick, O. (2012). Visual attention in the first years: Typical development and developmental disorders. *Developmental Medicine &amp Child Neurology*, *54*(7), 589–595. https://doi.org/10.1111/j.1469-8749.2012.04294.x

Bacher, L. F. (2010). Factors regulating eye blink rate in young infants. *Optometry and Vision Science*, *87*(5), 337–343. https://doi.org/10.1097/opx.0b013e3181d951b2

Barbaro, J., & Dissanayake, C. (2012). Early markers of autism spectrum disorders in infants and toddlers prospectively identified in the social attention and communication study. *Autism*, *17*(1), 64–86. https://doi.org/10.1177/1362361312442597

Beauchemin, M., Gonzalez-Frankenberger, B., Tremblay, J., Vannasing, P., Martinez-Montes, E., Belin, P., Beland, R., Francoeur, D., Carceller, A.-M., Wallois, F., & Lassonde, M. (2010). Mother and stranger: An electrophysiological study of voice processing in newborns. *Cerebral Cortex*, *21*(8), 1705–1711. https://doi.org/10.1093/cercor/bhq242

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, *7*(6), 1129–1159. https://doi.org/10.1162/neco.1995.7.6.1129

Belwafi, K., Gannouni, S., & Aboalsamh, H. (2021). Embedded brain computer interface: State-of-the-art in research. *Sensors*, *21*(13), 4293.

Blum, S., Mirkovic, B., & Debener, S. (2019). Evaluation of riemannian asr on ceegrid data: An artifact correction method for bcis. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 3625–3630. https://doi.org/10.1109/SMC.2019.8914319

Bouthillier, X., Tsirigotis, C., Corneau-Tremblay, F., Schweizer, T., Dong, L., Delaunay, P., Bronzi, M., Suhubdy, D., Askari, R., Noukhovitch, M., Xue, C., Ortiz-Gagné, S., Breuleux, O., Bergeron, A., Bilaniuk, O., Bocco, S., Bertrand, H., Alain, G., Serdyuk, D., … Beckham, C. (2021). *Epistimio/orion: Asynchronous Distributed Hyperparameter Optimization* (Version v0.2.1). Zenodo. https://doi.org/10.5281/zenodo.3478592

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). Lof: Identifying density-based local outliers. *SIGMOD Rec.*, *29*(2), 93–104. https://doi.org/10.1145/335191.335388

Bria, A., Marrocco, C., & Tortorella, F. (2021). Sinc-based convolutional neural networks for eeg-bci-based motor imagery classification. In A. Del Bimbo, R. Cucchiara, S. Sclaroff, G. M. Farinella, T. Mei, M. Bertini, H. J. Escalante, & R. Vezzani (Eds.), *Pattern recognition. icpr international workshops and challenges* (pp. 526–535). Springer International Publishing.

Buiatti, M., Pena, M., & DEHAENELAMBERTZ, G. (2009). Investigating the neural correlates of continuous speech computation with frequency-tagged neuroelectric responses. *NeuroImage*, *44*(2), 509–519. https://doi.org/10.1016/j.neuroimage.2008.09.015

Buiatti, M., Di Giorgio, E., Piazza, M., Polloni, C., Menna, G., Taddei, F., Baldo, E., & Vallortigara, G. (2019). Cortical route for facelike pattern processing in human newborns. *Proceedings of the National Academy of Sciences*, *116*(10), 4625–4630. https://doi.org/10.1073/pnas.1812419116

Chang, C.-Y., Hsu, S.-H., Pion-Tonachini, L., & Jung, T.-P. (2018). Evaluation of artifact subspace reconstruction for automatic eeg artifact removal, 1242–1245. https://doi.org/10.1109/EMBC.2018.8512547

Chaudhary, U., Birbaumer, N., & Ramos-Murguialday, A. (2016). Brain–computer interfaces for communication and rehabilitation. *Nature Reviews Neurology*, *12*(9), 513–525.

Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological Psychiatry*, *74*(3), 195–203. https://doi.org/10.1016/j.biopsych.2012.11.022

Chen, X., Peng, H., Yu, F., & Wang, K. (2017). Independent vector analysis applied to remove muscle artifacts in eeg data. *IEEE Transactions on Instrumentation and Measurement*, *66*(7), 1770–1779. https://doi.org/10.1109/TIM.2016.2608479

Chen, Y., Shi, W., Liu, Q., Chu, H., Chen, X., Yan, L., Wu, J., Li, L., & Gao, X. (2022). Eeg measurement for suppression in refractive amblyopia and push-pull perception efficacy. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *30*, 1321–1330. https://doi.org/10.1109/TNSRE.2022.3175177

Cho, K., van Merrienboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. https://doi.org/10.48550/ARXIV.1409.1259

Choi, H.-S., Park, S., Lee, J. H., Heo, H., Jeon, D., & Lee, K. (2021). Real-time denoising and dereverberation wtih tiny recurrent u-net. *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://doi.org/10.1109/icassp39728.2021.9414852

Cracco, E., Oomen, D., Papeo, L., & Wiersema, J. R. (2022). Using eeg movement tagging to isolate brain responses coupled to biological movements. *Neuropsychologia*, *177*, 108395.

De Clercq, W., Vergult, A., Vanrumste, B., Van Paesschen, W., & Van Huffel, S. (2006). Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram. *IEEE Transactions on Biomedical Engineering*, *53*(12), 2583–2587. https://doi.org/10.1109/TBME.2006.879459

Debnath, R., Buzzell, G. A., Morales, S., Bowers, M. E., Leach, S. C., & Fox, N. A. (2020). The maryland analysis of developmental EEG (MADE) pipeline. *Psychophysiology*, *57*(6). https://doi.org/10.1111/psyp.13580

de Heering, A., & Rossion, B. (2015). Rapid categorization of natural face images in the infant right hemisphere. *eLife*, *4*. https://doi.org/10.7554/elife.06564

Delorme, A. (2018). Github:clean$_r aw$.

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. https://doi.org/10.1016/j.jneumeth.2003.10.009

Delorme, A., Sejnowski, T., & Makeig, S. (2007). Enhanced detection of artifacts in eeg data using higher-order statistics and independent component analysis. *NeuroImage*, *34*(4), 1443–1449. https://doi.org/10.1016/j.neuroimage.2006.11.004

Di Giorgio, E., Frasnelli, E., Salva, O. R., Scattoni, M. L., Puopolo, M., Tosoni, D., Simion, F., & Vallortigara, G. (2016). Difference in visual social predispositions between newborns at low- and high-risk for autism. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep26395

DRYODE^{TM} - IDUN Technologies. (2016).

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175–191.

Feldman, D. H. (2012). Cognitive development in childhood. https://doi.org/10.1002/9781118133880.hop206008

Fifer, W. P., Byrd, D. L., Kaku, M., Eigsti, I.-M., Isler, J. R., Grose-Fifer, J., Tarullo, A. R., & Balsam, P. D. (2010). Newborn infants learn during sleep. *Proceedings of the National Academy of Sciences*, *107*(22), 10320–10323. https://doi.org/10.1073/pnas.1005061107

Fix, E., & Hodges, J. (1989). Discriminatory analysis, nonparametric discrimination: Consistency properties. technical report 4. *USAF School of Aviation Medicine, Randolph Field*. https://doi.org/10.2307/1403796

Fotiadou, E., Konopczyński, T., Hesser, J., & Vullings, R. (2020). End-to-end trained encoder–decoder convolutional neural network for fetal electrocardiogram signal denoising. *Physiological Measurement*, *41*(1), 015005. https://doi.org/10.1088/1361-6579/ab69b9

Gabard-Durnam, L., Mendez Leal, A., Wilkinson, C., & Levin, A. (2018). The harvard automated processing pipeline for electroencephalography (happe): Standardized processing software for developmental and high-artifact data [PMID: 29535597; PMCID: PMC5835235.]. *Front Neurosci*, *12*(97). https://doi.org/10.3389/fnins.2018.00097.

Georgieva, S., Lester, S., Noreika, V., Yilmaz, M. N., Wass, S., & Leong, V. (2020). Toward the understanding of topographical and spectral signatures of infant movement artifacts in naturalistic EEG. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.00352

Golub, G. H. (1969). Matrix decompositions and statistical calculations. *Statistical Computation*, 365–397.

Gramfort, A. (2013). MEG and EEG data analysis with MNE-python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267

Greco, A., Mammone, N., Morabito, F. C., & Versaci, M. (2006). Kurtosis, renyi's entropy and independent component scalp maps for the automatic artifact rejection from eeg data. *International Journal of Signal Processing*, *2*(4), 240–244.

Guermandi, M., Cardu, R., Scarselli, E. F., & Guerrieri, R. (2015). Active electrode ic for eeg and electrical impedance tomography with continuous monitoring of contact impedance. *IEEE transactions on biomedical circuits and systems*, *9*(1), 21–33.

Guo, Y., Bufacchi, R. J., Novembre, G., Kilintari, M., Moayedi, M., Hu, L., & Iannetti, G. D. (2020). Ultralow-frequency neural entrainment to pain. *PLoS biology*, *18*(4), e3000491.

Hakvoort, G., Reuderink, B., & Obbink, M. (2011). Comparison of psda and cca detection methods in a ssvep-based bci-system.

Heinrich, M. P., Stille, M., & Buzug, T. M. (2018). Residual u-net convolutional neural network architecture for low-dose CT denoising. *Current Directions in Biomedical Engineering*, *4*(1), 297–300. https://doi.org/10.1515/cdbme-2018-0072

Herrmann, C. S. (2001). Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental brain research*, *137*(3-4), 346–353.

Hervé, E., Mento, G., Desnous, B., & François, C. (2022). Challenges and new perspectives of developmental cognitive EEG studies. *NeuroImage*, *260*, 119508. https://doi.org/10.1016/j.neuroimage.2022.119508

Hoehl, S., Wiese, L., & Striano, T. (2008). Young infants' neural processing of objects is affected by eye gaze direction and emotional expression (G. F. Marcus, Ed.). *PLoS ONE*, *3*(6), e2389. https://doi.org/10.1371/journal.pone.0002389

Hollander, M., Wolfe, D. A., & Chicken, E. (2013). *Nonparametric statistical methods*. John Wiley & Sons.

Hutman, T., Rozga, A., DeLaurentis, A. D., Barnwell, J. M., Sugar, C. A., & Sigman, M. (2010). Response to distress in infants at risk for autism: A prospective longitudinal study. *Journal of Child Psychology and Psychiatry*, *51*(9), 1010–1020. https://doi.org/10.1111/j.1469-7610.2010.02270.x

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in r*. Springer. https://faculty.marshall.usc.edu/gareth-james/ISL/

Johnson, M. H. (2014). Autism: Demise of the innate social orienting hypothesis. *Current Biology*, *24*(1), R30–R31. https://doi.org/10.1016/j.cub.2013.11.021

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature*, *504*(7480), 427–431. https://doi.org/10.1038/nature12715

Kabdebon, C., Leroy, F., Simmonet, H., Perrot, M., Dubois, J., & Dehaene-Lambertz, G. (2014). Anatomical correlations of the international 10–20 sensor placement system in infants. *NeuroImage*, *99*, 342–356. https://doi.org/10.1016/j.neuroimage.2014.05.046

Kabdebon, C., Fló, A., de Heering, A., & Aslin, R. (2022). The power of rhythms: How steady-state evoked responses reveal early neurocognitive development. *NeuroImage*, *254*, 119150. https://doi.org/https://doi.org/10.1016/j.neuroimage.2022.119150

Kalenkovich, E., Shestakova, A., & Kazanina, N. (2022). Frequency tagging of syntactic structure or lexical properties a registered MEG study. *Cortex*, *146*, 24–38. https://doi.org/10.1016/j.cortex.2021.09.012

Kamphuisen, A., Bauer, M., & van Ee, R. (2008). No evidence for widespread synchronized networks in binocular rivalry: MEG frequency tagging entrains primarily early visual cortex. *Journal of Vision*, *8*(5), 4. https://doi.org/10.1167/8.5.4

Kartsch, V., Kumaravel, V., Benatti, S., Vallortigara, G., Benini, L., Farella, E., & Buiatti, M. (2022). Efficient low-frequency SSVEP detection with wearable EEG using normalized canonical correlation analysis. *Sensors*, *22*(24), 9803. https://doi.org/10.3390/s22249803

Kartsch, V., Tagliavini, G., Guermandi, M., Benatti, S., Rossi, D., & Benini, L. (2019). BioWolf: A sub-10-mW 8-channel advanced brain–computer interface platform with a nine-core processor and BLE connectivity. *IEEE Transactions on Biomedical Circuits and Systems*, *13*(5), 893–906. https://doi.org/10.1109/tbcas.2019.2927551

Klapwijk, E. T., van den Bos, W., Tamnes, C. K., Raschle, N. M., & Mills, K. L. (2021). Opportunities for increased reproducibility and replicability of developmental neuroimaging. *Developmental Cognitive Neuroscience*, *47*, 100902. https://doi.org/10.1016/j.dcn.2020.100902

Kothe, C. A., & Jung, T.-P. (Google Patent WO2015047462A9, Jun. 2014). Artifact removal techniques with signal reconstruction.

Kothe, C. A., & Makeig, S. (2013). BCILAB: A platform for brain–computer interface development. *Journal of Neural Engineering*, *10*(5), 056014. https://doi.org/10.1088/1741-2560/10/5/056014

Krol, L. R., Pawlitzki, J., Lotte, F., Gramann, K., & Zander, T. O. (2018). SEREEGA: Simulating event-related EEG activity. *Journal of Neuroscience Methods*, *309*, 13–24. https://doi.org/10.1016/j.jneumeth.2018.08.001

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (1st ed.). Springer.

Kumaravel, V. P. (2022). Github:near.

Kumaravel, V. P., Buiatti, M., & Farella, E. (2021). Hyperparameter selection for reliable eeg denoising using asr: A benchmarking study. *2021 IEEE International Conference on Bioin-*

formatics and Biomedicine (BIBM), 3638–3641. https://doi.org/10.1109/BIBM52615.2021.9669561

Kumaravel, V. P., Buiatti, M., Parise, E., & Farella, E. (2022). Adaptable and robust EEG bad channel detection using local outlier factor (LOF). *Sensors*, *22*(19), 7314. https://doi.org/10.3390/s22197314

Kumaravel, V. P., Farella, E., Parise, E., & Buiatti, M. (2022). NEAR: An artifact removal pipeline for human newborn EEG data. *Developmental Cognitive Neuroscience*, *54*, 101068. https://doi.org/10.1016/j.dcn.2022.101068

Kumaravel, V. P., Kartsch, V., Benatti, S., Vallortigara, G., Farella, E., & Buiatti, M. (2021). Efficient artifact removal from low-density wearable EEG using artifacts subspace reconstruction. https://doi.org/10.1109/embc46164.2021.9629771

Kumaravel, V. P., & Paissan, F. (2021). Github:cleaneegnet.

Kumaravel, V. P., Paissan, F., & Farella, E. (2021). Towards a domain-specific neural network approach for EEG bad channel detection. *2021 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 1–4. https://doi.org/10.1109/SPMB52430.2021.9672305

Landa, R. (2007). Early communication development and intervention for children with autism. *Mental Retardation and Developmental Disabilities Research Reviews*, *13*(1), 16–25. https://doi.org/10.1002/mrdd.20134

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., & Lance, B. J. (2018). Eegnet: A compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, *15*(5), 056013.

Leach, S. C., Morales, S., Bowers, M. E., Buzzell, G. A., Debnath, R., Beall, D., & Fox, N. A. (2020). Adjusting ADJUST: Optimizing the ADJUST algorithm for pediatric data using geodesic nets. *Psychophysiology*, *57*(8). https://doi.org/10.1111/psyp.13566

Lee, T.-W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation*, *11*(2), 417–441. https://doi.org/10.1162/089976699300016719

Li, L., Jamieson, K., Rostamizadeh, A., Gonina, E., Ben-Tzur, J., Hardt, M., Recht, B., & Talwalkar, A. (2020). A system for massively parallel hyperparameter tuning. *Proceedings of Machine Learning and Systems*, *2*, 230–246.

Lin, Z., Zhang, C., Wu, W., & Gao, X. (2006). Frequency recognition based on canonical correlation analysis for ssvep-based bcis. *IEEE transactions on biomedical engineering*, *53*(12), 2610–2614.

Lochy, A., Jacques, C., Maillard, L., Colnat-Coulbois, S., Rossion, B., & Jonas, J. (2018). Selective visual representation of letters and words in the left ventral occipito-temporal cortex with intracerebral recordings. *Proceedings of the National Academy of Sciences*, *115*(32), E7595–E7604.

Lopes da Silva, F. (2013). EEG and MEG: Relevance to neuroscience. *Neuron*, *80*(5), 1112–1128. https://doi.org/10.1016/j.neuron.2013.10.017

Lopez, K., Monachino, A., Morales, S., Leach, S., Bowers, M., & Gabard-Durnam, L. (2022). HAPPILEE: HAPPE in low electrode electroencephalography, a standardized pre-

processing software for lower density recordings. *NeuroImage*, *260*, 119390. https://doi.org/10.1016/j.neuroimage.2022.119390

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., Bishop, S., Esler, A., Hus, V., Oti, R., Richler, J., Risi, S., & Lord, C. (2009). The autism diagnostic observation schedule—toddler module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(9), 1305–1320. https://doi.org/10.1007/s10803-009-0746-z

Macari, S. L., Campbell, D., Gengoux, G. W., Saulnier, C. A., Klin, A. J., & Chawarska, K. (2012). Predicting developmental status from 12 to 24 months in infants at risk for autism spectrum disorder: A preliminary report. *Journal of Autism and Developmental Disorders*, *42*(12), 2636–2647. https://doi.org/10.1007/s10803-012-1521-0

Mahajan, R., & Morshed, B. I. (2015). Unsupervised eye blink artifact denoising of eeg data with modified multiscale sample entropy, kurtosis, and wavelet-ica. *IEEE Journal of Biomedical and Health Informatics*, *19*(1), 158–165. https://doi.org/10.1109/JBHI.2014.2333010

Makeig, S., Bell, A., Jung, T.-P., & Sejnowski, T. J. (1995). Independent component analysis of electroencephalographic data. In D. Touretzky, M. Mozer, & M. Hasselmo (Eds.), *Advances in neural information processing systems*. MIT Press. https://proceedings.neurips.cc/paper/1995/file/754dda4b1ba34c6fa89716b85d68532b-Paper.pdf

MATLAB. (2018). *Version 9.9.0.944444 (r2018b)*. The MathWorks Inc.

MATLAB. (2020). *Version 9.9 (r2020b)*. The MathWorks Inc.

Mayor-Torres, J. M., Ravanelli, M., Medina-DeVilliers, S. E., Lerner, M. D., & Riccardi, G. (2021). Interpretable sincnet-based deep learning for emotion recognition from eeg brain activity. *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 412–415. https://doi.org/10.1109/EMBC46164.2021.9630427

Messinger, D., Young, G. S., Ozonoff, S., Dobkins, K., Carter, A., Zwaigenbaum, L., Landa, R. J., Charman, T., Stone, W. L., Constantino, J. N., Hutman, T., Carver, L. J., Bryson, S., Iverson, J. M., Strauss, M. S., Rogers, S. J., & Sigman, M. (2013). Beyond autism: A baby siblings research consortium study of high-risk children at three years of age. *Journal of the American Academy of Child &amp Adolescent Psychiatry*, *52*(3), 300–308.e1. https://doi.org/10.1016/j.jaac.2012.12.011

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229–240. https://doi.org/10.1111/j.1469-8986.2010.01061.x

Moliner, E., & Valimaki, V. (2022). A two-stage u-net for high-fidelity denoising of historical recordings. *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. https://doi.org/10.1109/icassp43922.2022.9746977

Monroy, C., Domınguez-Martınez, E., Taylor, B., Marin, O. P., Parise, E., & Reid, V. M. (2021). Understanding the causes and consequences of variability in infant ERP editing practices. *Developmental Psychobiology*, *63*(8). https://doi.org/10.1002/dev.22217

Montagna, F., Buiatti, M., Benatti, S., Rossi, D., Farella, E., & Benini, L. (2017). A machine learning approach for automated wide-range frequency tagging analysis in embedded neuromonitoring systems. *Methods*, *129*, 96–107.

Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Jung, T.-P., & Cauwenberghs, G. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Transactions on Biomedical Engineering*, *62*(11), 2553–2567. https://doi.org/10.1109/tbme.2015.2481482

Nadig, A. S., Ozonoff, S., Young, G. S., Rozga, A., Sigman, M., & Rogers, S. J. (2007). A prospective study of response to name in infants at risk for autism. *Archives of Pediatrics &amp Adolescent Medicine*, *161*(4), 378. https://doi.org/10.1001/archpedi.161.4.378

Nolan, H., Whelan, R., & FASTER, R. R. (2010). Fully automated statistical thresholding for eeg artifact rejection [PMID: 20654646.]. *J Neurosci Methods*, *30;192(1):152-62*. https://doi.org/10.1016/j.jneumeth.2010.07.015.

Norcia, A. M., Appelbaum, L. G., Ales, J. M., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, *15*(6), 4. https://doi.org/10.1167/15.6.4

Nordin, A. D., Hairston, W. D., & Ferris, D. P. (2018). Dual-electrode motion artifact cancellation for mobile electroencephalography. *Journal of Neural Engineering*, *15*(5), 056024. https://doi.org/10.1088/1741-2552/aad7d7

Odabaee, M., Freeman, W. J., Colditz, P. B., Ramon, C., & Vanhatalo, S. (2013). Spatial patterning of the neonatal EEG suggests a need for a high number of electrodes. *NeuroImage*, *68*, 229–235. https://doi.org/10.1016/j.neuroimage.2012.11.062

Onton, J., Westerfield, M., Townsend, J., & Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience &amp Biobehavioral Reviews*, *30*(6), 808–822. https://doi.org/10.1016/j.neubiorev.2006.06.007

Ozonoff, S., Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., Rogers, S. J., Rozga, A., Sangha, S., Sigman, M., Steinfeld, M. B., & Young, G. S. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child &amp Adolescent Psychiatry*, *49*(3), 256–266.e2. https://doi.org/10.1016/j.jaac.2009.11.009

Paissan, F., Kumaravel, V., & Farella, E. (2022). Interpretable cnn for single-channel artifacts detection in raw eeg signals. *2022 IEEE Sensors Applications Symposium (SAS)*, 1–6. https://doi.org/10.1109/SAS54819.2022.9881381

Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, *23*(7), 728–733. https://doi.org/10.1177/0956797612438734

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Percio, C. D., Noce, G., Lopez, S., Tucci, F., Carlin, G., Lizio, R., Musat, A. M., Soricelli, A., Salvatore, M., Ferri, R., Nobili, F., Arnaldi, D., Famà, F., Buttinelli, C., Giubilei, F., Ma-

rizzoni, M., Güntekin, B., Yener, G., Stocchi, F., ... Babiloni, C. (2022). What a single electroencephalographic (EEG) channel can tell us about patients with dementia due to alzheimer's disease. *International Journal of Psychophysiology*, *182*, 169–181. https://doi.org/10.1016/j.ijpsycho.2022.10.011

Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. (2003). Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos. *International Journal of Audiology*, *42*(4), 177–219. https://doi.org/10.3109/14992020309101316

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

Pomiechowska, B., & Csibra, G. (2022). Nonverbal action interpretation guides novel word disambiguation in 12-month-olds. *Open Mind*, *6*, 51–76. https://doi.org/10.1162/opmi_a_00055

Reynolds, G. D., & Richards, J. E. (2005). Familiarization, attention, and recognition memory in infancy: An event-related potential and cortical source localization study. *Developmental Psychology*, *41*(4), 598–615. https://doi.org/10.1037/0012-1649.41.4.598

Ronga, I., Galigani, M., Bruno, V., Castellani, N., Sebastiano, A. R., Valentini, E., Fossataro, C., Neppi-Modona, M., & Garbarini, F. (2021). Seeming confines: Electrophysiological evidence of peripersonal space remapping following tool-use in humans. *Cortex*, *144*, 133–150. https://doi.org/10.1016/j.cortex.2021.08.004

Rossion, B., Torfs, K., Jacques, C., & Liu-Shuang, J. (2015). Fast periodic presentation of natural images reveals a robust face-selective electrophysiological response in the human brain. *Journal of Vision*, *15*(1), 18–18. https://doi.org/10.1167/15.1.18

Salvaro, M., Benatti, S., Kartsch, V. J., Guermandi, M., & Benini, L. (2019). A minimally invasive low-power platform for real-time brain computer interaction based on canonical correlation analysis. *IEEE Internet of Things Journal*, *6*(1), 967–977. https://doi.org/10.1109/JIOT.2018.2866341

Schiavone, P., Conti, F., Rossi, D., Gautschi, M., Pullini, A., Flamand, E., & Benini, L. (2017). Slow and steady wins the race? a comparison of ultra-low-power risc-v cores for internet-of-things applications, 1–8. https://doi.org/10.1109/PATMOS.2017.8106976

Schneider, C., Pereira, M., Tonin, L., & del R. Millan, J. (2019). "real-time eeg feedback on alpha power lateralization leads to behavioral improvements in a covert attention task". https://doi.org/10.18112/openneuro.ds002034.v1.0.1

Schneider, C., Pereira, M., Tonin, L., & del R. Millán, J. (2019). Real-time EEG feedback on alpha power lateralization leads to behavioral improvements in a covert attention task. *Brain Topography*, *33*(1), 48–59. https://doi.org/10.1007/s10548-019-00725-9

Sheppard, E., Pillai, D., Wong, G. T.-L., Ropar, D., & Mitchell, P. (2015). How easy is it to read the minds of people with autism spectrum disorder? *Journal of Autism and Developmental Disorders*, *46*(4), 1247–1254. https://doi.org/10.1007/s10803-015-2662-8

Talbott, E. O., Arena, V. C., Rager, J. R., Clougherty, J. E., Michanowicz, D. R., Sharma, R. K., & Stacy, S. L. (2015). Fine particulate matter and the risk of autism spectrum disorder. *Environmental Research*, *140*, 414–420. https://doi.org/10.1016/j.envres.2015.04.021

Tran, Y., Boord, P., Middleton, J., & Craig, A. (2004). Levels of brain wave activity (8–13 hz) in persons with spinal cord injury. *Spinal Cord*, *42*(2), 73–79. https://doi.org/10.1038/sj.sc.3101543

Tripathi, M. (2021). Facial image denoising using AutoEncoder and UNET. *Heritage and Sustainable Development*, *3*(2), 89–96. https://doi.org/10.37868/hsd.v3i2.71

Wan, M. W., Green, J., Elsabbagh, M., Johnson, M., Charman, T., & Plummer, F. (2012). Parent–infant interaction in infant siblings at risk of autism. *Research in Developmental Disabilities*, *33*(3), 924–932. https://doi.org/10.1016/j.ridd.2011.12.011

Wang, G., Teng, C., Li, K., Zhang, Z., & Yan, X. (2016). The removal of EOG artifacts from EEG signals using independent component analysis and multivariate empirical mode decomposition. *IEEE Journal of Biomedical and Health Informatics*, *20*(5), 1301–1308. https://doi.org/10.1109/jbhi.2015.2450196

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

Winkler, I., Brandl, S., Horn, F., Waldburger, E., Allefeld, C., & Tangermann, M. (2014). Robust artifactual independent component classification for BCI practitioners. *Journal of Neural Engineering*, *11*(3), 035013. https://doi.org/10.1088/1741-2560/11/3/035013

Yuen, K.-V., & Mu, H.-Q. (2012). A novel probabilistic method for robust parametric identification and outlier detection. *Probabilistic Engineering Mechanics*, *30*, 48–59. https://doi.org/https://doi.org/10.1016/j.probengmech.2012.06.002

Yuen, K.-V., & Ortiz, G. A. (2017). Outlier detection and robust regression for correlated data. *Computer Methods in Applied Mechanics and Engineering*, *313*, 632–646. https://doi.org/https://doi.org/10.1016/j.cma.2016.10.004

Zhang, H., Zhao, M., Wei, C., Mantini, D., Li, Z., & Liu, Q. (2021). Eegdenoisenet: A benchmark dataset for deep learning solutions of eeg denoising. *Journal of Neural Engineering*, *18*(5), 056057.

Zhang, Y., Xu, P., Liu, T., Hu, J., Zhang, R., & Yao, D. (2012). Multiple frequencies sequential coding for ssvep-based brain-computer interface. *PloS one*, *7*(3), e29519.

Zheng, X., Xu, G., Zhi, Y., Wang, Y., Han, C., Wang, B., Zhang, S., Zhang, K., & Liang, R. (2019). Objective and quantitative assessment of interocular suppression in strabismic amblyopia based on steady-state motion visual evoked potentials. *Vision Research*, *164*, 44–52. https://doi.org/https://doi.org/10.1016/j.visres.2019.07.003

Zhu, D., Bieger, J., Garcia Molina, G., & Aarts, R. M. (2010). A survey of stimulation methods used in ssvep-based bcis. *Computational intelligence and neuroscience*, *2010*.

Zhu, Q., Feng, J., & Huang, J. (2016). Natural neighbor: A self-adaptive neighborhood method without parameter k. *Pattern Recognition Letters*, *80*, 30–36. https://doi.org/10.1016/j.patrec.2016.05.007

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2004). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience*, *23*(2-3), 143–152. https://doi.org/10.1016/j.ijdevneu.2004.05.001

europass

Velu Prabhakar Kumaravel

Date of birth: 24/06/1991 | **Nationality:** Indian | (+39) 3923267459 |

velu.kumaravel@unitn.it | veluprabhakar@ymail.com |

Via Sommarive, 18, 38123, Trento, Italy

WORK EXPERIENCE

01/02/2022 – 31/07/2022 – Lausanne, Switzerland

VISITING PHD STUDENT – SWISS FEDERAL INSTITUTE OF TECHNOLOGY LAUSANNE (EPFL)

During my 6-months Ph.D. internship, I was involved in PEDESITE (Personalized Detection of Epileptic Seizure in the Internet of Things (IoT) Era), funded by the Fonds National Suisse de la Recherche Scientifique. I worked in Machine Learning approaches to enhance the detection of epileptic seizures.

EDUCATION AND TRAINING

01/11/2018 – CURRENT – Trento, Italy

PHD IN COGNITIVE AND BRAIN SCIENCES – Fondazione Bruno Kessler / Center for Mind-Brain Sciences (CIMeC), University of Trento

The goal of my Ph.D. is to bring mobile EEG solutions for investigating neural correlates of cognitive functions in human newborns/infants.

My tasks and responsibilities include

1. Develop an artifacts pre-processing pipeline for newborn/infant high-density EEG (offline processing based on EEGLAB).
2. Developing MATLAB-based software for pre-processing and neural feature extraction using our custom 8-electrode wearable EEG setup.
3. Developing firmware (embedded C) for the online artifacts processing specific to our custom EEG system.

Address Trento, Italy

Bologna, Italy

MASTER OF SCIENCE IN TELECOMMUNICATIONS ENGINEERING – University of Bologna

Address Bologna, Italy |

Thesis Experimental Evaluation of BITalino: A Low-cost Modular Platform for Biosignals Acquisition

Chennai, India

BACHELORS OF ENGINEERING IN ELECTRONICS AND COMMUNICATIONS ENGINEERING – Madras Institute of Technology, Anna University

Address Chennai, India |

Thesis Biomedical Image Segmentation using Combined Watershed and Level Set Method

DIGITAL SKILLS

My Digital Skills

EEG Signal Processing | Biomedical Signal Processing | Python | MATLAB/EEGLAB | Linux | Statistical Analysis | Machine Learning | LaTeX

FUNDED RESEARCH PROJECTS

01/11/2019 – 31/10/2021

Researcher in NeuroSoNew: Portable EEG-based screening of social predispositions in newborns

ERC-2018-PoC - ERC Proof of Concept Grant of EUR 149,945 (ID: 842243)

Earlier studies revealed that newborns with Autism-Spectrum Disorders (ASD) demonstrate different preferences to face processing compared to healthy populations. This project aims at designing a portable EEG for a rapid & reliable investigation of neural correlates of face processing in newborns to assist in an early diagnosis of ASD. My role in this project is to develop advanced signal-processing tools for pre-processing and extracting neural features from newborn data. I also collaborate with researchers from the University of Bologna to port the solutions into the resource-constrained MCU platform.
https://cordis.europa.eu/project/id/842243

HONOURS AND AWARDS

30/09/2022

Think Open @ CIMeC Awards 2022 – Center for Mind/Brain Sciences, University of Trento, Italy

This is awarded to my open-source toolbox (NEAR) that complies with all the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles and at the same time carrying high scientific relevance in the field of developmental cognitive science.
https://twitter.com/cimec_unitrento/status/1575797402061049857

27/01/2022

Third Selected PhD Talk – Center for Mind/Brain Sciences, University of Trento, Italy

My PhD project is chosen as the third best for the CIMeC Doctoral Day 2022.
https://twitter.com/cimec_unitrento/status/1486692452442558472

28/01/2021

CIMeC Doctoral Day - Best Poster Award – Center for Mind/Brain Sciences, University of Trento, Italy

Best poster award for presenting my research work as a part of CIMeC Doctoral Day.
https://github.com/vpKumaravel/dsday2021_poster

2015

UNIBO Azione 1 Grant Award – University of Bologna, Italy

Full tuition fee waiver granted by the University of Bologna to the International Students.

Star Award Winner for Two Consecutive Times (2014) – Larsen & Toubro Infotech

Quarterly-award dedicated to best performers in the Business Unit.

PUBLICATIONS

NEAR: An artifact removal pipeline for human newborn EEG data

SI: Developmental EEG Methods - A Tutorial Approach (Developmental Cognitive Neuroscience)
https://www.sciencedirect.com/science/article/pii/S1878929322000123 – 2022

Electroencephalography (EEG) is arising as a valuable method to investigate neurocognitive functions shortly after birth. However, obtaining high-quality EEG data from human newborn recordings is challenging because compared to adults and older infants, datasets are typically much shorter due to newborns' limited attentional span and much noisier due to non-stereotyped artifacts mainly caused by uncontrollable movements. Here we propose Newborn EEG Artifact Removal (NEAR), a novel pipeline for EEG artifact removal specifically designed for human newborns. NEAR is based on two key steps: 1) A novel bad channel detection tool relying on the Local Outlier Factor (LOF), a robust outlier detection algorithm; 2) A parameter calibration procedure for adapting Artifacts Subspace Reconstruction (ASR), a method primarily developed for artifact removal in mobile adult EEG, to newborn EEG data. NEAR is trained and validated on two different newborn EEG datasets recorded with a frequency-tagging paradigm, an experimental design that tackles the limitation of newborn's attentional span by maximizing the signal-to-noise ratio. Validation shows that NEAR outperforms existing methods for both bad channel detection and overall artifact removal, matching expert performance in obtaining statistically significant EEG responses from noisy datasets. NEAR consists of a set of freely available EEGLABbased custom scripts (https://github.com/vpKumaravel/NEAR).
https://github.com/vpKumaravel/NEAR

Adaptable and Robust EEG Bad Channel Detection Using Local Outlier Factor (LOF)

Sensors, MDPI Journal
https://www.mdpi.com/1424-8220/22/19/7314/htm – 2022

Electroencephalogram (EEG) data are typically affected by artifacts. The detection and removal of bad channels (i.e., with poor signal-to-noise ratio) is a crucial initial step. EEG data acquired from different populations require different cleaning strategies due to the inherent differences in the data quality, the artifacts' nature, and the employed experimental paradigm. To deal with such differences, we propose a robust EEG bad channel detection method based on the Local Outlier Factor (LOF) algorithm. Unlike most existing bad channel detection algorithms that look for the global distribution of channels, LOF identifies bad channels relative to the local cluster of channels, which makes it adaptable to any kind of EEG. To test the performance and versatility of the proposed algorithm, we validated it on EEG acquired from three populations (newborns, infants, and adults) and using two experimental paradigms (event-related and frequency-tagging). We found that LOF can be applied to all kinds of EEG data after calibrating its main hyperparameter: the LOF threshold. We benchmarked the performance of our approach with the existing state-of-the-art (SoA) bad channel detection methods. We found that LOF outperforms all of them by improving the F1 Score, our chosen performance metric, by about 40% for newborns and infants and 87.5% for adults.

Efficient Artifact Removal from Low-Density Wearable EEG using Artifacts Subspace Reconstruction

International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)
https://ieeexplore.ieee.org/document/9629771 – 2021

With the advent of hardware-software co-design, cost-effective portable and wireless EEG systems are made possible for both remote clinical monitoring and BCI (Brain-Computer Interfaces) applications. However, the quality of the collected data from such systems can be largely hampered by the presence of movement artifacts often leading to misclassification of neural features. A recently developed Artifacts Subspace Reconstruction (ASR) is a promising solution to remove non-stationary and non-stereotypical artifacts common in mobile EEG settings with high sensitivity and specificity. Since ASR has only been validated with high-density EEG systems (channels > 32), it was unclear whether it was equally efficient on low-density (channels = 8, or 16) mobile EEG systems. Cheng et al. empirically provided an empirical lower limit on the number of channels as 20 for an efficient artifact cleaning but a systematic validation was still missing. The technical merit of this paper is its first-time validation of ASR on clean and contaminated Steady-State Visually Evoked Potential (SSVEP) data acquired with an ultra-low-power custom EEG system 'BioWolf' with 8 dry electrodes. *Significance*: Results indicate that ASR can be used in low-density EEG

acquired with just eight electrodes to remove the non-stereotypical artifacts. Moreover, our chosen SSVEP paradigm is one of the most widely used in BCI applications. Altogether, the results promise a robust portable BCI system especially for recordings in natural context using ASR as a pre-processing technique in the near future.

Hyperparameter selection for reliable EEG denoising using ASR: a benchmarking study

IEEE International Conference on Bioinformatics and Biomedicine (BIBM)
https://ieeexplore.ieee.org/document/9669561 – 2021

Artifacts preprocessing in EEG is remarkably significant to extract reliable neural responses in the downstream analysis. A recently emerging powerful preprocessing tool among the EEG community is Artifacts Subspace Reconstruction (ASR). ASR is an unsupervised machine learning algorithm to identify and correct the transient-like non-stationary noisy samples. ASR is fully automatic, therefore, suitable for online applications. However, the performance of ASR is strongly dependent on the user-defined hyperparameter k. A poor choice of k might lead to severe performance degradation.In this work, we benchmark the performance of ASR against its parameter k. Toward this goal, we used the Temple University Hospital EEG Artifact Corpus (TUAR), which consists of 310 EEG files recorded in clinical settings from epileptic patients. Remarkably, these files are annotated for artifacts by trained personnel with a high inter-rater agreement score ($\kappa > 0.8$). Considering these reliable labels as ground truth, ASR has shown the best performance in artifacts cleaning with k ranging between 20 and 40.

Towards a Domain-specific Neural Network Approach for EEG Bad Channel Detection

IEEE Signal Processing in Medicine and Biology (SPMB)
https://ieeexplore.ieee.org/document/9672305 – 2021

Electroencephalogram (EEG) is prone to several artifacts that often lead to the misclassification of neural features in Brain-Computer Interfaces (BCI). Traditionally, detecting and removing bad EEG electrodes (or channels) is often the first and most critical step in cleaning the data. There are a few automated tools, and each uses its statistical signal processing techniques with tunable hyperparameters (e.g., the z-score threshold for amplitude-based outlier detection). To our knowledge, an objective deep-learning approach for this problem is still missing. This paper proposes *cleanEEGNet*, a Convolutional Neural Network, to identify the bad channels in EEG signals. We carefully chose the model hyperparameters (i.e., kernel size and stride) to mimic the conventional detection of bad channels performed via visual inspection. An open-source dataset from OpenNeuro with annotated bad channels is used to train and validate the network. For a benchmark comparison, we chose four state-of-the-art automated conventional methods for bad channel removal, including FASTER and HAPPE. Among them, HAPPE performed the best, achieving a balanced accuracy of 66%, while *cleanEEGNet* outperformed HAPPE by 17% with a balanced accuracy of 78%.

Interpretable CNN for Single-Channel Artifacts Detection in Raw EEG Signals

https://ieeexplore.ieee.org/document/9881381 – 2022

Electroencephalogram (EEG) signals recorded from the scalp are often affected by artifacts. Most existing artifact detection methods rely on multi-channel statistics such as inter-channel correlation. Recently, there has been a growing interest in realizing single-channel EEG systems to promote everyday use, demanding novel artifacts detection techniques. This paper presents validation results for single-channel artifacts detection in raw EEG signals using four neural architectures: a one-dimensional CNN (1D-CNN) - proposed by us, EEGNet, SincNet and EEGDenoiseNet. We used semi-synthetic EEG data corrupted with Ocular (EOG) and Myo-graphic (EMG) noise components to validate the approaches. Precisely, we contaminated the randomly chosen EEG channels with EOG and EMG artifacts in a controlled manner using a predefined Signal-to-Noise Ratio (SNR) such that the ground truth is known. We validated these models both in terms of classification performance and the interpretability of the learned features. Of the four models, 1D-CNN, EEGNet, and SincNet achieved a comparable classification accuracy (around 99%) and EEGDenoiseNet achieved as low as 64%. Analysing the learned filters for interpretability, we found both 1D-CNN and EEGNet clearly separates EOG (Delta and Theta) and EMG (Gamma) frequency bands from EEG. Instead, SincNet prioritized to learn EEG-specific features (Alpha and Beta) rather than artifact-related information still achieiving the comparable performance as the other two models. EEGDenoiseNet with kernel width of 3 was excluded from this evaluation as it is practically infeasible to perform FFT analysis. Finally, we also computed the number of training parameters for each model to evaluate which among them would be

suitable for a resource-constrained wearable device and we found that 1D-CNN and SincNet are the most parameter-efficient, although not by a large margin.

Chapter in the book "Health Monitoring Systems"

https://www.taylorfrancis.com/chapters/edit/10.1201/9780429113390-6/biomedical-signal-analysis-1-simone-benatti-victor-kartsch-fabio-montagna-elisabetta-farella-velu-kumaravel – 2019
My master thesis is included in the chapter "Biomedical Signal Analysis 1" of this book.

REVIEWER FOR INTERNATIONAL JOURNALS/CONFERENCES

I have been a direct or delegated reviewer for the following Journals and Conferences

- Journal of Neural Engineering
- Journal of Psychophysiology

- IEEE EMBC Conference
- DATE Conference
- IEEE SAS Symposium

LANGUAGE SKILLS

Mother tongue(s): **TAMIL**

Other language(s):

	UNDERSTANDING		SPEAKING		WRITING
	Listening	Reading	Spoken production	Spoken interaction	
ENGLISH	C1	C1	C1	C1	C1
ITALIAN	B1	A2	B1	B1	A2
FRENCH	A1	A1	A1	A1	A1

Levels: A1 and A2: Basic user; B1 and B2: Independent user; C1 and C2: Proficient user