Eye movement artifact correction in infant EEG

Diksha Srishyla

Integrated Program in Neuroscience, McGill University

August 2023

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Masters of Science

Table of Contents

Abstract (English)

Brain activity measured using an electroencephalogram (EEG) in infants could support the study of typical cognitive development and biomarkers for neurodevelopmental disorders. Independent Component Analysis (ICA) is a well-established approach to clean EEG and remove the impact of signals that are not of neural origin, such as muscle and eye movements. However, ICA may not perform as effectively for artifact removal in infants as in adults. Our study aims to compare ICA and Artifact blocking (AB), which has been shown to improve eye movement artifact correction in infant data. We analyzed EEG collected from 50 infants between 6 and 18 months of age as part of the International Infant EEG Data Integration Platform (EEG-IP), a longitudinal multi-site study. EEG recordings were made while infants sat on their caregivers' laps and watched videos. We corrected for eye movement artifacts in the EEG recordings using ICA and AB. We calculated the proportion of effectively corrected segments, signal-to-noise ratio (SNR), power-spectral density (PSD), and multiscale entropy (MSE) in segments with eye movement artifact and in clean segments. In the eye movement artifact segments, we found that AB corrected the artifact effectively for a significantly lower proportion of segments compared to ICA. The eye movement segments after correction by AB had a higher SNR than after correction by ICA. Since the noise term, in our definition of the SNR, captures what gets removed by the algorithm, the lower SNR of ICA indicates better artifact rejection for this algorithm. The eye movement segments after correction by both AB and ICA had a significantly lower PSD compared to the original signal, however, the difference in PSD between ICA and the original signal was much greater, indicating more effective correction. The eye movement segments after correction by both AB and ICA showed an overall higher MSE, indicating equally effective correction. In the clean segments, SNR following correction by AB was greater than that after correction by ICA, suggesting less distortion by AB. PSD following correction by both AB and ICA was lower than that of the original signal, suggesting distortion by both algorithms. MSE following correction by AB was similar to that of the original signal in most channels. Results for MSE following ICA with respect to that of the original signal, were mixed. Further investigation using SNR calculated on simulated EEG, PSD changes in higher frequency bands caused by ICA and MSE calculated on the residuals of EEG signal, is required. The results will guide the choice of artifact rejection methods in future studies that involve infant EEG.

Abstract (French)

L'activité cérébrale mesurée par l' électroencéphalogramme (EEG) en bas âge pourrait soutenir l'étude du développement cognitif typique et des biomarqueurs des troubles neurodéveloppementaux. L'analyse en composantes indépendantes (*independent component analysis*, ou *ICA*, en anglais) est une approche bien établie pour le prétraitement de l'EEG, particulièrement pour supprimer l'impact des signaux qui ne sont pas d'origine cérébrale, tels que ceux provenant des mouvements musculaires et oculaires. Cependant, l'ICA peut ne pas fonctionner aussi efficacement pour l'élimination des artefacts chez les nourrissons que chez les adultes. Notre étude vise à comparer la correction d'artéfact oculaires à l'aide de l'ICA et d'une seconde méthode nommée le blocage d'artefacts (AB). Cette dernière a démontrée des performances encourageantes pour les données provenant de nourrissons. Nous avons analysé l'EEG recueilli auprès de 50 nourrissons âgés de 6 à 18 mois dans le cadre de l'*International Infant EEG Data Integration Platform* (*EEG-IP*), une étude longitudinale multisite. L'enregistrements de l'EEG a été réalisé pendant que les nourrissons étaient assis sur les genoux de leurs parents et regardaient des vidéos. Nous avons corrigé les artefacts oculaires dans les enregistrements EEG en utilisant ICA et AB. Nous avons calculé la proportion des segments corrigés correctement, le rapport signal sur bruit (SNR), la puissance spectrale (PSD) et l'entropie multi-échelle (MSE) pour des segments contenant ou non des artéfacts oculaires. Pour les segments contenant des artefacts oculaires, nous avons constaté que AB corrigeait corectement une proportion significativement plus faible de segments par rapport à ICA. La correction des segments contenant des artéfacts oculaires résultait en un SNR plus élevé pour AB que pour ICA. Puisque nous avons défini le bruit dans notre SNR comme étant la portion de signal retiré par les algorithmes, un SNR plus faible pour des segments avec artéfacts indique un filtrage plus efficace. ICA semble donc mieux performer selon ce critère. Les segments avec mouvement oculaire avaient un PSD significativement inférieur par rapport au signal d'origine après correction par AB et ICA. Cependant, cet effect était plus grand pour l'ICA, indiquant une correction plus efficace. MSE était globalement plus élevée après correction par AB et ICA des segments avec mouvement oculaire, indiquant une correction similiaire pour ces deux algorithmes. Pour les segments sans artéfacts, le SNR après correction par AB était supérieur à celui après correction par ICA, suggérant moins de distorsion par AB. Le PSD après correction par AB et ICA était réduite dans des proportions similaires, suggérant une distorsion équivalente pour les deux algorithmes. La MSE après correction par AB était similaire à celle du signal d'origine dans la plupart des canaux. La comparaison des résultats pour la MSE avant et après l'ICA étaient cependant mitigée. Une enquête plus approfondie sur les propriétés des artefacts en fonction de l'échelle, à l'aide de MSE, est nécessaire. Les résultats suggèrent que les deux algorithmes provoquent une distorsion des signaux cérébraux. Une enquête plus approfondie sur les algorithmes de suppression des artefacts de mouvement oculaire chez les nourrissons est nécessaire. Les résultats guideront le choix des méthodes d'élimination des artefacts pour les futures études EEG en bas âge.

Acknowledgements

I would like to thank my supervisors Drs. Mayada Elsabbagh and Christian O'Reilly for their mentorship and feedback during the course of this thesis. I would also like to thank my committee, Drs. Julie Scorah and Tim Smith for their feedback on my thesis proposal and my presentations at the advisory committee meeting and seminar.

I would like to acknowledge my colleagues James Desjardins and Scott Huberty for their feedback on my preliminary results and encouragement.

I thank my colleagues at Dr. Elsabbagh's lab for creating an amiable work environment during these past two years and providing feedback on my preliminary results.

I would like to thank Professor Joseph Rochford for helping me plan the statistical analyses as my project for his course 'Statistics for the Neurosciences'.

I would like to acknowledge the research teams at the University of London and the University of Washington, that collected the data reported in this thesis.

The following sources of funding that made this project possible, are acknowledged: Diksha Srishyla is supported by the Maysie Macsporran fellowship, Faculty of medicine grant. EEG-IP is supported by funding from Brain Canada and the Azrieli Centre for Autism Research

List of figures and tables

Introduction

EEG recordings contain high amounts of signals not related to cortical processing, such as muscular and ocular artifacts. This problem is particularly acute with recordings from infants who are not yet responsive to instructions aimed at limiting such artifacts, e.g., directives to sit still and avoid blinking. As a result, recordings from infants are often not long enough for well-established approaches such as Independent Component Analysis (ICA) to work effectively [1]. For example, to apply ICA on EEG collected from 128-channels with a 250 Hz sampling rate, a minimum recording length of 3.2 minutes has been suggested [2]. Further, some studies suggest that ICA may not be the best solution for the removal of ocular artifacts in infant data [1]. Pre-processing pipelines have made significant progress in addressing these issues, to make infant EEG compatible with ICA [2]–[7]. However, existing solutions still require improvement and efforts to effectively account for different types of artifacts are continuing. In this study, we compare ICA with Artifact Blocking (AB) [8], an alternative approach to remove high amplitude eye movement artifacts from infant EEG. A previous study suggested that AB distorts clean EEG less than ICA, while effectively correcting for high amplitude artifacts. However, this demonstration was qualitative and used only one EEG recording. Our study aims to improve on the state-of-the-art by quantitatively benchmarking the performance of AB and ICA.

We will analyze resting state EEG recordings from 50 typically developing infants. We will annotate saccadic eye movement artifacts by visual inspection, apply ICA and AB, and compare the effectiveness of these algorithms based on 1) visual inspection, 2) signal-to-noise-ratio (SNR), 3) power spectral density (PSD), and 4) multiscale entropy (MSE) [8]. We aim to provide empirical evidence to help researchers decide on which artifact correction method to use for future studies.

Background

Electroencephalogram (EEG) studies in infants provide important information regarding typical brain development. This information can be used to improve our understanding of the development of cognitive processes in the context of neurodevelopmental disorders, such as

Autism Spectrum Disorder (ASD) [9]. EEG studies have found that brain activity patterns change differently during the first three years of life in infants who have an older sibling with ASD, compared to those who do not [12, 44]. However, findings are not consistent. EEG recordings contain a mix of neural signals and non-neural activity, or artifacts due to non-physiological sources (such as power line noise) and physiological sources (such as muscle contraction and eye movements). To avoid biasing the analysis, the recordings must be pre-processed to remove these artifacts prior to analysis. One potential reason for inconsistencies in research findings is that different research groups employ different techniques to pre-process the EEG.

In many studies, an experimenter visually inspects the recording to identify artifacts and reject segments of the EEG recording containing artifacts [10]–[14]. When segments of a recording containing eye movement artifacts are rejected, information regarding the neural activity of interest is also lost. For this reason, approaches aiming to correct rather than reject recordings with artifacts are often recommended [15].

Independent Component Analysis (ICA) is a well-established approach for artifact identification and correction, incorporated into several pre-processing pipelines for adult EEG [2]–[7]. ICA works by using statistical properties to separate a signal into independent components, which may then be classified by experts or automated algorithms into different types of brain activity or artifacts. Examples of these components are shown in Fig.1

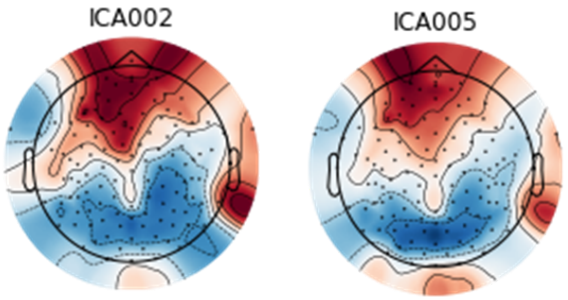


Fig 1: Examples of independent components corresponding to eye movements, from one of the recordings included in this study

In adult EEG, artifact components are relatively easily distinguished from brain activity. However, independent components in infant EEG often contain a mixture of brain signal and artifacts [16][17] and artifacts from eye movements is often spread across multiple components [15]. Similar to approaches relying on visual inspection OR although an improvement to approaches relying on visual inspection, ICA can make specific artifact components hard to identify and remove and, hence, leads to the removal of brain signal along with artifact [1]. Eye movement artifacts specifically, are challenging to identify and correct with ICA. Identification of independent components associated with artifacts relies on pre-identified, characteristic spatial distributions of EEG power over the scalp. Spectral power due to eye movements tends to be concentrated in the frontal region of the scalp in adults [18]. In infants, there is more variation in the topography of components and therefore an increased risk for misidentification of eye movements as brain activity, or vice versa [1].

In response to these recognized challenges, a number of pre-processing pipelines for infant EEG have been developed often using complementary methods that address different challenges and they each have strengths and limitations. Among these:

The Maryland Analysis of Developmental EEG (MADE) pipeline applies an automated processing step, following ICA, to remove residual eye movement artifacts [4]. The first step is to epoch the continuous data. EOG channels across epochs are then rejected when any electrooculogram (EOG)

channel exceed a certain voltage threshold. Epochs are then examined and rejected if more than 10% of EEG channels exceed a fixed voltage threshold. For the remaining epochs, EEG channels exceeding this threshold are interpolated.

The EEG Integrated Platform Lossless (EEG-IP-L) pipeline utilizes a procedure for removing artifacts that is compatible with short recording durations, such as that seen in infant EEG. It also applies a voltage-based criterion to identify time periods in the data that would not be corrected effectively by ICA, but also tests for bad channels based on extreme correlations between nearest neighbors. Further, the pipeline ensures a robust re-referencing to a consistent average across different montages. It then applies multiple ICA decompositions and semi-automatic component classification for artifact removal [5]. The classification is semi-automatic in that although the initial identification of artifacts is performed automatically, the user can review and edit automated artifact annotations for quality control.

The Newborn EEG Artifact Removal (NEAR) pipeline proposes another approach called Artifact Subspace Reconstruction (ASR). It detects bad channels and adapts ASR, originally designed for mobile adult EEG, to infant EEG. This approach was shown to be more effective than ICA to remove artifacts such as eye movements that do not have a fixed pattern of scalp topography and temporal dynamics [6]. However, the NEAR pipeline addresses specifically non-stereotyped artifacts and, hence, must be followed by another approach (e.g., ICA) for stereotyped artifacts such as eye movements. This may prove problematic in high-density recordings.

The Multiple Artifact Rejection Algorithm (iMARA) pipeline is trained specifically with components from infant EEG, to be able to automatically classify them into neural and artifactual categories after ICA decomposition [3]. However, the iMARA pipeline only works effectively for frequencies between 1-20 Hz.

The ADJUST pipeline also automatically identifies independent components but instead of using a training dataset, it simultaneously calculates multiple features associated with the topographical and spectral properties of eye movement artifacts.

The HAPPE (Harvard Automated Processing Pipeline for Electroencephalography) pipeline uses wavelet-thresholding ICA, which is said to mitigate distortions to spectral power estimates

caused by ICA [19]. However, in high density recordings (for example, 128 channels), not all channels may be included in analyses if the recording is not long enough for reliable decomposition for ICA. This length in seconds is specified by the formula [30*(no.of channels)^2/sampling frequency]. In a recording with a sampling rate of 500 Hz, this equates to 983.04 s, which is not feasible when recording EEG from infants.

Complementary approaches have been used to augment capacity of such pipelines to specifically detect eye movement artifcacts. One approach to identify components separated by ICA is to compare their properties during periods of saccades and fixations, using eye-tracking data [20]. However, this approach relies on the availability of accurate eye-tracking data. Identification may be biased by the presence of microsaccadic eye movements during fixations. Additionally, the method is effective only when there is a considerable number of high-amplitude eye movements in the data.

Another approach is to apply regression-based approaches to correct for eye movement artifacts by using recording from EOG channels, with electrodes placed above and below each eye[21]. This approach enables calculation of the correlation coefficient between signal recorded on the EOG and EEG electrodes, respectively, and correct for eye movements accordingly. A limitation is that regression-based methods rely on the usage of electrodes placed close to the eyes, which may not be feasible, depending on available resources in a lab or the level of cooperation of participants (e.g., children with ASD tend not to tolerate well the typical EOG electrode placement). They also may not be effective to correct for artifacts in recordings with short lengths, or if there are issues with estimating signals from EOG placed on either side of the eyes.

A systematic review of EEG and MEG studies of functional connectivity [57] found very few consistencies across a large number of studies. These differences were attributed to the following four factors: sample heterogeneity, acquisition, signal processing, hypothesis testing.

The pipelines discussed above constitute a significant aspect of signal processing and the standardization of EEG pre-processing pipelines, which will in turn enable consistencies across studies investigating biomarkers for ASD and other neurodevelopmental disorders, such as the ones above related to functional connectivity.

Hence, the choice of an eye movement artifact correction methods should consider the pros and cons of the available methods and the experimental design. Here, we propose another approach that has not been widely used in infant EEG, but has high potential to address known challenges. Artifact Blocking (AB) is an algorithm that offers a potential solution to the distortion of brain signals by ICA in infant EEG. It relies on an amplitude-based threshold to remove artifacts. The algorithm is described in detail in section 2.5. AB has been systematically evaluated and applied in the context of event-related potential (ERP) and high amplitude artifacts in infant data [15] [22] [23] but not with continuous EEG or eye movement artifacts specifically. When applied to infant EEG, AB has been reported to remove high amplitude artifacts, while causing significantly lower distortion to non-artifactual segments of data than ICA [8], [15].

Rationale for the study

A previous study presented a qualitative comparison of AB and ICA, using data from one infant [8]. The comparison demonstrated that AB removed artifact without distorting clean signal. However, this conclusion is not robust as it pertains to the data of only one infant. This study aims to generate quantitative metrics comparing the performance of AB and ICA in removing saccadic eye movement artifacts using a dataset of 50 infant EEG recordings. Further, we also propose a systematic benchmark of these two approaches relying on multiple metrics, for reference by researchers attempting to determine the best artifact correction method to use.

Hypotheses and Specific Aims

We expect that AB and ICA will achieve a similar level of eye movement artifact correction but that AB will not distort the clean signal as much as ICA [15]. Correction for artifacts and distortion of clean signal will be evaluated based on four outcomes. Three of these measures are commonly used in prior studies comparing algorithms for ocular artifact correction [24]–[27]. The fourth one, multiscale entropy, is an emerging metric. To our knowledge, this study is the first one where multiscale entropy has been used to evaluate the outcome of artifact correction. Multiscale entropy provides information in the non-linear domain, not provided by the other metrics. The four metrics are explained below:

i) **Visual inspection of segments containing saccadic eye movements:** A prior inspection of AB-corrected and ICA-corrected signal showed that AB was equally effective at removing eye movement artifacts compared to ICA [15]. We hypothesize that the proportion of eye movement segments that are effectively corrected by ICA will be the same as for AB.

ii) **Signal to noise ratio (SNR) in segments containing saccade eye movements and in clean segments:** In this ratio, the signal is defined as the EEG after correcting for artifacts and the noise refers to the portion removed from the original EEG while correcting for artifacts. Based on the expectations that AB and ICA will achieve a similar level of artifact correction (1) and AB will distort the clean signal less than ICA (2), we hypothesize that:

   a. In segments with eye movement artifacts, $SNR_{AB} \approx SNR_{ICA}$
   b. In clean segments, $SNR_{AB} > SNR_{ICA}$

iii) **Power spectral density (PSD) after applying AB and ICA:** An effective ocular artifact removal that does not distort neural signals should be associated with a decrease in spectral power in the eye movement segments, but no change in the clean segments [24], [25], [27]. Our corresponding hypotheses are formulated in Table 1.

Table 1. Hypotheses for the effect of AB and ICA on spectral power

	Eye movement segments	Clean segments
ICA	$PSD_{ICA} < PSD_{ORIGINAL}$	$PSD_{ICA} \neq PSD_{ORIGINAL}$
AB	$PSD_{AB} \neq PSD_{ORIGINAL}$	$PSD_{AB} \approx PSD_{ORIGINAL}$

iv) **Multiscale Entropy (MSE):** Multiscale entropy (MSE) is a measure of complexity, or information contained in the signal [28]. Certain artifact removal methods use MSE to identify artifact-related components [29]. Since low-information artifacts can overshadow high-information neural signals, segments containing artifacts have a lower entropy compared to non-artifact segments [26]. Based on the expectations that AB and ICA will achieve the same level of artifact correction (1) and AB will not distort the clean signal, but ICA will (2), we hypothesize that:

Table 2. Hypotheses for the effect of AB and ICA on entropy

	Eye movement segments	Clean segments
ICA	$Entropy_{ICA} > Entropy_{ORIGINAL}$	$Entropy_{ICA} < Entropy_{ORIGINAL}$
AB	$Entropy_{AB} > Entropy_{ORIGINAL}$	$Entropy_{AB} \approx Entropy_{ORIGINAL}$

Materials and methods

2.1 Participants

Recordings from the International Infant EEG Data Integration Platform (EEG-IP) were used for our analysis [5], [30]. EEG-IP includes EEG recordings from longitudinal infant-sibling studies conducted separately at two sites, London and Seattle [10], [11], [31]–[33]. EEG was recorded at ages 6-7 months and 12 months at both sites, and at 18 months in Seattle. Each site contributed raw EEG and behavioral data to this repository. Participants in these studies were grouped based on familial risk for ASD, that is, whether they had an older sibling diagnosed with ASD (elevated likelihood for ASD) or not (control group). The repository comprises a total of 420 recordings from 191 (94 females) infants. Table A1 presents a summary of recordings by participant age, risk group, and site.

For this study, we used recordings from the control group to benchmark these algorithms on a sample representative of the overall infant population. Taking into consideration the time involved in manual annotations of artifacts and revisions of analyses (approximately 0.75h/recording), it was not feasible in this thesis to include all recordings from the control group. Fifty recordings were randomly chosen; fifty was considered an adequate sample size based on a review of prior studies that evaluated ocular artifacts. Table A2 presents a summary of recordings by participant age, risk group, and site.

2.2 Recordings and pre-processing

EEG was recorded while infants sat on their caregivers' laps and watched videos on a monitor. An EGI 128 channel Hydrocel net was used. Videos consisted of either brightly colored toys producing sounds or an adult woman singing nursery rhymes.

Data were pre-processed and standardized to be maximally compatible for cross-site analysis. To establish an extendable standardized data state, open source solutions to technical constraints that typically impede successful integration were employed, including the adoption of the Brain Imaging Data Structure (BIDS) extension to EEG [34], [35] and standardized pre-processing using the EEG-IP Lossless Pipeline [5].

The EEG-IP Lossless pipeline includes systematic pre-processing procedures for identifying unreliable EEG signals and building comprehensive data annotation regarding signal quality. It

harmonizes data recordings by implementing data quality assessment procedures that are sensitive to differences in EEG data acquisition across sites and systems. The pipeline first addresses differences across datasets by executing *staging* scripts that are specific to each project, including procedures for the co-registration of electrode coordinates to a common shared head surface, a robust average reference, and a 1Hz high pass and notch filter (49-51hz in the London dataset and 59-61hz in Seattle dataset). The staging scripts then flag noisy time periods and channels based on consistently outlying variance values to make files more comparable to one another for later stages of the pipeline. Following the staging scripts, the pipeline performs signal quality assessments that use confidence intervals of signal properties within each file to flag unusual time periods and channels. Each time that channels are flagged as problematic, the data are re-referenced to interpolated channels on the shared co-registered head surface. Following the scalp channel assessment, a robust Adaptive Mixture Independent Component Analysis (AMICA) procedure is performed and components are automatically classified as being of neural origin or from artifacts. The automatic classification is based on a crowd-sourced database of labelled components. The database, ICLabels, includes 200,000 independent components from more than 6000 EEG recordings. It can be accessed at the website: https://iclabel.ucsd.edu. The automated classification is followed by a manual quality control step.

After their preprocessing with the EEG-IP Lossless pipeline, EEG recordings that retained enough signal to be included in post-processing were assessed for comparability [30]. The proportion of time removed from data due to artifact was similar across sites, as was the distribution of data removed due to different properties (extreme voltage variance, low correlation with neighboring channels, artifact identified by ICA decomposition). The average channel retention (which ranged from 77% to 82%), and the spatial variance in both the retained and rejected independent components was also similar across datasets. Finally, a power spectrum profile of the EEG recordings showed that the EEG-IP Lossless Pipeline resulted in similar profiles across datasets.

2.3 Annotations of eye movement artifact segments and clean segments

2.3.1 Eye movement artifact segments

An interactive plot generated using MNE-Python (version 1.3.1) [36], [37] was used to manually annotate (i.e., by visual inspection) the EEG recordings for eye movement and clean segments (Fig. 2). Segments with eye movement artifacts are indicated by orange windows in Fig. 2. They were annotated using reference images from prior studies that also examined eye movement artifacts in the EEG [38]. We also consider specifically channels that were most susceptible to eye movement artifacts in the EGI 128-channel Hydrocel Geodesic net (E1, E8, E14, E21, E32, E125, E126, E127, and E128; Fig. 3).

Video recordings or direct measurement of eye movements were not available. To verify accuracy of the annotated eye movement windows, we plotted scalp topographic plots of amplitude. The amplitude for true ocular artifacts is expected to be higher in the frontal regions compared to other regions of the scalp [39]. Lateral saccades are also expected to show as a left-right polarity inversion. Final selection was based on both temporal and topological profiles. An example is shown in Fig. 4.

2.3.2 Clean segments

Clean segments are indicated in Fig.2 by blue windows. We used an adult EEG atlas and a guide for labelling independent components corresponding to a variety of artifact types and brain activity [40]. Clean segments were evaluated by also considering the activity on channels Fz (E11), Cz, Pz (E62), and Oz (E75). Segments without any discernible artifacts, were marked as clean. We followed the recommendation from a previous study evaluating artifacts in pediatric

EEG [39] and included a minimum of 30s of clean segments per recording.

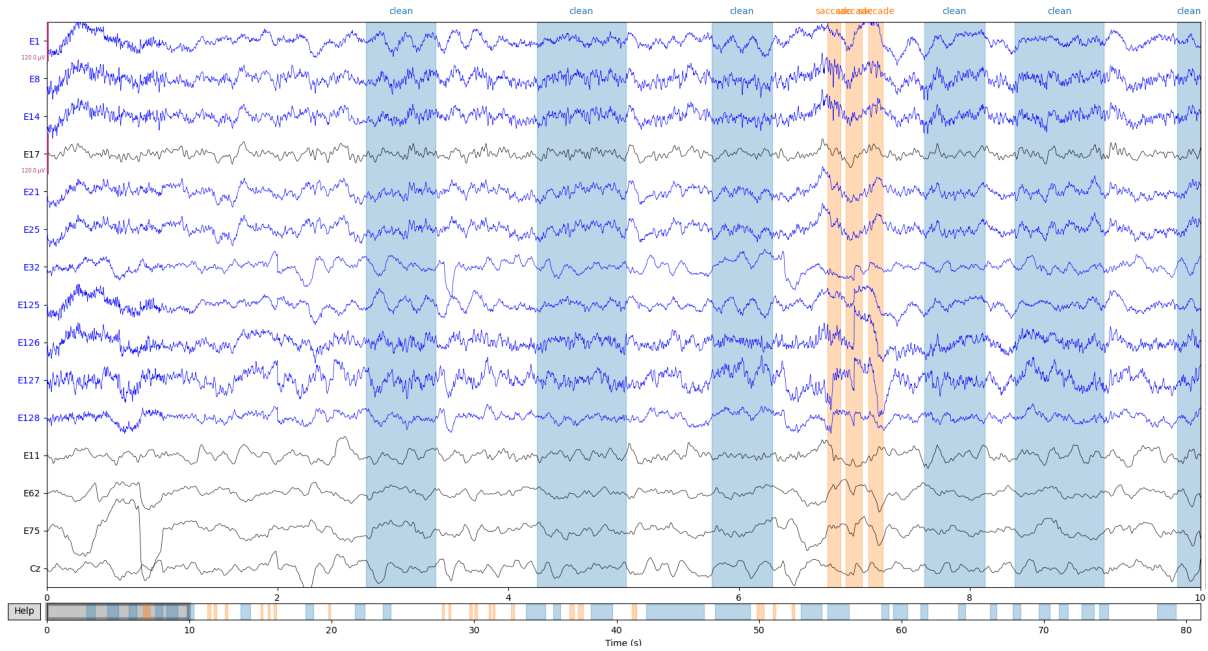


Fig. 2: Window displaying annotations of eye movement segments (light orange sections) and clean segments (blue sections). Channels used to annotate eye movement are indicated by the blue lines, while clean segment reference channels are indicated by the black lines**.**

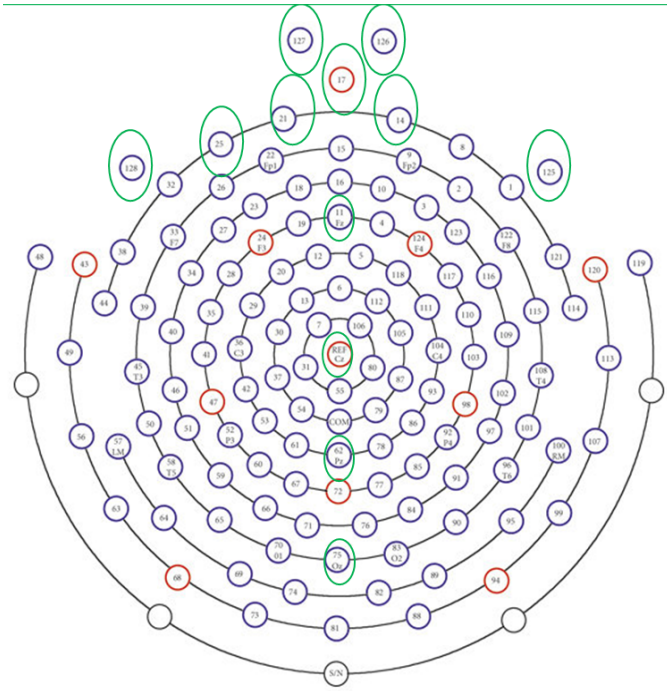


Fig.3: EGI 128-channel Hydrocel Geodesic montage. Channels used for annotations of eye movements and clean segments circled in green. Figure modified with green circles from [15].

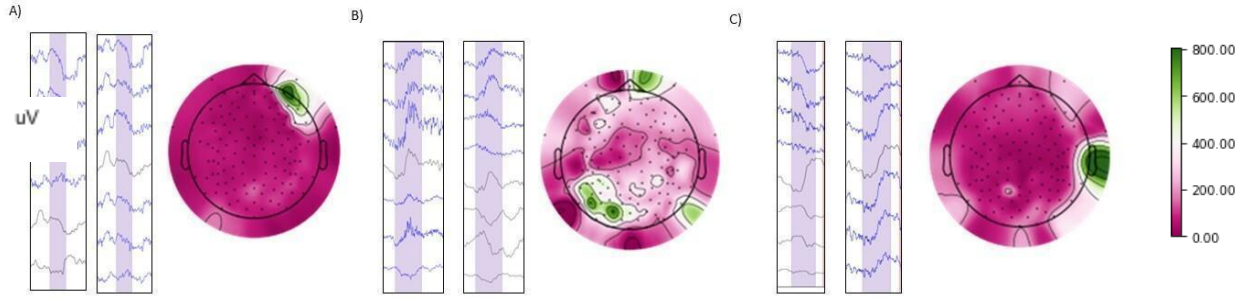


Fig.4: Topographic plots used to verify three eye-movement segments: A) high amplitude in the frontal region, as expected; B) high amplitude in the frontal as well as posterior regions; C) high amplitude in the lateral central region overshadowing patterns in the frontal central region. In all three cases, the eye movement segment annotations were retained, because the artifact can be clearly distinguished as eye movement by visual inspection.

2.4 ICA-based artifact correction

ICA can be used to correct artifacts but depends on three assumptions: i) sources (S) of brain activity and artifact activity detected at the electrodes are linearly mixed and statistically

independent; ii) the number of channels is greater than the number of sources (components); iii) there are negligible propagation delays between source of the signal and the scalp [41].

The linear model for sources S obtained through ICA is given as:

$$X = AS \quad (1)$$

where X is the observed signals (i.e., the EEG) defined as a linear mixture of source signals s_i in $S = (s_1, s_2, s_3, \ldots, s_n)$ and A is the unknown mixing matrix. The sources S are approximated by finding the unmixing matrix A^{-1} :

$$S = A^{-1}X \quad (2)$$

such that the sources the sources s_i have a minimal mutual information.

There are several different implementations of ICA. We used a robust adaptive mixture ICA (AMICA) [42]. These ICA have already been computed and visually validated when the EEG-IP-L pipeline was run prior to this project. All components indicative of artifacts were selected for rejection (i.e., this ICA artifact rejection is not specific to eye movement artifacts). AMICA is available as part of EEGlab [18].

Fig.5 shows an example ICA decomposition of signal contaminated with eye movement artifact.

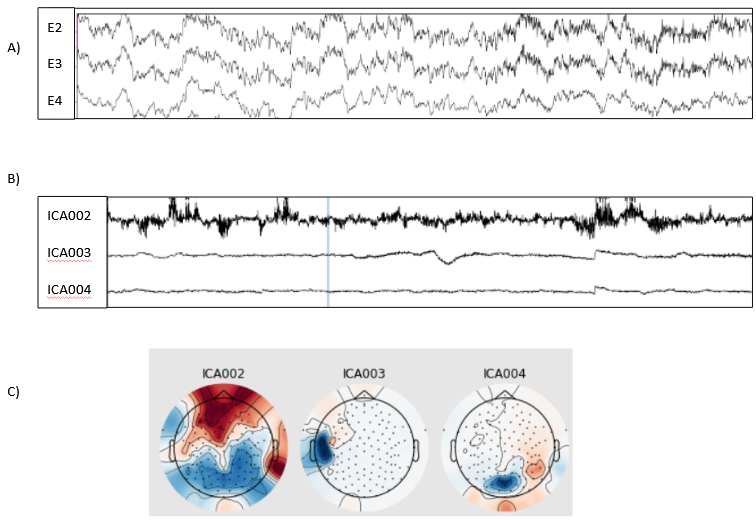


Fig.6. An example of ICA decomposition of ocular artifacts in infant EEG. A) Original signal contaminated with eye movement artifact. B) Independent components decomposing the signal into sources. C) Relative projection strengths of the components over all the channels.

2.5 AB-based artifact correction

The AB algorithm relies on a threshold θ to distinguish between brain signal and high amplitude artifacts. Since such a threshold is likely to be recording- and subject-dependent (e.g., due to differences in electrode impedance or head tissue conductivities causing systematic biases in EEG amplitude), we developed a procedure to calculate the optimal threshold θ. This is defined as the voltage at which eye movements would be maximally removed while minimizing distortion to clean segments. Prior to starting calculation, channels used for annotating eye movement and clean segments were examined for each recoding, so that only the ones with sufficient data for analysis would be used (Fig.3). Channels that remained noisy after interpolation (less than ~30s clean EEG) were not included in the calculation of optimal threshold. Table S2 lists the channels that were selected for each participant. For each channel, the following procedure was followed:

1. The maximum amplitude in each annotated segment (eye movement and clean) was calculated. These maximum amplitudes were used to construct a range of amplitudes, as

thresholds to be tested. The range consisted of all thresholds, in increments of 1 μV, from the lowest to the highest maximum amplitude recorded across channels.

2. For each threshold, the mean absolute error (MAE) for each annotation was calculated as follows:

$$e_{ti} = \frac{1}{N_i} \sum_{j=1}^{N_i} |x_O^{ij} - x_{AB}^{ij}| \quad (3)$$

where *e* is the total error, *t* is the threshold index, *i* denotes the index of the annotation, N_i stands for number of time samples in annotation *i*, *j* is time sample index, x_O and x_{AB} represent the amplitude of the original and the AB-processed signal, respectively.

3. The average MAE across all eye movement annotations and across all clean annotations were calculated.
4. The average MAE for eye movement and clean segments was normalized to be on the same scale (Fig.6):

$$\widetilde{e_{ti}} = \frac{(e_{ti} - min_t e_{ti})}{(max_t e_{ti} - min_t e_{ti})} \quad (4)$$

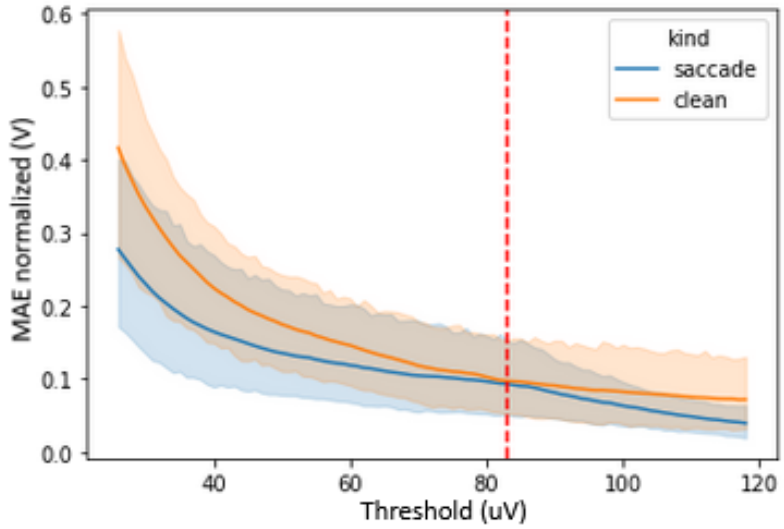


Fig.7: Normalized MAE across tested thresholds, for one channel in one recording. The shaded region represents the 95% confidence interval.

5. The total error at each threshold takes into account the MAE at that threshold for both eye movement and clean segments. Note that a large "error" for eye movement segments is desirable (i.e., the cleaned signals should be different than the raw signals) while a small error is desirable for the clean segments (i.e., the artifact rejection should not distort the signals in such epochs), hence the subtraction of the former and the addition of the latter type of errors. Also, note that an equal weight for the contribution of both types of errors is implicitly used by not further weighting these two types of errors.

$$E_t = 1 - \frac{1}{N_S}\sum_{i\in\Theta_S}\tilde{e_{ti}} + \frac{1}{N_C}\sum_{i\in\Theta_C}\tilde{e_{ti}} \quad (5)$$

6. The optimal threshold for the channel is the threshold corresponding to the minimum of all total errors calculated (Fig.7):

$$t_{opt} = \arg min_t E_t \quad (6)$$

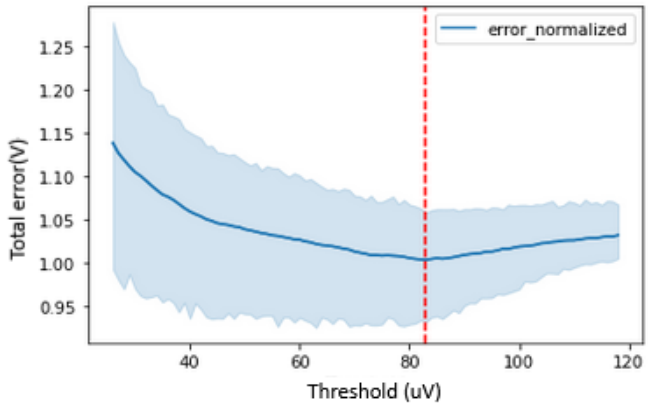


Fig.8: The optimal threshold for this channel is 83 µV, i.e., the threshold corresponding to the minimum of all total errors calculated.

Since the original algorithm uses a common threshold for all channels, the optimal threshold for the recording was calculated as the average optimal thresholds of all selected channels.

Following the setting of the optimal threshold θ, a matrix *Y* is constructed from the original matrix *X* by zeroing values exceeding θ. *X* and *Y* are two-dimensional matrices of sizes *n x T*, with n being the number of channels and *T* being the number of time points [19], [20]. A matrix B_{opt}, referred to as the blocking matrix, is constructed as follows:

$$B_{opt} = argmin_B E \|Bx_t - y_t\|_2^2 \quad (7)$$

where x_t and y_t refer to the t^{th} column of matrices *X* and *Y*, respectively. *E* stands for the expectation operator [43]. Additional details on the calculation of *B* are available in [15].

By performing this calculation, the algorithm produces an output where samples exceeding θ are removed, while those not exceeding θ remain un-distorted. This output may be examined for over-smoothing, and the threshold in the first step may be adjusted accordingly.

The MATLAB script for AB was obtained from the original authors [8]. Output from MATLAB was exported to Python for comparison with output from ICA. Portions of code written in MATLAB were integrated into the Python pipeline by running them from Python using the MATLAB engine.

2.6 Performance metrics

2.6.1 Visual inspection to compare correction by AB and ICA

Visual inspection of the eye movement epochs was done for one channel in each recording that displayed eye movement artifacts. For each recording, the channel where eye movement artifacts were most visible, was chosen. Table A3 displays the number of recordings for which each channel has been used for visual inspection. The inspection incorporated a blinded scoring of the level of correction achieved by AB and ICA. To make this annotation procedure systematic and efficient, we implemented a custom interactive window displaying the corresponding time series and accepting sequential annotations through keystrokes (Fig.8). The following procedure was used:

1) For each annotated eye movement segment, time series plots of the original signal, signal corrected with ICA, and signal corrected with AB were overlaid.

2) For each segment, AB- and ICA-corrected traces were randomly assigned one of two colors, red or blue

3) The red and blue plots were rated as having under-corrected or effectively corrected for the eye movement artifact. The authors discussed various examples and arrived at a consensus on criteria for effective correction vs under-correction. If an eye movement artifact was clearly visible on the plot after correction, it was rated as an under-correction.

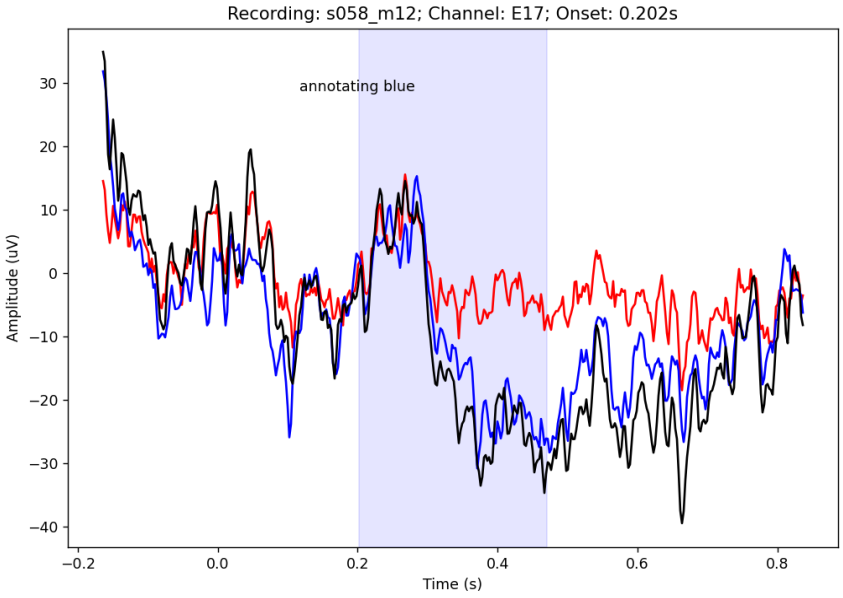
4) At the end of the scoring procedure, blinded scoring was automatically associated with their true labels (i.e., ICA or AB) based on a log of the random draws. Proportions of segments where AB and ICA corrected effectively or under-corrected for the eye movements were calculated.

The eye movement artifacts were not as distinctive in every recordings. Examples of ratings for such recordings are shown in Fig. S1.

Clean segments were not manually annotated as the ideal outcome is known (i.e., it is identical to the original signal since these segments are free from artifacts). Any deviation will be

considered overcorrection, and this deviation can be measure quantitatively, e.g., through the SNR.

A)



B)

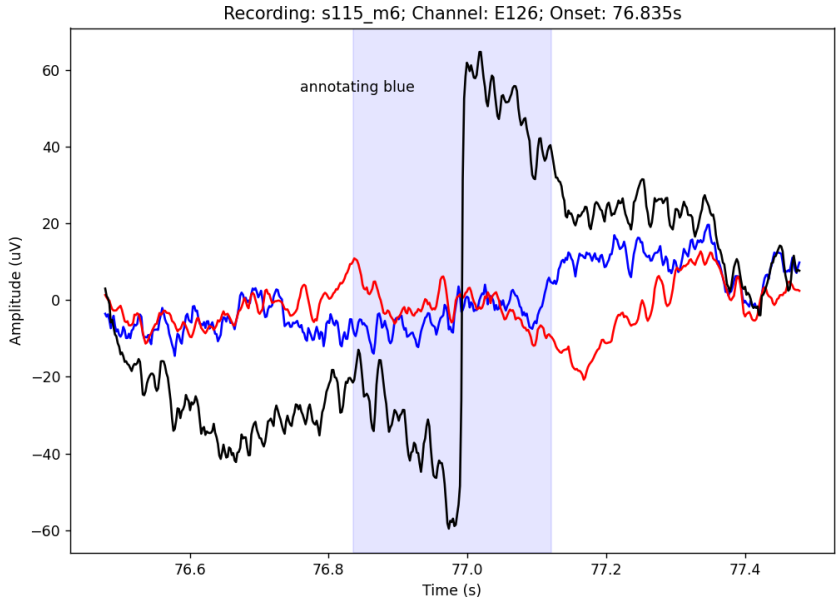


Fig.9: Example of visual inspection of an eye movement segment. Scoring was blinded by randomly assigning red or blue to signal corrected by AB and ICA. Black indicates the original signal. Ratings for correction were based on whether an eye movement was clearly visible after correction. In A), both algorithms have undercorrected for the eye movement. In B), both algorithms have effectively corrected for the eye movement.

2.6.2 Signal-to-noise ratio (SNR)

The SNR is a measure of signal quality of a time series. In our calculations, the signal refers to the processed EEG after applying a correction algorithm, while noise refers to the portion of EEG removed during preprocessing. We calculated SNR for each annotated eye movement and clean segment. Following the application of the algorithms, we expect the SNR to be lower in the eye movement segments than in the clean segments, because more noise is removed and hence the denominator is larger.

SNR was calculated for each eye movement and clean segment. To calculate the SNR, the root-mean-square-error (RMSE) for the segment processed with AB or ICA (signal) was first calculated using its classic definition:

$$RMSE_{Signal} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} y_i^2} \tag{8}$$

with y_i being the processed segment (signal) and n the number of samples in the segment. Then, the RMSE for noise removed in the segment was calculated:

$$RMSE_{Noise} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - x_i)^2} \tag{9}$$

with x_i being the original, unprocessed segment.

Finally, the SNR was calculated as the quotient of the two RMSE values, using a base 10 logarithmic scale and multiplying by 10 to express this ratio is decibels (dB), as is usual:

$$SNR = 10 * log(RMSE_{signal}/RMSE_{noise}) \tag{10}$$

A value of 0 dB indicates a signal and noise RMSE values of equal amplitude. A value of 10 dB (i.e., a signal 10 time as large as the noise) or more indicates a large SNR. Since we defined the noise as what was removed from the raw data by the artifact correction algorithm, a large "noise" (hence a low SNR) is desirable when assessing the performance of the algorithm on epochs with artifact, while this "noise" would ideally be zero (hence large SNR) for epochs

without artifacts, indicating that the algorithm did not distort clean signals. SNR calculations were implemented in Python.

2.6.3 Power spectral density (PSD)

The PSD is a measure of energy contained in the signal at different frequencies. It was calculated for eye movement and clean segments, for all channels and recording. For each recording, we first cropped individual eye movement and clean segments from the signal. They were then concatenated in two separate files, one for eye movement and one for clean segments. An average reference was used. The Welch method (the 'psd_welch' function in MNE-Python [24]) was used to estimate the spectrum. In short, a Hamming window taper was applied on the segmented windows with 50% overlap. A periodogram for each window was calculated and averaged over the epochs in each of multiple frequency bands, at each channel. PSD values were log10 transformed to account for skewed distributions.

The following frequency bands were chosen for analysis: delta (2-4 Hz), theta (4-6 Hz), low alpha (6-9 Hz), high alpha (9-13 Hz), beta (13-30 Hz), and gamma (30-48 Hz) [44].

2.6.4 Multiscale entropy (MSE)

MSE is a non-linear measure of signal complexity. We calculated MSE by first coarse graining the EEG signal for multiple scales [28]:

$$y_j^{\tau} = \frac{1}{\tau_{i=(j-1)_{\tau+1}}} \sum_{j\tau} x_i, 1 \leq j \leq \frac{N}{\tau} \tag{11}$$

where y_j is the coarse-grained series, $\boldsymbol{\tau}$ is the scale factor, and x_i is the original signal. A list of $\boldsymbol{\tau}$ was determined as a range between 0 and $\boldsymbol{\tau}_{max}$, determined as follows:

$$\tau_{max} = \frac{N}{m + 10} \tag{12}$$

where *N* is the length of the signal and *m* is the embedding dimension and is equal to 2.

Then, sample entropy [45], [46] was calculated for the coarse-grained signal at each scale:

$$SampEn(m,r,N) = -log\frac{\sum_{i=1}^{N-m}\sum_{j=1,j\neq i}^{N-m} freq_d\left[|x_{m+1}(j) - x_{m+1}(i)|\right] < r}{\sum_{i=1}^{N-m}\sum_{j=1,j\neq i}^{N-m} freq_d\left[|x_m(j) - x_m(i)|\right] < r} \quad (13)$$

where *r* is known as the tolerance and *d* the distance between time series *i* and *j*. MSE was then calculated as the sum of sample entropy across all scales. MSE calculations were implemented using the 'entropy_multiscale' function of the neurokit2 package on Python [47]. Statistical analyses were performed using the python libraries StatsModels and SciPy [48], [49].

Results

We report on results for the four metrics we previously defined to assess the efficacy of algorithms in removing eye movement artifacts without distorting clean signal.

3.1 Visual inspection

We performed a blinded comparison of the level of correction achieved by AB and ICA for eye movement segments. To compare the difference in proportions of under-corrected and effectively corrected by each algorithm, we performed a McNemar's test for difference in proportions. This non-parametric test was used as the data did not meet the assumption for normality.

In segments with eye movements, there was a lower proportion of instances where AB effectively corrected for ocular artifacts compared to ICA (Fig. 10). This difference was statistically significant (χ^2=62.77, df=1, p=1.182 x 10^{-8}.

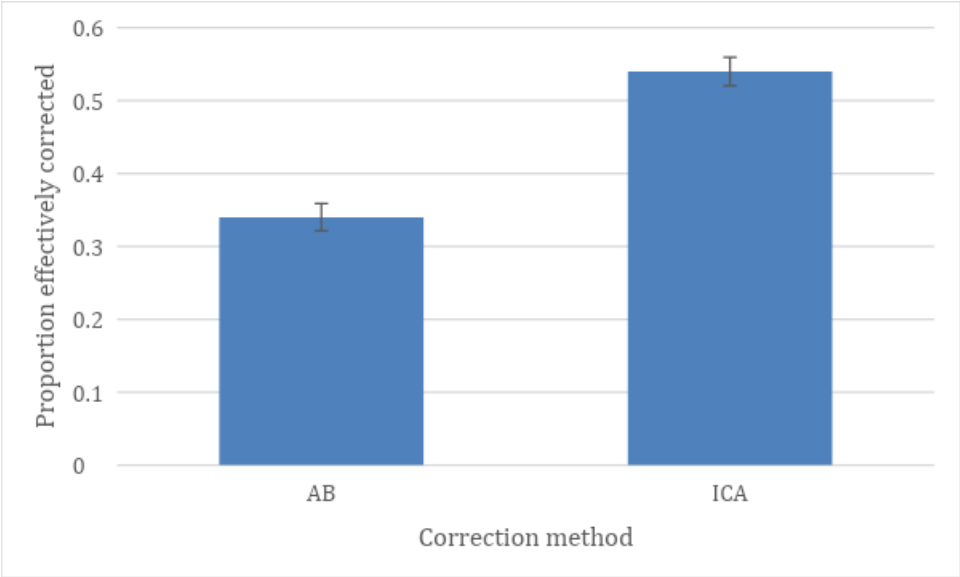


Fig.10: Proportion of all eye movement segments, where AB and ICA effectively corrected for the eye movement, based on visual inspection. Error bars indicate the standard error.

3.2 SNR

To assess the efficacy of ICA and AB in the time domain, we calculated SNR in eye movement segments and clean segments. As the assumption of normality was violated, a nonparametric Wilcoxon-signed rank test was used for testing difference in means. A cluster-based permutation test was applied for multiple comparisons across channels.

In both the eye movement and clean epochs, SNR was higher following AB than ICA (Fig.10). The difference is most pronounced in the left frontal, central and right-posterior regions. A cluster-based permutation test found that there was one significant cluster with a probability of $p<0.0001$. These findings do not support our predictions that both AB and ICA would correct equally effectively for eye movement segments. The trends suggest that ICA corrects for eye movement better than AB.

The higher SNR for AB in clean segments supports our prediction that this algorithm causes less distortion of clean EEG compared to ICA.

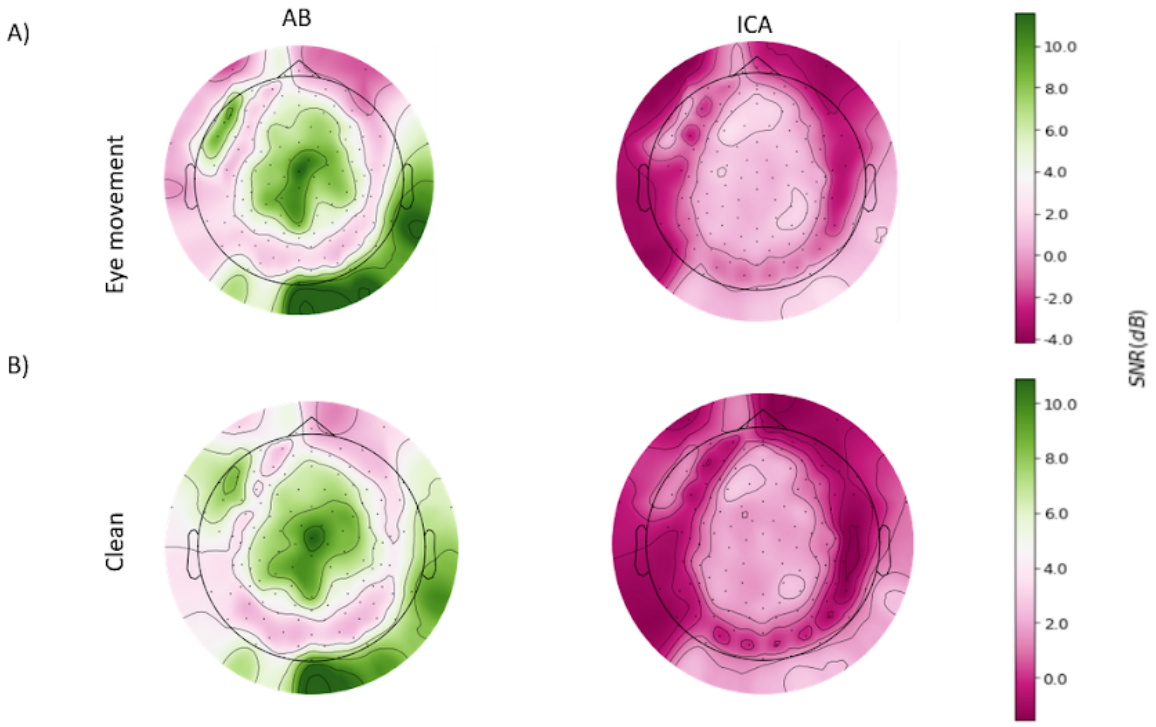


Fig.11: Scalp topographic plots of average SNR per channel across all participants, following AB-correction and SNR following ICA-correction in A) eye movement segments B) clean segments.

3.3 PSD

To assess the efficacy of AB and ICA in the frequency domain, we calculated log10PSD in the eye movement and clean segments of the original, AB-corrected, and ICA-corrected signals across different frequency bands between 0 and 48Hz. For each recording, eye movement and clean segments were concatenated prior to calculating log10PSD. The average length of concatenated eye movement segments was 3.29s, and that of concatenated clean segments was 27.37s. As the assumption for normality was violated, a Wilcoxon signed-rank test was performed. A cluster-based permutation test was applied for multiple comparisons across channels.
Fig.12 shows the difference in $\log_{10}$PSD between the original signal and the AB and ICA-corrected signals, for eye movement and clean segments.

In the eye movement segments, across all frequency bands, the $\log_{10}$PSD following AB is slightly lower than for the original signal (Fig.12A). Though the difference is small, it is statistically significant across all frequency bands. A cluster-based permutation test showed one cluster, with a positive sum of w values and a probability of $p<0.0001$, for each frequency band. The $\log_{10}$PSD following ICA is noticeably lower than that of the original signal (Fig.12A). A cluster-based permutation test showed one cluster, with a positive sum of w values and a probability of $p<0.0001$, for each frequency band. As a greater decrease was observed following ICA, the findings suggest that ICA corrects more effectively for eye movement artifacts. This does not support our prediction that the algorithms would equally effectively correct for eye movement segments.

In the clean segments, $\log_{10}$PSD following AB is slightly lower than that of the original signal across all frequency bands and regions of the scalp (Fig.12B). The $\log_{10}$PSD for ICA is also lower than that of the original signal across all frequency bands, more noticeably in the beta and gamma bands (Fig.12B). A cluster-based permutation test showed one cluster, with a probability of $p<0.0001$, for each frequency band. The findings do not support our prediction, that correction by AB would not alter $\log_{10}$PSD in clean segments. However, they do support the prediction that correction with ICA would cause distortion to clean segments.

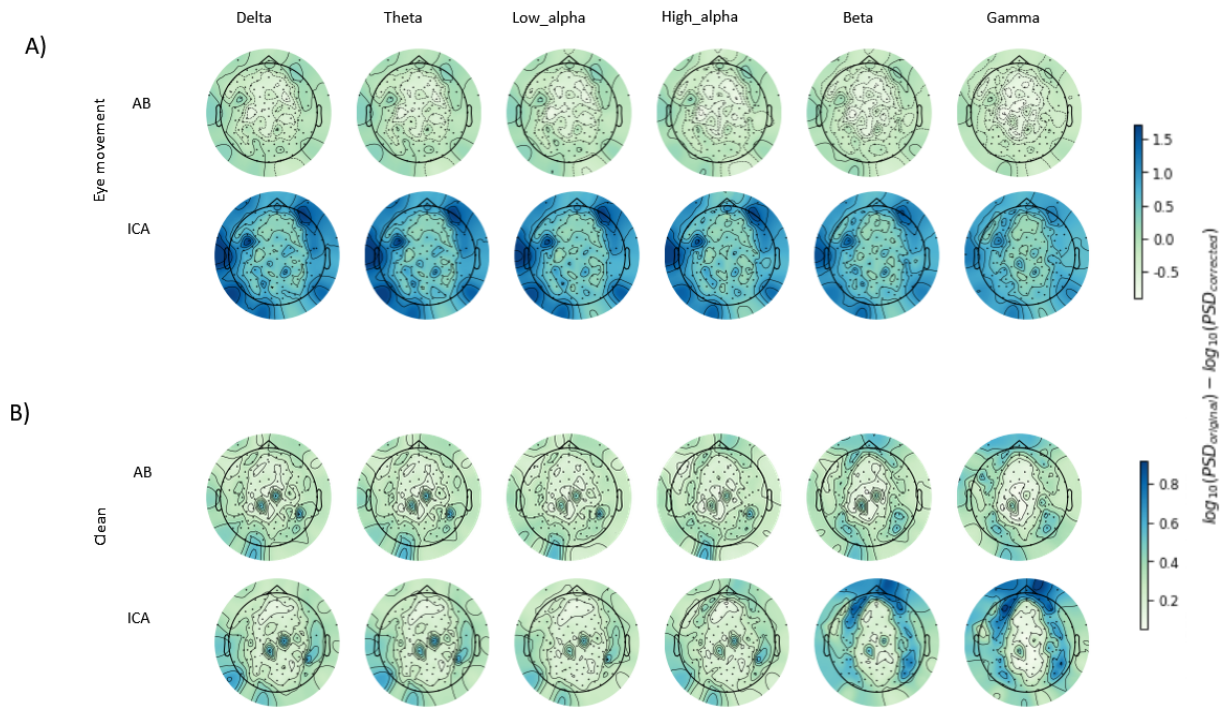


Fig.12: Scalp topographic plots depicting the differences in log_{10}PSD between original signal and processed signal in A) eye movement epochs and B) clean epochs. Log_{10}PSD values have been calculated by averaging across 50 participants, per frequency band, per channel.

3.4 MSE

We calculated MSE as a non-linear measure of signal efficacy, in selected channels (Fz, F7, F8, Fp1, Fp2, Pz, Cz). As the assumption for normality was violated, a Wilcoxon signed-rank test was performed. A cluster-based permutation test was applied for multiple comparisons.

In the eye movement segments, across all channels, MSE following both AB and ICA increases compared to the original signal, across 25 scales, indicating an equal amount of artifact correction. This increase is more pronounced in channels F7, F8, Oz and T6 (Fig 13). We calculated the mean MSE across all scales, for each of the select channels. Mean MSE following both AB and ICA increased, compared to the original signal. Cluster-based permutation tests for multiple comparisons found that the differences were statistically significant (p values for the difference

in MSE between original signal and AB: 0.0039, 0.0125, 0.2416, p values for the difference in MSE between original signal and ICA: 0.0206, 0.0084, 0.245).

Hence, these results support our prediction that AB and ICA would equally effectively correct for eye movement segments.

In the clean segments, we found mixed results. In eight channels (Fz, F7, F8, Fp1, Fp2, Pz, Cz, Oz), the MSE of AB across 25 scales is similar to that of the original signal, indicating negligible distortion. In two of the channels (T5, T6), MSE of AB is greater than that of the original signal, indicating some distortion. In five channels, the MSE following ICA across 25 scales is greater than that of the original signal, indicating some distortion. In one channel, the MSE following ICA across 25 scales was similar to that of the original signal, indicating negligible distortion (Fig 14). Mean MSE increased following AB compared to the original signals, for all channels, except Fz. However, a cluster-based permutation test for multiple comparisons showed that the difference was not statistically significant (p's > 0.05). Mean MSE following ICA increased in half the channels and decreased in half the channels. The cluster-based permutation test for multiple comparisons yielded zero clusters. The findings support our prediction that MSE would not significantly change following AB, and hence the hypothesis that AB would not distort clean segments. However, the findings regarding ICA are inconclusive, with no clusters found.

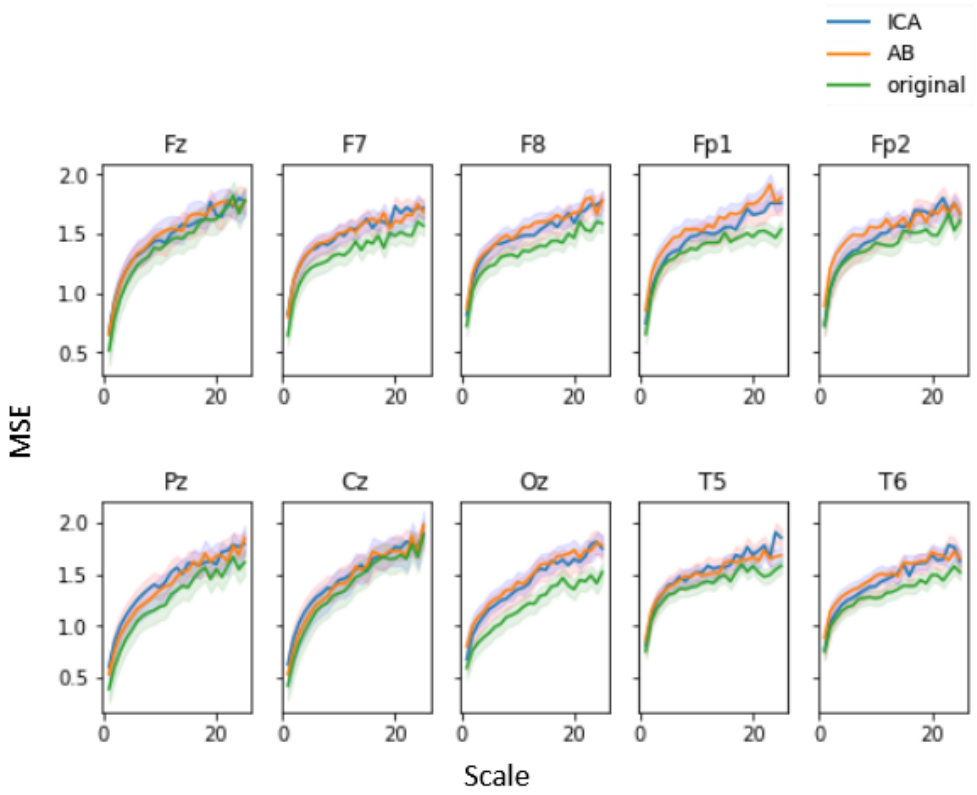


Fig 13: Scale-wise MSE following correction by ICA and AB, compared to the original signal, in eye movement segments.

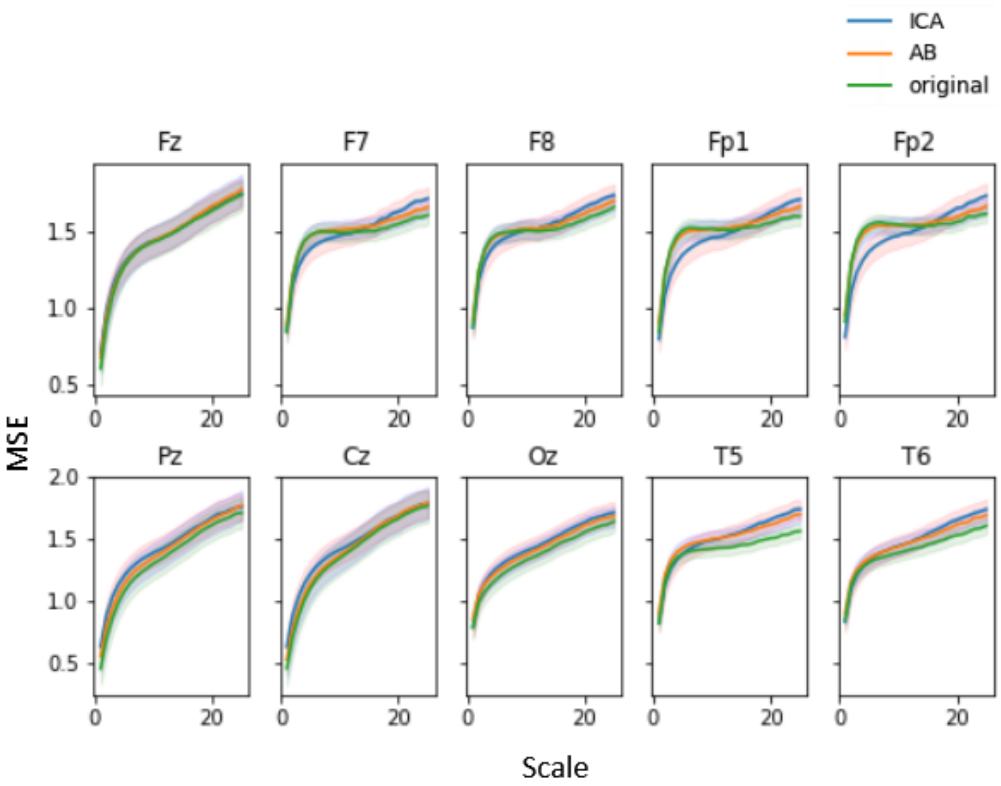


Fig 14: Scale-wise MSE following correction by ICA and AB, compared to the original signal, in clean segments.

Discussion

We compared AB and ICA for their effectiveness in the removal of eye movement artifacts from infant EEG (i.e., sensitivity) while not distorting clean signal (i.e., specificity). We relied on four key metrics to assess these outcomes: proportion of effectively corrected segments, signal-to-noise ratio (SNR), power spectral density (PSD) and multiscale entropy (MSE). Comparing the MSE and PSD following the two algorithms supported the first hypothesis we tested, that AB would correct for eye movements just as effectively as ICA. Comparing the SNR following the two algorithms supported the second hypothesis that AB would not distort brain activity, while ICA would. Table 3 below summarizes these findings.

Table 3: Summary of findings from the four metrics

	Proportion of effectively corrected segments	**SNR**	**PSD**	**MSE**
Hypothesis 1: AB will correct for eye movements just as effectively as ICA	No	No	Partial	Yes
Hypothesis 2: ICA but not AB will not distort brain activity	NA	Yes	No	Partial

More specifically, with regards to the removal of eye movement artifacts, we found that ICA was superior to AB as demonstrated by the proportion of effectively corrected eye movement segments, SNR and PSD. ICA effectively corrected for a higher proportion of eye movement segments compared to AB, as assessed by visual inspection. ICA-corrected eye movement segments had a lower SNR compared to AB-corrected eye movement segments. Our SNR calculation involved signal in the numerator and noise removed, in the denominator. Hence, a lower SNR indicates that more noise was removed by ICA. The PSD of both AB and ICA-corrected segments decreased compared to the original signal, though by a much greater amount for ICA, indicating that ICA was more superior. The scale-wise MSE of both AB and ICA-corrected segments increased by similar amounts, compared to the original signal, indicating either that there was equally effective correction by both algorithms. Hence, MSE was not sensitive to differences between AB and ICA in the processing of eye movement segments, indicating that MSE might be a robust measure of distortion in the clean segments.

With regards to the distortion of clean signal, we found that AB-corrected segments had a higher SNR compared to ICA-corrected segments, indicating more distortion by ICA. Both AB-corrected and ICA-corrected segments had a lower PSD compared to the original signal, indicating distortion by both AB and ICA. Scale-wise MSE following AB did not change significantly compared to the original signal, supporting the prediction that AB would not distort clean segments.

Our results from visual inspection are contrary to a prior visual comparison of AB and ICA conducted using real infant EEG that found that AB was more effective at removing eye

movement artifacts compared to ICA [15]. However, since this previous study relied only on one recording, these results are likely unreliable.

Our findings also suggest that the performance of the artifact correction algorithms is dependent on specific regions of recording. We obtained a lower SNR in the frontal channels compared to the central, parietal, and temporal channels, meaning that more artifacts were removed from the frontal channels. This observation is in agreement with the expectation that eye movement artifacts are dominant in the frontal regions. Many studies that have evaluated ICA and other algorithms for ocular artifact removal have calculated the SNR on simulated rather than real EEG, because the ground truth is unknown in real EEG. While a comparison with real EEG is essential because of its ecological validity, simulations provide complementary evidence because they allow to verify how well the algorithm performs knowing the ground truth about the (simulated) brain activity and eye activity [50]. A potential next step to strengthen our analyses is to compare AB and ICA using simulated EEG generated using ocular artifacts extracted from a real dataset. Simulated recordings would be generated by extracting segments of eye artifacts from real EEG recordings [51]. Our lab is currently collecting synchronized eye-tracking and EEG data. This will provide for a robust way to know which periods are associated with eye movements, and extract those, when creating simulated EEG recordings.

One other study has used SNR to evaluate the performance of different types of ICA algorithms for EOG artifact removal in real EEG [52]. In this study, the authors evaluated the performance of a novel ICA-based method on removing different types of physiological artifacts from EEG collected from adults. In their calculation of SNR, they regarded signal as a segment of variable

length immediately following a beep tone used to cue an eyeblink, and noise as the segment associated with 200 ms prior to the blink artifact. In our calculation, we defined the signal as the original EEG, and the noise as the difference between this original and the corrected segments. Due to the difference in calculation of SNR, we are unable to compare our results for SNR with this study.

PSD is the most commonly evaluated metric in studies using real EEG to evaluate artifact removal algorithms. We found that ICA reduced PSD compared to original signals in the eye movement segments. This reduction was notable in frontal channels, across all frequencies. The difference was approximately 1-1.5 $\mu V^2/Hz$, which is close to the magnitude of the difference between ICA-corrected and original signal in a prior study that evaluated ICA to correct ocular artifacts in EEG collected from adults [50]. However, we also found that ICA reduced power noticeably in the beta and gamma bands, in clean segments, in the frontal channels. This is contrary to the findings of the prior study [50]. The reduction in power at these higher frequencies may be related to components associated with other sources of noise, such as muscle movement [53]. Unfortunately, the manual selection of independent components to be rejected for cleaning the signals were done prior to this work and was not specific to EOG artifacts.

To our knowledge, this is the first study that calculated the difference in MSE between the original and corrected versions of eye movement artifact segments. However, there are prior studies that have calculated the MSE of eye movement artifacts, calculated MSE of infant EEG, or made recommendations with respect to the calculation of MSE in infant EEG. These recent studies have introduced different approaches and choices of parameters in the calculation of

MSE in infant EEG. One recommendation is to recalculate *r* at each scale, whereas we used a fixed r of 0.2 across all scales [54]. Another approach is to decompose each channel into six frequency bands prior to coarse-graining and calculating multiscale entropy [55]. A final recommendation is to calculate entropy on the residuals of the EEG signal rather than on the EEG signal itself [54]. This process is designed for use in scenarios where not sufficient data is available, as with infant EEG recordings. The residuals of the EEG signal for given participant are found as per the following procedure: 1) selecting the trials where the global field power (GFP) is equal to the median GFP for a given participant 2) subtracting the within-person average response across trials within each condition of the experiment.

This current study can support future decisions on whether to use ICA or AB in artifact-removal pipelines. To our knowledge, none of the pipelines discussed in the paper used AB. This is also the first study where four metrics have been used to evaluate eye movement artifact removal and non-distortion of clean EEG. Furthermore, consideration of the above methodological improvements will make a significant contribution to efforts to standardize EEG pre-processing pipelines. Such standardization is important for consistency in biomarker research.

One limitation to our methodology is that there was no independent measure of eye movements available [56]. Hence, annotations of eye movement segments were performed based on templates of the characteristic appearance of these artifacts. Moreover, only one rater performed these annotations. The visual inspection for proportion of effectively corrected eye movement segments were discussed and validated between the authors , and the annotation process iteratively improved. A potential next step to arrive at conclusive results, is to involve additional raters in the visual inspection process and assess the inter-rater agreement. A future

study involving eye-tracking with infants might provide a stronger basis to compare ocular artifact removal techniques.

Conclusion

This study applied AB and ICA on infant EEG to compare their effectiveness for removing eye movement artifact without distorting neural signals in EEG. While the results on eye movement artifact removal did not allow to select a clear winner between these two algorithms, our study showed that both algorithms distort clean EEG and that a better alternative must be devised. The findings yielded some insights regarding the benchmarking of artifact removal in real infant EEG. We identified venues to better utilize SNR for this application. We also identified ways to extract more information from MSE calculations. Future work should incorporate these modifications and either utilize infant EEG with direct recordings of eye movement or involve multiple raters for artifact annotation.

Appendices

Figures A1-A3 show further examples of visual inspection of an eye movement segments. Scoring was blinded by randomly assigning red or blue to signal corrected by AB or ICA. Black indicates the original signal. Ratings for correction were based on whether an eye movement was clearly visible after correction.

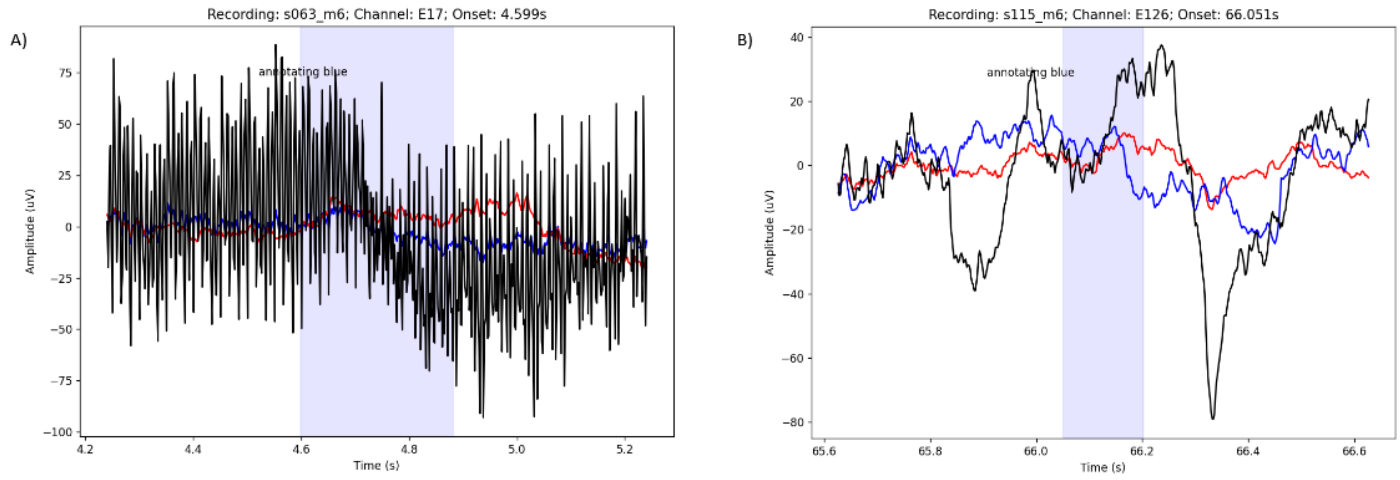


Fig. A1: Both algorithms effectively corrected for the eye movement.

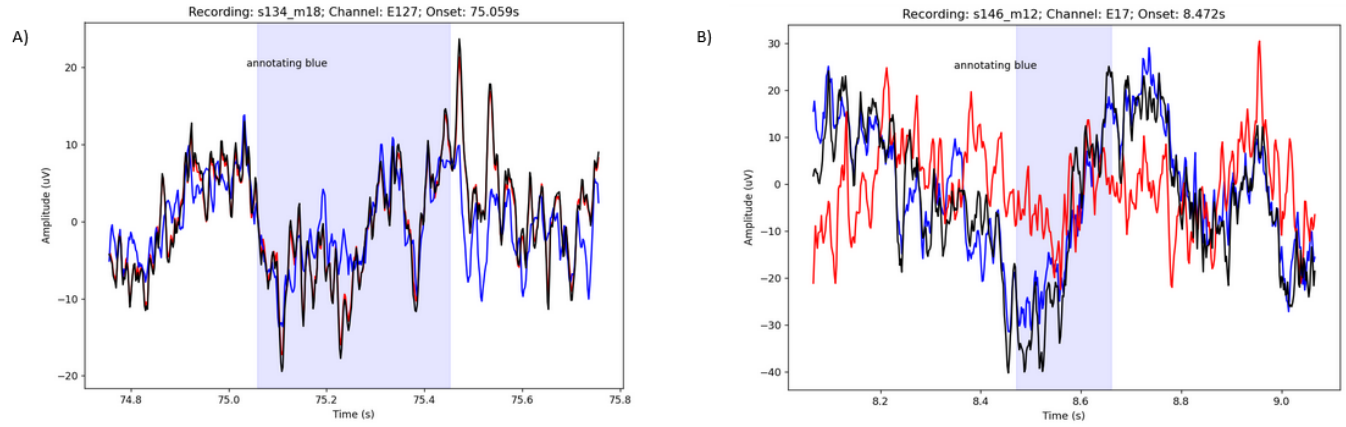


Fig. A2 Both algorithms undercorrected for the eye movement.

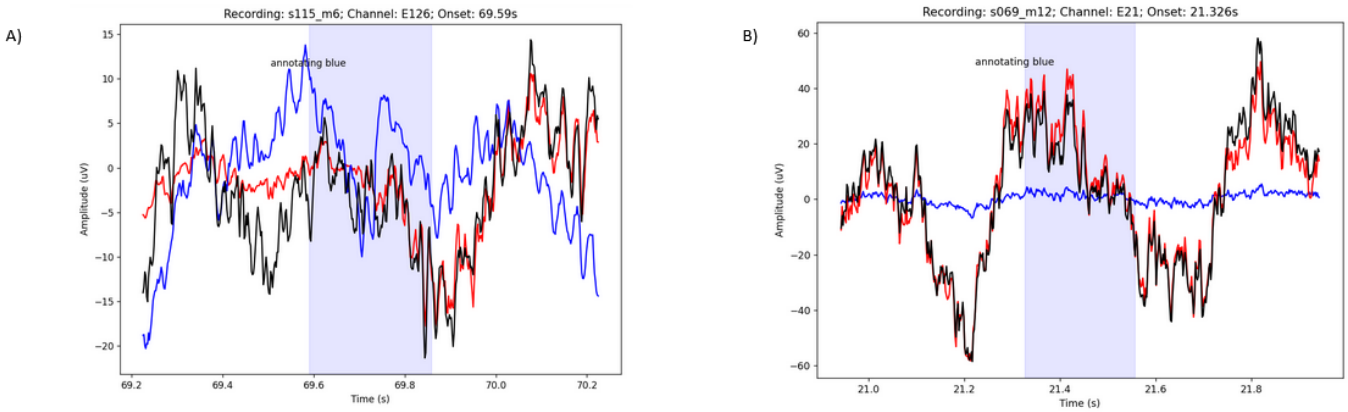


Fig. A3 Blue algorithm effectively corrected, while the red algorithm undercorrected for eye movement.

Table A.1: Number of recordings by participant age, risk group and site

Site	6m EEG	12m EEG	18m EEG
At-Risk	91	84	36
Seattle	39	37	36
London	52	47	
Control	88	81	40
Seattle	42	38	40
London	46	43	
Total	179	165	76

Table A.2: Channels used for optimal threshold calculation, per recording.

Recording ID	Channels
s058_m12	E1,E8,E14,E17,E21,E25,E32,E125,E126,E127,E128,E11,E62,E75,Cz
s059_m12	E1,E17,E21,E25,E32,E128,E11,E62,E75,Cz
s060_m12	E17,E21,E25,E32,E11,E62,E75,Cz
s062_m12	E14,E17,E21,E25,E32,E128,E11,E62,E75,Cz
s063_m6	E17,E21,E25,E32,E128,E11,E62,E75,Cz
s063_m12	E17,E32,E128,E11,E62,E75,Cz
s069_m12	E17,E21,E25,E32,E127,E128,E11,E62,E75,Cz
s134_m18	E17,E21,E25,E32,E127,E128,E11,E62,E75,Cz
s067_m18	E17,E32,E125, E128,E11,E62,E75,Cz
s085_m12	E17,E21,E25,E32,E128,E11,E62,E75,Cz
s102_m18	E1,E8,E14,E17,E21,E32,E125,E126,E128,E11,E62,E75,Cz
s106_m18	E17,E21,E25,E32,,E128,E11,E62,E75,Cz
s115_m6	E1,E8,E17,E21,E25,E32,E125,E126,E127,E128,E11,E62,E75,Cz
s138_m12	E1,E17,E21,E25,E32,E125,E128,E11,E62,E75,Cz
s146_m12	E17, E32, E128,E11,E62,E75,Cz
s713_m6	E17, E11, E62, E75, Cz
s741_m6	E1,E17,E128,E11,E62,E75,Cz
s702_m6	E17, E11, E62, E75, Cz
s147_m6	E1,E8,E14,E17,E21,E25,E32,E125,E126,E127,E128,E11,E62,E75,Cz
s147_m12	E17,E21,E25,E32, E127,E128,E11,E62,E75,Cz
s161_m6	E17,E21,E32,E125,E128,E11,E62,E75,Cz
s166_m6	E17,E21,E32,E128,E11,E62,E75,Cz

s937_m12	E17,E21,E32,E128,E11,E62,E75,Cz
s131_m18	E17,E21,E25, E32,E128,E11,E62,E75,Cz
s137_m6	E17,E21,E25, E32,E128,E11,E62,E75,Cz
s164_m6	E17,E21,E25,E32,E128,E11,E62,E75,Cz
s129_m12	E17,E21,E25,E32,E128,E11,E62,E75,Cz
s143_m12	E17,E21,E25,E32,E126,E127,E128,E11,E62,E75,Cz
s136_m6	E17,E21,E25,E32,E11,E62,E75,Cz
s107_m6	E1, E17, E32, E11,E62,E75,Cz
s730_m6	E32,E128,E11, E62, E75, Cz
s711_m6	E1,E8,E14,E17,E21,E25,E32,E125,E126,E127,E128,E11,E62,E75,Cz
s110_m18	E17,E21,E25,E32,E128,E11,E62,E75,Cz
s102_m12	E1,E17,E21,E125, E128,E11,E62,E75,Cz
s163_m6	E17,E21,E32,E11,E62,E75,Cz
s051_m6	E17,E21,E32,E128,E11,E62,E75,Cz
s937_m18	E32,E128,E11,E62,E75,Cz
s915_m6	E21, E32,E128,E11,E62,E75,Cz
s062_m18	E1,E8,E14,E17,E21,E25,E32,E125,E126,E127,E128,E11,E62,E75,Cz
s057_m12	E1,E8,E14,E17,E21,E25,E32,E125,E126,E127,E128,E11,E62,E75,Cz
s065_m6	E17,E21,E25,E32,E128,E11,E62,E75,Cz
s128_m6	E1,E8,E14,E17,E21,E25,E32,E125,E126,E128,E11,E62,E75,Cz
s742_m6	E17,E125,E11,E62,E75,Cz
s712_m6	E1,E32,E125,E128,E11,E62,E75,Cz
s747_m6	E17,E128,E11,E62,E75,Cz
s721_m6	E1,E8,E17,E25,E32,E125,E126,E127,E128,E11,E62,E75,Cz
s729_m6	E17,E11,E62,E75,Cz
s734_m6	E17,E125, E11,E62,E75,Cz

s716_m6	E1, E17, E11,E62,E75,Cz
s166_m12	E1,E8,E14,E17,E21,E25,E32,E125,E126,E127,E128,E11,E62,E75,Cz

Table A.3: Channels used for visual inspection of eye movement segments.

Channel	Number of recordings where the channel was used for visual inspection
E1	4
E8	1
E17	11
E21	6
E25	3
E32	2
E125	6
E126	3
E127	3
E128	11

References:

[1] V. Noreika, S. Georgieva, S. Wass, and V. Leong, “14 challenges and their solutions for conducting social neuroscience and longitudinal EEG research with infants,” *Infant Behavior and Development*, vol. 58, p. 101393, Feb. 2020, doi: 10.1016/j.infbeh.2019.101393.

[2] “Frontiers | The Harvard Automated Processing Pipeline for Electroencephalography (HAPPE): Standardized Processing Software for Developmental and High-Artifact Data,” Oct. 24, 2022. https://www.frontiersin.org/articles/10.3389/fnins.2018.00097/full (accessed Oct. 23, 2022).

[3] I. Marriott Haresign *et al.*, “Automatic classification of ICA components from infant EEG using MARA,” *Developmental Cognitive Neuroscience*, vol. 52, p. 101024, Dec. 2021, doi: 10.1016/j.dcn.2021.101024.

[4] R. Debnath, G. A. Buzzell, S. Morales, M. E. Bowers, S. C. Leach, and N. A. Fox, “The Maryland analysis of developmental EEG (MADE) pipeline,” *Psychophysiology*, vol. 57, no. 6, p. e13580, 2020, doi: 10.1111/psyp.13580.

[5] J. A. Desjardins, S. van Noordt, S. Huberty, S. J. Segalowitz, and M. Elsabbagh, “EEG Integrated Platform Lossless (EEG-IP-L) pre-processing pipeline for objective signal quality assessment incorporating data annotation and blind source separation,” *Journal of Neuroscience Methods*, vol. 347, p. 108961, Jan. 2021, doi: 10.1016/j.jneumeth.2020.108961.

[6] V. P. Kumaravel, E. Farella, E. Parise, and M. Buiatti, “NEAR: An artifact removal pipeline for human newborn EEG data,” *Developmental Cognitive Neuroscience*, vol. 54, p. 101068, Apr. 2022, doi: 10.1016/j.dcn.2022.101068.

[7] “Adjusting ADJUST: Optimizing the ADJUST algorithm for pediatric data using geodesic nets - Leach - 2020 - Psychophysiology - Wiley Online Library,” Oct. 24, 2022. https://onlinelibrary.wiley.com/doi/abs/10.1111/psyp.13566 (accessed Oct. 23, 2022).

[8] N. Mourad, J. P. Reilly, H. de Bruin, G. Hasey, and D. MacCrimmon, “A Simple and Fast Algorithm for Automatic Suppression of High-Amplitude Artifacts in EEG Data,” in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP ’07*, Apr. 2007, p. I-393-I–396. doi: 10.1109/ICASSP.2007.366699.

[9] J. Wang, J. Barstein, L. E. Ethridge, M. W. Mosconi, Y. Takarae, and J. A. Sweeney, “Resting state EEG abnormalities in autism spectrum disorders,” *J Neurodev Disord*, vol. 5, no. 1, Art. no. 1, Sep. 2013, doi: 10.1186/1866-1955-5-24.

[10] E. V. Orekhova *et al.*, “EEG hyper-connectivity in high-risk infants is associated with later autism,” *Journal of Neurodevelopmental Disorders*, vol. 6, no. 1, p. 40, Nov. 2014, doi: 10.1186/1866-1955-6-40.

[11] E. J. H. Jones, K. Venema, R. Lowy, R. K. Earl, and S. J. Webb, “Developmental changes in infant brain activity during naturalistic social experiences,” *Dev Psychobiol*, vol. 57, no. 7, pp. 842–853, Nov. 2015, doi: 10.1002/dev.21336.

[12] A. L. Tierney, L. Gabard-Durnam, V. Vogel-Farley, H. Tager-Flusberg, and C. A. Nelson, “Developmental trajectories of resting EEG power: an endophenotype of autism spectrum disorder,” *PLoS One*, vol. 7, no. 6, p. e39127, 2012, doi: 10.1371/journal.pone.0039127.

[13] G. Righi, A. L. Tierney, H. Tager-Flusberg, and C. A. Nelson, “Functional connectivity in the first year of life in infants at risk for autism spectrum disorder: an EEG study,” *PLoS One*, vol. 9, no. 8, p. e105176, 2014, doi: 10.1371/journal.pone.0105176.

[14] B. C. N. Müller, N. Kühn-Popp, J. Meinhardt, B. Sodian, and M. Paulus, “Long-term stability in children’s frontal EEG alpha asymmetry between 14-months and 83-months,” *Int J Dev Neurosci*, vol. 41, pp. 110–114, Apr. 2015, doi: 10.1016/j.ijdevneu.2015.01.002.
[15] T. Fujioka, N. Mourad, C. He, and L. J. Trainor, “Comparison of artifact correction methods for infant EEG applied to extraction of event-related potential signals,” *Clinical Neurophysiology*, vol. 122, no. 1, Art. no. 1, Jan. 2011, doi: 10.1016/j.clinph.2010.04.036.
[16] D. Kampis, E. Parise, G. Csibra, and Á. M. Kovács, “On potential ocular artefacts in infant electroencephalogram: a reply to comments by Köster,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 283, no. 1835, Art. no. 1835, Jul. 2016, doi: 10.1098/rspb.2016.1285.
[17] M. A. Bell and K. Cuevas, “Using EEG to Study Cognitive Development: Issues and Practices,” *J Cogn Dev*, vol. 13, no. 3, pp. 281–294, Jul. 2012, doi: 10.1080/15248372.2012.691143.
[18] “SCCN: Independent Component Labeling.” https://labeling.ucsd.edu/tutorial/labels (accessed Nov. 16, 2022).
[19] L. J. Gabard-Durnam, C. Wilkinson, K. Kapur, H. Tager-Flusberg, A. R. Levin, and C. A. Nelson, “Longitudinal EEG power in the first postnatal year differentiates autism outcomes,” *Nat Commun*, vol. 10, no. 1, p. 4188, Sep. 2019, doi: 10.1038/s41467-019-12202-9.
[20] M. Plöchl, J. Ossandón, and P. König, “Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data,” *Frontiers in Human Neuroscience*, vol. 6, 2012, Accessed: Jan. 10, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2012.00278
[21] S. B. Agyei, F. R. R. van der Weel, and A. L. H. van der Meer, “Longitudinal study of preterm and full-term infants: High-density EEG analyses of cortical activity in response to visual motion,” *Neuropsychologia*, vol. 84, pp. 89–104, Apr. 2016, doi: 10.1016/j.neuropsychologia.2016.02.001.
[22] E. Flaten, S. A. Marshall, A. Dittrich, and L. J. Trainor, “Evidence for top-down metre perception in infancy as shown by primed neural responses to an ambiguous rhythm,” *Eur J Neurosci*, vol. 55, no. 8, pp. 2003–2023, Apr. 2022, doi: 10.1111/ejn.15671.
[23] A. Leleu *et al.*, “Maternal odor shapes rapid face categorization in the infant brain,” *Developmental Science*, vol. 23, no. 2, p. e12877, 2020, doi: 10.1111/desc.12877.
[24] “Frontiers | A Novel Method Based on Combination of Independent Component Analysis and Ensemble Empirical Mode Decomposition for Removing Electrooculogram Artifacts From Multichannel Electroencephalogram Signals,” Oct. 24, 2022. https://www.frontiersin.org/articles/10.3389/fnins.2021.729403/full (accessed Oct. 23, 2022).
[25] S. Romero *et al.*, “Evaluation of an automatic ocular filtering method for awake spontaneous EEG signals based on independent component analysis,” *Conf Proc IEEE Eng Med Biol Soc*, vol. 2004, pp. 925–928, 2004, doi: 10.1109/IEMBS.2004.1403311.
[26] S. Mariani, A. F. T. Borges, T. Henriques, A. L. Goldberger, and M. D. Costa, “Use of multiscale entropy to facilitate artifact detection in electroencephalographic signals,” *Annu Int Conf IEEE Eng Med Biol Soc*, vol. 2015, pp. 7869–7872, 2015, doi: 10.1109/EMBC.2015.7320216.
[27] B. Fortune, X. Zhang, D. C. Hood, S. Demirel, and C. A. Johnson, “Normative ranges and specificity of the multifocal VEP,” *Doc Ophthalmol*, vol. 109, no. 1, pp. 87–100, Jul. 2004, doi: 10.1007/s10633-004-3300-5.

[28] M. Costa, A. L. Goldberger, and C.-K. Peng, “Multiscale entropy analysis of complex physiologic time series,” *Phys Rev Lett*, vol. 89, no. 6, p. 068102, Aug. 2002, doi: 10.1103/PhysRevLett.89.068102.
[29] R. Mahajan and B. I. Morshed, “Unsupervised Eye Blink Artifact Denoising of EEG Data with Modified Multiscale Sample Entropy, Kurtosis, and Wavelet-ICA,” *IEEE Journal of Biomedical and Health Informatics*, vol. 19, no. 1, pp. 158–165, Jan. 2015, doi: 10.1109/JBHI.2014.2333010.
[30] S. van Noordt *et al.*, “EEG-IP: an international infant EEG data integration platform for the study of risk and resilience in autism and related conditions,” *Mol Med*, vol. 26, no. 1, Art. no. 1, May 2020, doi: 10.1186/s10020-020-00149-3.
[31] M. Elsabbagh *et al.*, “Infant neural sensitivity to dynamic eye gaze is associated with later emerging autism,” *Curr Biol*, vol. 22, no. 4, pp. 338–342, Feb. 2012, doi: 10.1016/j.cub.2011.12.056.
[32] “Reduced engagement with social stimuli in 6-month-old infants with later autism spectrum disorder: a longitudinal prospec/tive study of infants at high familial risk - Google Search.” https://www.google.com/search?q=Reduced+engagement+with+social+stimuli+in+6-month-old+infants+with+later+autism+spectrum+disorder%3A+a+longitudinal+prospec%2Ftive+study+of+infants+at+high+familial+risk&rlz=1C1GCEA_enCA981CA981&oq=Reduced+engagement+with+social+stimuli+in+6-month-old+infants+with+later+autism+spectrum+disorder%3A+a+longitudinal+prospec%2Ftive+study+of+infants+at+high+familial+risk&aqs=chrome..69i57.301j0j9&sourceid=chrome&ie=UTF-8 (accessed Mar. 16, 2023).
[33] E. J. H. Jones, K. Venema, R. K. Earl, R. Lowy, and S. J. Webb, “Infant social attention: an endophenotype of ASD-related traits?,” *J Child Psychol Psychiatry*, vol. 58, no. 3, pp. 270–281, Mar. 2017, doi: 10.1111/jcpp.12650.
[34] C. R. Pernet *et al.*, “EEG-BIDS, an extension to the brain imaging data structure for electroencephalography,” *Scientific Data*, vol. 6, no. 1, Art. no. 1, Jun. 2019, doi: 10.1038/s41597-019-0104-8.
[35] K. J. Gorgolewski *et al.*, “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments,” *Scientific Data*, vol. 3, no. 1, Art. no. 1, Jun. 2016, doi: 10.1038/sdata.2016.44.
[36] E. Larson *et al.*, “MNE-Python.” Zenodo, Feb. 23, 2023. doi: 10.5281/zenodo.7671973.
[37] A. Gramfort *et al.*, “MEG and EEG data analysis with MNE-Python,” *Frontiers in Neuroscience*, vol. 7, 2013, Accessed: Mar. 16, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnins.2013.00267
[38] A. López, F. J. Martin, and O. Postolache, “An Affordable Method for Evaluation of Ataxic Disorders Based on Electrooculography,” *Sensors*, vol. 19, p. 3756, Aug. 2019, doi: 10.3390/s19173756.
[39] K. McEvoy, K. Hasenstab, D. Senturk, A. Sanders, and S. S. Jeste, “Physiologic artifacts in resting state oscillations in young children: methodological considerations for noisy data,” *Brain Imaging Behav*, vol. 9, no. 1, Art. no. 1, Mar. 2015, doi: 10.1007/s11682-014-9343-7.
[40] “https://eegatlas-online.com/index.php/en/.”
[41] A. J. Bell and T. J. Sejnowski, “An information-maximization approach to blind separation and blind deconvolution,” *Neural Comput*, vol. 7, no. 6, pp. 1129–1159, Nov. 1995, doi: 10.1162/neco.1995.7.6.1129.
[42] J. Palmer, K. Kreutz-Delgado, and S. Makeig, “AMICA: An Adaptive Mixture of Independent Component Analyzers with Shared Components,” Jan. 2011.

[43] “Parent-delivered early intervention in infants at risk for ASD: Effects on electrophysiological and habituation measures of social attention: Intervention in infants at risk for ASD - Google Search.” https://www.google.com/search?q=Parent-delivered+early+intervention+in+infants+at+risk+for+ASD%3A+Effects+on+electrophysiological+and+habituation+measures+of+social+attention%3A+Intervention+in+infants+at+risk+for+ASD&rlz=1C1GCEA_enCA981CA981&oq=Parent-delivered+early+intervention+in+infants+at+risk+for+ASD%3A+Effects+on+electrophysiological+and+habituation+measures+of+social+attention%3A+Intervention+in+infants+at+risk+for+ASD&aqs=chrome..69i57.307j0j9&sourceid=chrome&ie=UTF-8 (accessed Mar. 16, 2023).
[44] “Association between spectral electroencephalography power and autism risk and diagnosis in early development - Huberty - 2021 - Autism Research - Wiley Online Library,” Oct. 24, 2022. https://onlinelibrary.wiley.com/doi/full/10.1002/aur.2518 (accessed Oct. 23, 2022).
[45] J. S. Richman and J. R. Moorman, “Physiological time-series analysis using approximate entropy and sample entropy,” *Am J Physiol Heart Circ Physiol*, vol. 278, no. 6, pp. H2039-2049, Jun. 2000, doi: 10.1152/ajpheart.2000.278.6.H2039.
[46] A. Delgado-Bonal and A. Marshak, “Approximate Entropy and Sample Entropy: A Comprehensive Tutorial,” *Entropy (Basel)*, vol. 21, no. 6, p. 541, May 2019, doi: 10.3390/e21060541.
[47] D. Makowski *et al.*, “NeuroKit2: A Python toolbox for neurophysiological signal processing,” *Behav Res*, vol. 53, no. 4, pp. 1689–1696, Aug. 2021, doi: 10.3758/s13428-020-01516-y.
[48] S. Seabold and J. Perktold, “Statsmodels: Econometric and Statistical Modeling with Python,” presented at the Python in Science Conference, Austin, Texas, 2010, pp. 92–96. doi: 10.25080/Majora-92bf1922-011.
[49] P. Virtanen *et al.*, “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nat Methods*, vol. 17, no. 3, Art. no. 3, Mar. 2020, doi: 10.1038/s41592-019-0686-2.
[50] S. Romero, M. A. Mañanas, and M. J. Barbanoj, “A comparative study of automatic techniques for ocular artifact reduction in spontaneous EEG signals based on clinical target variables: a simulation case,” *Comput Biol Med*, vol. 38, no. 3, pp. 348–360, Mar. 2008, doi: 10.1016/j.compbiomed.2007.12.001.
[51] Q. Shi *et al.*, “High-speed ocular artifacts removal of multichannel EEG based on improved moment matching,” *J Neural Eng*, vol. 18, no. 5, Sep. 2021, doi: 10.1088/1741-2552/ac1d5a.
[52] G. Tamburro, P. Fiedler, D. Stone, J. Haueisen, and S. Comani, “A new ICA-based fingerprint method for the automatic removal of physiological artifacts from EEG recordings,” *PeerJ*, vol. 6, p. e4380, 2018, doi: 10.7717/peerj.4380.
[53] S. Georgieva, S. Lester, V. Noreika, M. N. Yilmaz, S. Wass, and V. Leong, “Toward the Understanding of Topographical and Spectral Signatures of Infant Movement Artifacts in Naturalistic EEG,” *Front Neurosci*, vol. 14, p. 352, 2020, doi: 10.3389/fnins.2020.00352.
[54] M. H. Puglia, J. S. Slobin, and C. L. Williams, “The automated preprocessing pipe-line for the estimation of scale-wise entropy from EEG data (APPLESEED): Development and validation for use in pediatric populations,” *Dev Cogn Neurosci*, vol. 58, p. 101163, Dec. 2022, doi: 10.1016/j.dcn.2022.101163.
[55] W. J. Bosl, H. Tager-Flusberg, and C. A. Nelson, “EEG Analytics for Early Detection of Autism Spectrum Disorder: A data-driven approach,” *Sci Rep*, vol. 8, no. 1, Art. no. 1, Dec. 2018, doi: 10.1038/s41598-018-24318-x.

[56] “The EEGLAB News.” https://sccn.ucsd.edu/eeglab/eeglab_news/14/Q_and_A.php (accessed Apr. 17, 2023).
[57] C. O’Reilly, J. D. Lewis, and M. Elsabbagh, “Is functional brain connectivity atypical in autism? A systematic review of EEG and MEG studies,” *PLoS One*, vol. 12, no. 5, p. e0175870, 2017, doi: 10.1371/journal.pone.0175870.