Brain Stimulation 17 (2024) 10–18

Contents lists available at ScienceDirect

Brain Stimulation

journal homepage: http://www.journals.elsevier.com/brain-stimulation

Measuring the accuracy of ICA-based artifact removal from TMS-evoked potentials

Iiris Atti [a,1], Paolo Belardinelli [b,c], Risto J. Ilmoniemi [a], Johanna Metsomaa [a,*,1]

[a] Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Finland
[b] Center for Mind/Brain Sciences - CIMeC, University of Trento, Italy
[c] Department of Neurology & Stroke, University of Tübingen, Germany

ARTICLE INFO

Keywords:
Artifact
Electroencephalography
Event-related potentials
Independent component analysis
Transcranial magnetic stimulation

ABSTRACT

Background: The analysis and interpretation of transcranial magnetic stimulation (TMS)-evoked potentials (TEPs) relies on successful cleaning of the artifacts, which typically mask the early (0–30 ms) TEPs. Independent component analysis (ICA) is possibly the single most utilized methodology to clean these signals.
Objective: ICA-based cleaning is reliable provided that the input data are composed of independent components. Differently, in case the underlying components are to some extent dependent, ICA algorithms may yield erroneous estimates of the components, resulting in incorrectly cleaned data. We aim to ascertain whether TEP signals are suited for ICA.
Methods: We present a systematic analysis of how the properties of simulated artifacts imposed on measured artifact-free TEPs affect the ICA results. The variability of the artifact waveform over the recorded trials is varied from deterministic to stochastic. We measure the accuracy of ICA-based cleaning for each level of variability.
Results: Our findings indicate that, when the trial-to-trial variability of an artifact component is small, which can result in dependencies between underlying components, ICA-based cleaning biases towards eliminating also non-artifactual TEP data. We also show that the variability can be measured using the ICA-derived components, which in turn allows us to estimate the cleaning accuracy.
Conclusion: As TEP artifacts tend to have small trial-to-trial variability, one should be aware of the possibility of eliminating brain-derived EEG when applying ICA-based cleaning strategies. In practice, after ICA, the artifact component variability can be measured, and it predicts to some extent the cleaning reliability, even when not knowing the clean ground-truth data.

1. Introduction

Transcranial magnetic stimulation (TMS) used jointly with electroencephalography (EEG) has proven to be an effective tool in studying neuronal excitability and connectivity of the brain [13,1]. It is well known that TMS does not exclusively activate cortical neurons; it can also cause activation of cranial muscles and other types of unwanted processes, which results in large artifacts in EEG recordings [18]. These artifacts are a major challenge for the interpretation of the obtained EEG signals and their neuronal origins.

Following a TMS pulse, the EEG data display different kinds of artifacts, especially within the first 30 ms after the stimulus [21]. These high-amplitude artifacts temporally overlap with TMS-evoked cortical potentials (TEPs), and thus, they may mask the underlying data of interest. In addition to the artifacts time-locked to the stimulus, the recordings may contain disturbances occurring at any point of time, such as those caused by eye blinks. The occurrence of this latter type of artifacts is unpredictable and not time-locked to the TMS pulse delivery.

In addition to the direct cortical perturbation, TMS delivery causes somatosensory and auditory stimulation for the subject, which gives rise to cortical neural activity recorded by EEG as so-called peripheral-evoked potentials (PEPs). Disentangling between the PEPs and the true TEPs is important to avoid misinterpretations because the latter ones typically represent the data of interest [4]. From this perspective, we can also view the neural processes representing PEPs as artifact generators which arise due to the TMS pulses.

* Corresponding author at: Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, P.O. Box 12200, FI-00076 Aalto, Finland.
E-mail address: johanna.metsomaa@aalto.fi (J. Metsomaa).
[1] These authors have equally contributed to the article.

https://doi.org/10.1016/j.brs.2023.12.001
Received 4 June 2023; Received in revised form 1 November 2023; Accepted 1 December 2023
Available online 10 December 2023

Independent component analysis (ICA) is a blind source separation method that looks for the underlying components of the data. In the case of EEG, these components can represent cortical neural activity, muscle signals, and, for example, interference from the surrounding electronics [12,24]. ICA has been commonly used to extract and remove extracranial artifacts from TMS–EEG data [14,22,10,11]. Additionally, ICA has been proposed to separate peripheral-evoked components from TEPs [3]. The components can be successfully identified provided that they are non-Gaussian and show statistical independence. These conditions are referred to as the ICA assumption [9]. It is of crucial importance to examine if the ICA assumption is met when using ICA. The goal of this work is to clarify and quantify in which scenarios ICA can be properly applied, and in which cases it fails, possibly leading to erroneous data analysis and misinterpretations of the final results.

Here, we investigate how the stimulus-induced nature of the artifacts affects the independence and the ICA performance. Earlier publications have argued that the ICA assumption of independence might not hold for TMS–EEG data sets [17,16,7]. Here, we illustrate how ICA succeeds in separating and removing the artifact components when different types of artifacts are present in TEPs. This set-up is obtained by systematically changing the statistical properties of the artifacts. To balance between realistic EEG properties and a controllable artifact scenario, we combine measured clean evoked EEG with simulated artifacts whose properties are precisely tuned. After artifact removal by ICA, the success of the removal can be measured by comparing the original known ground truth data and the data after the correction. We also examine whether increasing the number of trials can be used to improve the data cleaning accuracy in the cases when the artifact identification becomes increasingly challenging due to the broken ICA assumption.

In practice, a major limitation in the preprocessing of measured artifactual TEPs is the lack of any ground truth information to which comparisons of extracted components or cleaned data could be made. Multiple pipelines exist for TMS–EEG cleaning, which are known to yield different corrected data [2]. However, it is not possible to know which corrected data are closer to the true clean EEG. Therefore, we also pursue practical measures to quantify the correctness of any single estimated independent component. If the reliability of the ICA-based cleaning for a given artifact component can be estimated, it enables informed decision making about the component removal even in cases when the comparison with the ground truth data is not possible.

2. Methods

We first briefly give an overview on the ICA methodology. Then, we explain how the data measurement was performed and how the simulated artifact data were generated. These artifacts were mixed with the clean measured EEG, after which the data were cleaned with ICA. Finally, we measured how accurately the artifact removing uncovered the original artifact-free EEG.

2.1. Independent component analysis

ICA is a blind source separation technique used to separate statistically independent components from multidimensional data, such as artifacts and neuronal signals in EEG and MEG recordings [8,10,22].

The ICA assumption assumes the components to be statistically independent and non-Gaussian for correct IC detection. Statistical independence is defined so that the joint probability density function (pdf) $p(\mathbf{s})$ of N components $\mathbf{s} = [s_1, \ldots, s_N]^{\mathrm{T}}$, can be factorized into the product of the marginal pdf's of the components, *i.e.*, $p(\mathbf{s}) = p(s_1)\cdots p(s_N)$. Since the components are statistically independent, they are also uncorrelated, and their covariance matrix can be assumed to be diagonal.

ICA algorithms, such as FastICA and Infomax, show a certain degree of flexibility with respect to the ICA assumption; they are able to successfully find the independent components when the hidden components are 'close to' statistical independence and their degree of deviation from Gaussianity is sufficiently high. In practice, it is very difficult to ensure that the data fulfill the assumption of statistical independence.

The measured data can be modeled in the following way, where $\mathbf{Y}$ is the $M\times T$ data matrix, M being the number of channels and T the time instants:

$$\mathbf{Y} = \mathbf{AS}, \tag{1}$$

where $\mathbf{A}$ is the $M\times N$ mixing matrix, and $\mathbf{S}$ is the $N\times T$ time-course matrix of the components. The columns $\mathbf{A}(:,j)$ of the mixing matrix are the topographies of the hidden data generators (components). The row vector $\mathbf{S}(j,:)$ represents the time course (waveform) of component j. The column vectors $\mathbf{S}(:,t)$, $t = 1,\ldots,T$, are considered random samples of a random vector variable $\mathbf{s}$, with statistically independent components s_i, each with unit variance. Noise is not considered here for simplicity. The matrix $\mathbf{S}$ is normalized so that the covariance matrix,

$$\mathrm{Cov}(\mathbf{S}) \approx \frac{1}{T}\mathbf{S}\mathbf{S}^{\mathrm{T}} = \mathbf{I}, \tag{2}$$

equals the $N\times N$ identity matrix $\mathbf{I}$. The goal of ICA is to find the matrices $\mathbf{A}$ and $\mathbf{S}$.

As mentioned above, ICA is able to find the matrices when the components are statistically independent and non-Gaussian. The requirement of non-Gaussianity is needed because the joint probability density of signals with Gaussian variables is completely symmetric assuming that the mixing matrix is orthogonal. Therefore, the joint pdf of the signals contains no information on the directions of the columns of the mixing matrix $\mathbf{A}$, and the mixing matrix cannot be estimated for Gaussian variables.

FastICA measures independence by approximating the *non-Gaussianity* of the component [8]. According to the Central Limit Theorem, the distribution of a sum of independent random variables tends towards a Gaussian distribution. Thus the distribution of a sum of two variables is usually closer to a Gaussian, than the distributions of the single variables. This result is used so that the algorithm finds the independent components by maximizing the non-Gaussianity property.

FastICA approximates the non-Gaussianity by estimating the *negentropy* of each component. Negentropy $J_G(s_k)$ for component k is defined in the following way:

$$J_G(s_k) = \left[\frac{1}{T}\sum_{t=1}^{T} G(s_k(t)) - c\right]^2 = \left[\frac{1}{T}\sum_{t=1}^{T} G(\mathbf{w}_k^{\mathrm{T}}\mathbf{Y}_{\mathrm{w}}(:,t)) - c\right]^2, \tag{3}$$

where $G(\cdot)$ is the contrast function, $s_k(t)$ is the amplitude of component k at time index t, and c is the expectation of $G(u)$, u being a normalized Gaussian variable. For each component k, there is a separate weight vector (spatial filter) $\mathbf{w}_k$ whose entries FastICA optimizes, so that the maxima of Eq. (3) are found at $\mathbf{w}_k$, $k = 1,..,N$. $\mathbf{Y}_{\mathrm{w}}$ is a whitened data matrix. Whitening is a preprocessing step used in many blind source separation approaches to set the weight vectors orthogonal to each other.

The algorithm searches for the maxima of the negentropy function with the help of a fixed-point algorithm [9,8]. After finding the local maxima, the estimated whitened mixing matrix is constructed as $\widehat{\mathbf{A}}_{\mathrm{w}} = [\mathbf{w}_1, .., \mathbf{w}_N]$, and due to its orthogonality, the estimated waveform matrix is $\widehat{\mathbf{S}} = \widehat{\mathbf{A}}_{\mathrm{w}}^{\mathrm{T}}\mathbf{Y}_{\mathrm{w}}$. The estimated mixing matrix is then obtained in the following way since the independent components are uncorrelated:

$$\widehat{\mathbf{A}} = \mathbf{Y}\widehat{\mathbf{S}}^{\mathrm{T}}\frac{1}{T} = \mathbf{A}\frac{1}{T}\mathbf{S}\widehat{\mathbf{S}}^{\mathrm{T}} \approx \mathbf{AI} = \mathbf{A}. \tag{4}$$

In this work, we chose 'log(cosh(·))' as a contrast function $G(\cdot)$ in Eq. (3), and the symmetric mode of optimization, where all weight vectors are optimized simultaneously. Prior to ICA separation, we used principal component analysis (PCA) and compressed the data with the help of principal components. PCA finds the signal space directions (*i.e.*, topographies) where most of the data power lies, so it can be used to eliminate the data with poor signal-to-noise-ratio. The cutoff value of

the number of components was selected at a point where the decreasing eigenvalues had reached a relatively steady level, presumably representing noise. For the computations in this work, we used the FastICA package for MATLAB (https://research.ics.aalto.fi/ica/fastica/).

2.2. Data collection

The study participant was an adult right-handed female, with no history of neurological or psychiatric pathologies. She gave a written informed consent. All the experimental procedures were approved by the local ethics committee of the medical faculty of the University of Tübingen (protocol 716/2014BO2). The study was conducted in accordance with the Declaration of Helsinki.

The hand representation of left M1 was targeted. For the TMS neuronavigation, high-resolution 3D T1-weighted magnetic resonance images (3T Siemens Prisma with a 32-channel head coil, GRE pulse sequence, TE 2.22 ms, TR 2400 ms, FA 8°, FoV 256, Phase FoV 93.8%) were acquired. To deliver biphasic stimulation pulses of 260 μV in width, a TMS stimulator (PowerMAG Research 100, MAG & More, Munich, Germany) was used. The stimulator was connected to a passively cooled figure-of-eight coil (PMD70-pCool, 70-mm winding diameter, MAG & More, Munich, Germany). EEG signals were recorded at 5-kHz sampling rate with a TMS-compatible 128-channel device (NeuroOne, Bittium, Kuopio, Finland). Ag/AgCl-sintered ring electrodes preset in an elastic cap (EasyCap BC-TMS-128, EasyCap, Herrsching, Germany) according to the International 10–5 system were used for recording EEG. Electrode impedances were kept below 5 kΩ. We used a 24-bit biosignal amplifier for both EEG and electromyography (EMG) recordings (NeurOne Tesla with Digital Out Option, Bittium Biosignals Ltd, Finland). EMG responses were also recorded at the sampling rate of 5 kHz. Motor-evoked potentials (MEPs) were measured from the abductor pollicis brevis (APB) and first dorsal interosseous (FDI) muscles of the right hand in a bipolar belly–tendon montage using a 0.16-Hz – 1.25-kHz band-pass filter. The head position of the subject was maintained by a vacuum pillow (Vacuform, Salzbergen, Germany), and the TMS coil by a mechanical arm (Fisso, Baitella, Zürich, Switzerland). A stereoscopic neuronavigation system (Localite, St Augustin, Germany) was used to target and to maintain electric field produced by the coil in a consistent position with respect to the participant's head throughout the session. With the same system, the locations of the EEG electrodes were recorded.

The optimal coil placement to stimulate the motor hotspot was defined as the coil position and orientation producing the largest MEP amplitudes in the right FDI. The resting motor threshold (RMT) was determined as the minimum stimulation intensity eliciting MEPs with an amplitude exceeding 50 μV peak-to-peak in 50% of stimulation pulses [6], [23].

After preparing EEG and EMG, as well as pinpointing the EEG electrode locations, the optimal coil position was determined. The participant was seated and was instructed to fixate a visual target (fixation cross approximately 1 m in front of her). 1,000 single TMS pulses were delivered with an inter-stimulus interval of 2 s and a jitter of ± 0.25 s to avoid habituation effects. As a stimulation intensity, 110% of RMT was set. The measurement consisted of a single session lasting for about 3 hours.

2.3. Data preprocessing

EEG data were extracted from the time window ranging in the interval of [−1000 ms, 1000 ms] with respect to the stimulus. For estimating low-frequency drifts in these epochs without biasing effects from phase-locked evoked EEG, the data within the time window of [−4 ms, 400 ms] were removed and replaced with interpolated signals using Fourier-based interpolation (Matlab function 'interpft') for each channel and trial separately. Using these surrogate data, slow trends were then estimated within the [−1000 ms, 1000 ms] time windows by applying a Laplacian-based trend detection; see [15,7] for more detailed explanation. Finally, to eliminate the slow drifts, the estimated trend lines were subtracted from the original epoched EEG data.

To identify noisy channels and trials, we estimated the channel-wise uncorrelated noise signals using the DDWiener method [20]. The noise standard deviation in each channel and trial was computed to estimate the respective noise level. To avoid biased noise-level estimates, we normalized them by dividing by the noise-level estimates obtained using a spherical-head-model lead-field matrix as input data for DDWiener. The noise standard deviation median was computed excluding the values above the 98th percentile. If the noise level in a channel exceeded 10 times the median in over 10% of the trials, the channel was considered noisy and removed from further analysis. 10 channels were excluded. We also computed the noise range in each channel and trial, taking the maximum range across all channels in each trial to detect high-amplitude baseline changes in the signal: A trial was excluded if its maximum range exceeded two times the median of all maxima (computed excluding values above the 98th percentile). 199 trials were removed from the data set. After rejecting noisy channels and trials, the average reference was applied to the included EEG channels.

2.4. Simulations

The simulations were performed on MATLAB (The Mathworks Inc., Natick, MA, USA). The EEGlab toolbox [5] was used for visualizations. We first simulated data $\mathbf{Y}_a$ due to an artifact component, which were added to non-artifactual data $\mathbf{Y}_n$. The neural and artifactual data can be presented as a composition of N_n neural components and $N_a = 1$ artifact component as

$$\mathbf{Y}_a = \mathbf{a}_a \mathbf{s}_a^T$$

$$\mathbf{Y}_n = \mathbf{A}_n \mathbf{S}_n ,$$

where $\mathbf{a}_a$ is the $M \times 1$ artifact topography and $\mathbf{s}_a$ the $T \times 1$ artifact waveform. Similarly, $\mathbf{S}_n$ is the $N_n \times T$ neural waveform matrix, and $\mathbf{A}_n$ the respective $M \times N_n$ mixing matrix. After the addition of the artifact, the data becomes

$$\mathbf{Y}_{tot} = \mathbf{Y}_n + \mathbf{Y}_a .$$

We aimed to extract the artifact component by ICA from $\mathbf{Y}_{tot}$. This allowed us to reconstruct an estimate for $\mathbf{Y}_a$, after which data cleaning became a simple subtraction.

The neural data $\mathbf{Y}_n$ were obtained as TMS-evoked EEG signals in the interval of [100 ms, 300 ms] after the stimulus onset. Within this interval no stimulus-evoked spiky artifacts, including the muscle artifacts, were present. These data contained realistic TMS-evoked neural activity, spontaneous resting-state EEG, and measurement noise. The signal sources within these data were the hidden neural (and noise) components when running ICA. For the simulations, we simply shift the time-frame to [0 ms, 200 ms], and consider $\mathbf{Y}_n$ our clean ground truth EEG response.

To simulate artifact data, the topography $\mathbf{a}_a$ was first set. Artifactual TEP epochs within the time interval of [7 ms, 60 ms] were extracted. The data were first compressed into a subspace spanned by the 50 principal vectors representing the greatest eigenvalues of the data covariance matrix. Then, the compressed data were given as input for FastICA. Out of the 50 estimated components, one artifact component was then selected based on its waveform, which was a short-lived, 'sharp' time course, and topography, which was spatially non-smooth and resembled activity of extra-cranial origin. The L2-norm of the topography was set to 1. Here, it is not essential whether the artifact was retrieved fully accurately since this round of ICA was performed only for setting the simulated artifact topography.

The simulated data set consisted of 1,000 trials and 1,001 time instants in each trial. There were 116 channels, after the removal of bad channels, we estimated the data to contain 45 components in total based

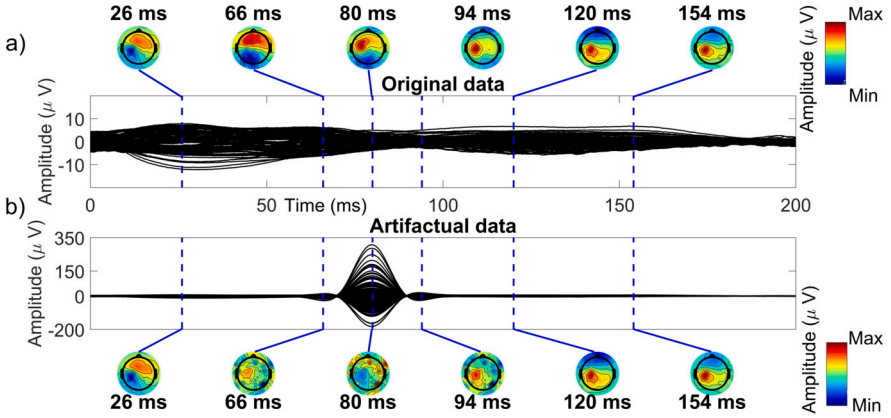

Fig. 1. a) The butterfly plot and the topographies of the original clean data, at time instants 26 ms, 66 ms, 80 ms, 94 ms, 120 ms and 154 ms. The color-maps of the topographies are scaled individually for each snapshot. b) Topographies of the total data are shown. The added artifactual data are simulated and added to the clean data as described in Section 2.4. At the peak latency of 80 ms, the neural data is obscured by the artifact. Note that the vertical axes are scaled differently in the panels a) and b).

on the rank of the data. To summarize, $N_{\mathrm{tr}} = 1000$, $T = 1001$, $M = 116$, and $N = 45$. We verified that the artifact topography was linearly independent of the topographies in the clean data by concatenating the first 45 principal vectors of the clean EEG with the artifact topography column-wise, after which the rank of the concatenated matrix was verified being equal to 46, *i.e.*, having increased by one.

The artifact waveform, $\mathbf{s}_{\mathrm{a}}^{\mathrm{T}} = [s_{\mathrm{a}}(1), \ldots, s_{\mathrm{a}}(T)]$, was set to follow an amplitude-modulated sinusoidal function $s_{\mathrm{a}}(t) = w(t) \cdot \sin(\omega t + \phi)$, where ω is the angular frequency, and ϕ the phase shift. A Gaussian window was used as a windowing function $w(t)$ to modulate the component latency and life-time, which yielded the waveform:

$$s_{\mathrm{a}}(t) = A \cdot \exp(-\frac{(t-t_0)^2}{\sigma^2}) \cdot \sin(\omega t + \phi), \tag{5}$$

where A sets the amplitude, σ is the standard deviation of the Gaussian window, and t_0 is the latency, *i.e.*, the moment of time when the artifact reaches its maximum amplitude. To maintain the shape of the waveform with varying latencies, we set $\phi = \frac{\pi}{2} - t_0 \frac{\pi}{100}$. Unless otherwise stated, we set here $A = 200$ μV, $\sigma = 10$ ms, $\omega = \frac{\pi}{100}$, and $t_0 = 80$ ms to produce a stereotypical waveform shape.

In Fig. 1 we see the difference in the topographies when the data contained no artifact, compared to the data containing an artifact that was simulated using Eq. (5). By changing the artifact waveform properties, *i.e.*, by randomizing the phase shift or latency of the signal in each trial withing chosen limits, we could specify how variable the artifact was. In the following Sections 2.4.1 and 2.4.2, we describe how the variables were systematically modulated to change the artifact variability. Thereafter, we applied ICA to study how well it uncovers artifacts with different variabilities.

We point out that the ICA algorithms also assume that the component waveform samples are non-Gaussian. This assumption is met since the amplitude distribution of the modulated sine wave does not follow a Gaussian distribution, which one can easily verify.

2.4.1. Varying the phase of artifact waveform

Firstly, the variation was created by adding a phase jitter which was randomized separately for each trial. We split the simulated sine wave into two parts: The first one, $s_{\mathrm{det}}(t)$, had a completely deterministic (constant) phase shift $\phi = \frac{\pi}{2} - t_0 \frac{\pi}{100}$ in Eq. (5). The second part, $s_{\mathrm{stoch}}(t)$, had a random phase shift, evenly distributed in the interval of $[0, 2\pi)$. Both sine waves, $s_{\mathrm{det}}(t)$ and $s_{\mathrm{stoch}}(t)$, were simulated using Eq. (5), but with different phase shifts. The proportions of random-phase and deterministic sine waves were systematically modified from completely deterministic to purely random-phase signal by as

$$s_{\mathrm{a}}(t) = (1-\alpha) \cdot s_{\mathrm{det}}(t) + \alpha \cdot s_{\mathrm{stoch}}(t), \tag{6}$$

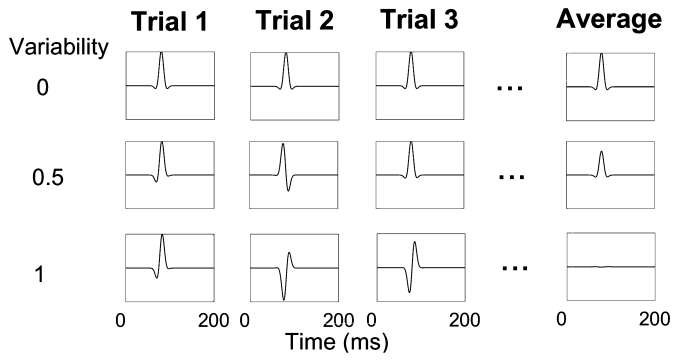

Fig. 2. Examples of artifact waveforms with different variabilities in terms of phase. Columns represent separate trials, and the right-most column shows the average of 1000 trials, only three of which are depicted for clarity. The phase variability increases towards the bottom row, meaning that the phase of the sinusoid becomes more random. The average of the trials approaches zero as the phase variability increases.

where the coefficient α ranged from zero to one. When α equals zero, the artifact waveform is completely deterministic. On the contrary, when α equals one, the artifact waveform is thought to be fully stochastic. We use the term *phase variability* to express the degree of randomness of the artifact phase. In other words, the phase variability of the signal increases as the value of α increases.

The effect of the phase variability on the artifact waveform to the averaged signal can be seen in Fig. 2. The average of the trials approaches zero as the phase variability of the signal increases. Thus large phase variability is not relevant in context of TMS artifacts, which tend to have large averaged amplitudes.

To perform systematic simulation, we set $\alpha = \{0, 0.1, \ldots, 1\}$. In addition to iteratively modifying the phase variability of the artifact waveform, we also modified the number of trials used for the simulation of the artifact. We used the values 50, 150, 300, 500, 750 and 1000 for the numbers of trials. In Section 3, we show how the ICA separation succeeds depending on the degree of the phase variability, and when the number of trials is diminished.

2.4.2. Varying the artifact latency

We also controlled the variability of the artifact waveform by randomizing the latency of the artifact within preset limits. This was obtained by controlling the width of the time window in which the artifact latency t_0 appears. We used Eq. (5) to simulate the artifact. In each trial, the time instant of the artifact peak (latency) was randomly chosen within the limits of the time window. The time window was consis-

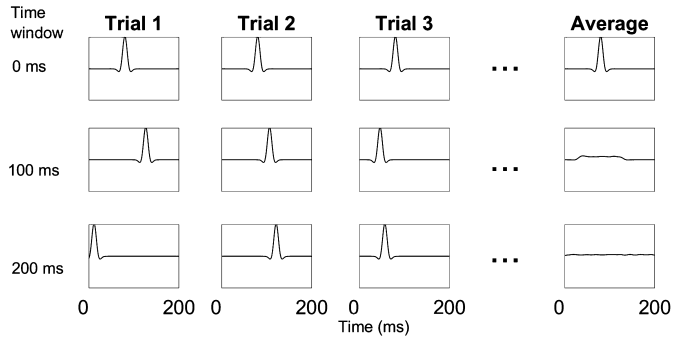

Fig. 3. Examples of artifact waveforms over trials (columns) with different latency variabilities. The latency is randomized in each trial separately within the time window indicated on the left, which widens towards the bottom row. The average of the trials, in the right-most column, decreases as the window width increases.

tently centered at 80 ms. The width of the time window was iteratively modified between 0 ms, meaning constant latency, and 200 ms, meaning completely random latency, *i.e.*, the artifact could occur at any point in time.

In the simulations, we used time window widths of 0, 0.4, 1, 1.6, 2.4, 3.2, 5, 10, 20, 40, 100, and 200 ms. When the value of the time window increased, so did the variability of the artifact waveform. Thus, we termed the time window width as the *latency variability*: 0 ms means identical waveform in every trial, *i.e.*, no latency variability and 200 ms corresponds to maximal latency variability and unpredictable timing of the peak artifact amplitude. In this case, the phase shift relative to the latency was always kept constant by choosing $\phi = \frac{\pi}{2} - t_0 \frac{\pi}{100}$ in (5). Artifact waveforms with different latency variabilities are demonstrated in Fig. 3: As the time window grows longer, the resulting average signal becomes increasingly flat.

2.5. Measuring the accuracy of ICA-based TEP cleaning

The artifact data $\mathbf{Y}_a$ were added to the recorded EEG data. After the addition, we aimed to determine and remove the artifact with the help of FastICA. Here, we knew that the artifact was represented by one single component. Thus, after estimating independent components, we looked for the one whose waveform best matched with that of the simulated artifact. This was obtained by calculating the correlation coefficients between the added artifact waveform and all the components returned by FastICA. The artifact component $\hat{\mathbf{s}}_a$, with the largest absolute correlation coefficient, was then removed from the data. The removal was implemented in the following way:

$$\hat{\mathbf{Y}} = \mathbf{Y}_{tot} - \hat{\mathbf{Y}}_a = \mathbf{Y}_{tot} - \hat{\mathbf{a}}_a \hat{\mathbf{s}}_a^T,$$

where $\hat{\mathbf{s}}_a$ is the estimated artifact waveform returned by FastICA, and $\hat{\mathbf{a}}_a$ the FastICA-estimated topography of the artifact.

If FastICA returns the exactly correct values for the estimated vectors $\hat{\mathbf{s}}_a$ and $\hat{\mathbf{a}}_a$, $\hat{\mathbf{Y}}$ coincides with the original clean data $\mathbf{Y}_n$, yielding perfect cleaning. In practice, there is always some error in the ICA results. To measure the estimation accuracy, after the removal of the artifactual component, the original, non-artifactual data were compared with the data cleaned with FastICA. We quantified the success of ICA removal by evaluating the difference between the original data and those cleaned with ICA, yielding $\mathbf{Y}_n - \hat{\mathbf{Y}}$. We measured *Relative Error* with the following equation:

$$\text{Relative Error} = \frac{||\mathbf{Y}_n - \hat{\mathbf{Y}}||_F}{||\mathbf{Y}_n||_F} \cdot 100\%, \tag{7}$$

where $||\mathbf{B}||_F$ denotes the Frobenius norm of matrix $\mathbf{B}$. The minimum value for Relative Error is zero, in case of perfect correction, and there is no maximum.

The data simulations were randomly generated 100 times for each chosen combination of parameters for the artifact waveform in Eqs. (5) and (6). The resulting artifactual data were cleaned, and Relative Error was then averaged over the 100 repetitions for the final results.

2.6. Measuring the artifact variability from estimated components

When simulating the artifact component, as described above, we can explicitly set the latency and phase variabilities of the artifact component. In fact, both features describe how predictable the component is as a function of time in any of the trials. As the latency or the phase variability increases, the component variance over trials also increases, and it becomes increasingly difficult to predict the exact amplitude of the component at a certain time after the stimulus.

This property can be quantified from the estimated components retrieved from ICA even though we do not know the original true components. Here, we give two options for assessing the trial-to-trial variability of an estimated component, whose amplitude at time index t and trial r is now denoted by $\hat{s}_r(t)$. Type 1 variability for an estimated component is defined as

$$\text{Esimated Variability Type 1} = \frac{\langle (\hat{s}_r(t) - \langle \hat{s}_r(t) \rangle_r)^2 \rangle_{r,t}}{\langle (\hat{s}_r(t))^2 \rangle_{r,t}}, \tag{8}$$

where $\langle \cdot \rangle_{\ldots}$ denotes the sample mean computed over variables listed in the subscript. Type 2 variability, on the other hand, is given by

$$\text{Esimated Variability Type 2} = 1 - \frac{\langle \langle \hat{s}_r(t) \rangle_r^2 \rangle_t}{\langle (\hat{s}_r(t))^2 \rangle_{r,t}}. \tag{9}$$

For both types of estimate, variability reaches its minimum at 0 when the component waveform is exactly the same over all trials, whereas the maximum, at 1, indicates that there is no repeatability over the trials, producing an averaged waveform of zero.

3. Results

3.1. Example cases of artifact elimination by ICA

We simulated different types of artifacts, as described in Section 2.4. Examples of the cleaning outcomes of the two types of artifacts, varying with respect to phase and latency, are depicted in Fig. 4. Panels a) and b) contain data that were cleaned from artifacts simulated according to Section 2.4.1, where the random phase proportion was determined by parameter α. Panels c) and d) show data cleaned from an artifact with variable peak-amplitude latency, described in Section 2.4.2.

According to the examples in Fig. 4, Relative Error tended higher with decreasing phase and latency variability, *i.e.*, decreasing α and the peak-artifact time window, respectively. The error caused by ICA cleaning clearly appears due to excessive data removal as shown by the attenuated EEG amplitudes after artifact subtraction. For example, in panel d), with Relative Error of about 60%, the ICA-corrected data almost gets to zero in amplitude within the time interval [78.4 ms, 81.2 ms], during which the artifact occurs.

The decrease of α seems to have similar effects on the amplitude of the data as compared to that of the shortening of the time window, as we can see by comparing panels a) and b): As the value of α decreases from 0.6 to 0.5, Relative Error grows from 15% to 35%. The peak artifact latency t_0 was 80 ms, which is also the time during which the amplitude of the ICA-corrected data is suppressed.

Based on the example cases, we set the threshold of 20% of Relative Error as an acceptable level of bias, which seems to correspond to reasonable cleaning outcomes. We use this threshold in the depictions, to describe results, make conclusions, and to give practical recommendations. Naturally, some other threshold could also be justified.

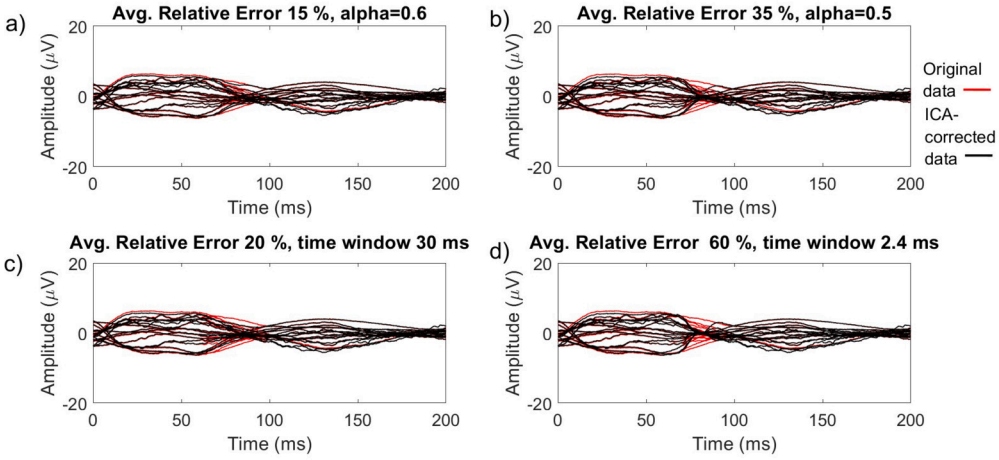

Fig. 4. Examples of cleaning outcomes of ICA-corrected data. Red lines depict the original clean EEG data, and the black ones depict the ICA-corrected EEG data. Relative Error from Eq. (7) is in the title above each examples case. In the case of perfect cleaning outcome, the black lines exactly match the red ones, giving Relative Error of 0%. Top row, a), b): Data cleaned from an artifact with variable phase over trials. The phase variability coefficient 'alpha' is given above each example illustration (see Section 2.4.1 for details). Bottom row, c), d): Data cleaned from an artifact with variable latency over trials. The peak artifact latency is varying over the time window whose width is given in the title of each example case (see Section 2.4.2 for details). Here, the phase and latency variabilities were adjusted such that illustrative examples were obtained, and the values do not exactly match with those used in the systematic simulations.

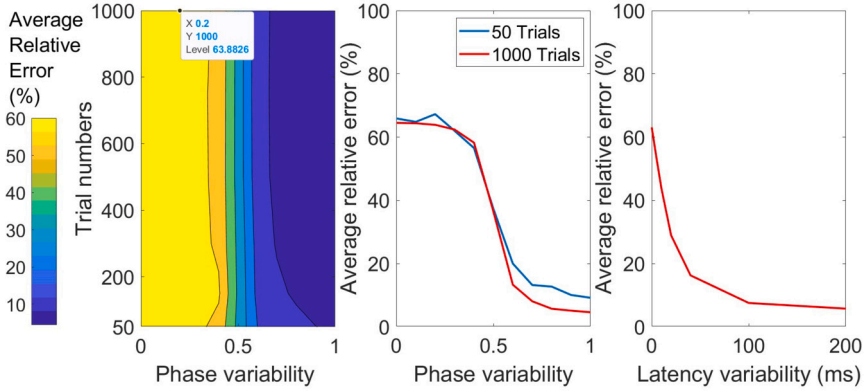

Fig. 5. The outcome of ICA-based EEG artifact cleaning depends on the variability of the artifact waveform over the trials. Here, a simulated artifact was superimposed on clean evoked responses, ICA-based data cleaning was then applied, and the accuracy of the cleaning was estimated with Relative Error given by (7). The phase variability and the latency variability of the artifact waveform over trials were systematically modified to study their effects on the cleaning accuracy. We also tested how the number of data samples affects the Relative Error. Left: Both the number of trials and the phase variability were changed in an iterative manner. As the phase variability increases, Relative Error clearly decreases at all numbers of trials. Increasing the number of trials slightly decreases Relative Error if the phase variability is higher than around 0.6. Middle: Relative Error as a function of phase variability extracted at the two extreme numbers of trials. Right: Relative Error decreases as the latency variability of the artifact increases.

3.2. Accuracy of data cleaning by systematic mapping of artifact properties

We used Relative Error to systematically measure the effectiveness of the artifact removal. Relative Error was calculated with Eq. (7) for the two types of artifact variations described in Sections 2.4.1 and 2.4.2, and the results are depicted in Fig. 5. In panel a), we tested removing the artifact with different proportions of phase variability (see Section 2.4.1) as tuned by α (Eq. (6)), as well as varying number of trials. The values of Relative Error stay above 60% until they rapidly decrease when the phase variability falls below 0.5. This type of an artifact requires the phase variability to be roughly 0.6 or larger in order to achieve Relative Error of less than 20%.

Panel b) corresponds to the data cleaning results as extracted from panel a) at 50 and 1,000 trials. It highlights the fact that increasing the number of trials from 50 to 1,000 does not help in improving the ICA accuracy when the phase variability is too small for effective component separation. In cases where there was sufficient phase variation, and ICA was successful even with the input of 50 trials, the addition of 950 trials decreased Relative Error by 5% (in absolute terms), bringing it down to around half of the original level.

Panel c) shows the data cleaning results when the artifact peaked within a preset time window, whose width (latency variability) was changed iteratively as explained in Section 2.4.2. Here, 1,000 trials were used. Artifacts with latency variability (Section 2.4.2) required the peak amplitude to vary within a time window of at least 40 ms to achieve Relative Errors of less than 20%, as seen in panel c). The values of Relative Error decreased rapidly as the latency variability increased. After the latency variability reached 100 ms, Relative Error stayed rather fixed at slightly below 10%.

Both types of artifact variations lead to similar results in terms of Relative Error. Both increasing the phase and latency variability resulted in decreasing Relative Error. In both artifact variations, FastICA was able to successfully find the artifact as long as the artifact component showed sufficient amount of trial-to-trial randomness.

To further validate the results, we repeated similar analysis of testing ICA outcome as a function of phase variability when using other artifact latencies and artifact offset phases. These analysis and results are presented in Supplementary Material, where we have also compared the ICA-estimated topography to the true simulated artifact topography. In all tested cases, Relative Error decreases in a as the artifact variabil-

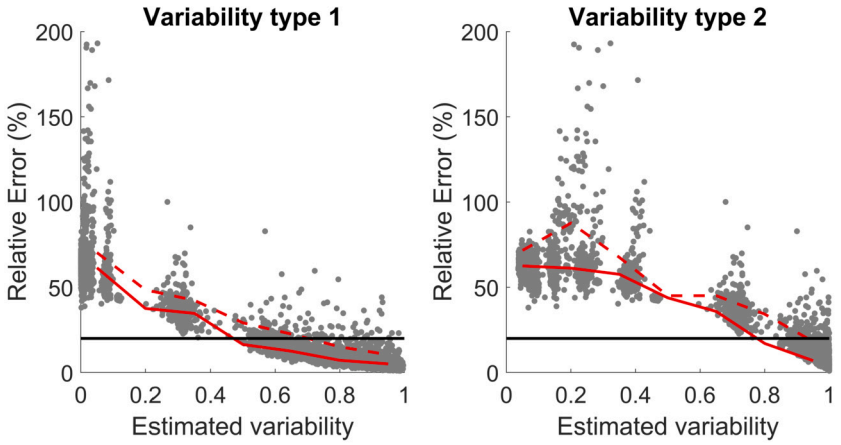

Fig. 6. Error of data cleaning as a function of estimated variability of the artifact waveform retrieved by ICA. Two types of variability estimates are illustrated, which we refer to as type 1 (left) and type 2 (right), defined by Eqs. (8), and (9), respectively. After each run of data cleaning, Relative Error was measured as a function variability estimate (gray dots). Median of Relative Error is shown as the red solid line, and the 95-percentile as the red dashed line. For both types of estimated variability, Relative Error decreased monotonically with increasing variability. The black line indicates Relative Error of 20%.

ity is increased. Furthermore, the error in the estimated topography decreases analogically.

3.3. Predictive value of the estimated artifact variability for the ICA accuracy

In each of the simulations described above, variability of the extracted artifact component was estimated using Eqs. (8) and (9). Thus, here we combined the data from Sections 2.4.1 and 2.4.2. Altogether, there were $100 \times 6 \times 7 + 100 \times 5 = 4700$ runs of ICA after simulations with different numbers of trials, and variabilities. For each estimated, artifact component from all data sets, we computed the variability estimates as well as Relative Errors. The outcome is depicted in Fig. 6, where single-run cleaning results can be seen as gray dots. The two types of estimated variabilities can be observed in the separate panels.

In addition to the single-run results, we computed the median and 0.95-percentile values for Relative Error as a function of Estimated variability. These statistics, in Fig. 6, were obtained using a sliding window of 0.3 (A.U.) in width within the horizontal axis, and with an overlap of 0.1 between each window frame. It shows that both types of variability estimates were clearly predictive of the resulting cleaning success, as measured by Relative Error (Eq. (7)). The larger the estimated variability got, the smaller became the resulting error of the cleaned EEG. There are some noteworthy differences between the variability estimation types: For type 1 estimate, Relative Error reached largest values for the very small variability estimates of less than 0.2. Additionally, the 0.95-percentile of Relative Error fell below 20% at the estimated variability of around 0.7. For type 2 variability estimate, high Relative Errors were obtained at larger variability estimates, and Relative Error also got very small ($\lesssim 0.1$) at higher estimated variability values as compared to type 1.

4. Discussion

4.1. Reliability of ICA-estimated components and artifact removal

By simulating differently behaving artifacts, we have demonstrated here that ICA-based cleaning of TMS-evoked EEG signals is reliable if the components show random behavior over trials in terms of their latencies and phases (Fig. 5). Importantly, phase- and time-locked components, are not accurately separable by ICA. These components have time courses which are reoccurring roughly with the same pattern and at same latency over multiple epochs. TMS-related EEG artifacts tend to have such a behavior, which is easily demonstrated by averaging the evoked EEG over trials. In the case of artifacts with randomly fluctuating phase, averaging effectively reduces the amplitude of the signal, while the amplitude remains roughly unmodulated if the signal is phase- and time-locked (Fig. 2).

When analyzing evoked responses, the ICA bias arises because both artifacts and neural signals are time-dependent. Thus, time axis can be considered as a so-called latent (hidden) variable, through which dependencies and correlations between components arise [16]. Due to common time-dependency, artifact and neural components can get combined into a mixed component, which includes all EEG-generating activity within the latency where the components activate. When eliminating this mixed component, we can easily remove a relevant part of the data within the time window of activity. This can be observed in the example cases presented in Fig. 4 where, due to combined ICA-estimated components, the cleaning also eliminates neural signals when the artifact variability is small.

For ICA to reliably uncover a component, the trial-to-trial behavior of the component waveform does not need to be completely random. However, it was clearly demonstrated here that when the artifact is increasingly time- and phase-locked (deterministic) with respect to the stimulus, the estimation accuracy clearly decreases. In terms of phase, the change in cleaning accuracy is clearly changing between poor and good when the component is 50% random and 50% fixed phase (Fig. 5). Regarding artifact latency, the cleaning error stayed low when the peak-amplitude latency occurred randomly within a time window of at least 40 ms, which was about the same as the life-time of the artifact (non-zero part of its waveform). It is noteworthy that time-locking alone does not hamper the ICA performance, nor does the phase locking alone. This was demonstrated in the results since properties were tested such that one of them was always fixed over trials, while the other one varied.

4.2. Practical implications for the ICA applied to TEPs

Broadly speaking, we may state that random-phase or random-latency artifacts can be effectively estimated from the TEP data. Such artifacts include randomly occurring eye blinks, (non-TMS-related) muscle twists, electrical activity of the heart (electrocardiogram signal), and 50-Hz noise (provided that its phase is not coupled with the TMS pulse. We note here that the artifact also needs to have a fixed topography. For example, if 50-Hz noise appears in several varying spatial patterns, several ICs are required to completely model this artifact.

Typically, TMS artifacts arise within a narrow time window and rather fixed phases, meaning small latency and phase variabilities. Thus, with TEP artifacts, we need to consider the cases where both the phase and variabilities are small. One must be aware of the risk of eliminating neural EEG data, when removing stimulation-elicited artifacts

from the TEPs. TEPs also contain artifacts with a notable amount randomness in the waveform shapes or timings within epochs. Such artifact components include, *e.g.*, eye blinks, and some external electromagnetic disturbances.

Artifacts that cannot be reliably estimated by ICA need to be eliminated by other means. For example, Signal Space Projection (SSP) requires very few prior assumptions [19]. Beamforming-based filtering techniques should be considered as an option, when there is a good sample size for estimating the covariance matrix [7].

Another important entity of artifacts are the neural responses to the auditory and somatosensory stimuli also delivered by TMS and recorded also by EEG. They could be expected to have more trial-to-trial variability compared to, for example, the muscle artifacts. However, they also include deterministic features as can be demonstrated by averaging over such peripheral-evoked responses. Quantification of the variability of the underlying sources would be needed to answer whether these uninteresting brain responses can be reliably separated by ICA.

4.2.1. Number of needed trials

According to the simulation results reported in Fig. 5, the number of trials does not seem to play as significant a role for ICA cleaning results as the variability of the artifact waveform: Relative Error with 50 trials was almost exactly matching the values obtained with 1000 trials when the variability of phase fell below 0.5. This means that increasing the number of trials does not compensate for the problem of components getting dependent when the trial-to-trial variability is too small.

On the other hand, when the variability is high enough for accurate ICA-based cleaning, increasing the number of trials is beneficial: increasing the number of trials from 50 to 200 results in Relative Error decreasing from 10% to around 5%. Using more than 200 trials does not seem to improve the cleaning accuracy.

4.3. Utility of measuring component variability

As discussed, based on our simulations, we found that 50% of randomness in the phase was sufficient for separating components by ICA. Thus, knowing the preset phase variability proportion α in Eq. (6) in a simulation setting, we could already estimate with some confidence if ICA-based cleaning is reliable.

Having such information available for measured data would be highly useful, but in practice, we are not able to measure the variability of the true components underlying measured TMS–EEG data. Therefore, here we introduced two variability estimates for estimated components to approximate the same information as the α coefficient or the peak-amplitude time window, which controlled the variability in the simulations.

According to the results (Fig. 6), both of these variability estimates correlated nicely with the Relative Error, which suggests that we can use such a measure to assess if the estimated component is reliably separated. In practice, the decision of eliminating an IC is done either automatically or by the researcher based on the component topography, time-course, and frequency spectrum. The estimated variability could be added as an informative parameter describing whether removal of the component provides accurate outcome or not. For example, if the median error of 20% was tolerated, the minimum estimated variability threshold could be set to 0.85 for Variability estimate type 1.

During TEP cleaning, deciding which artifacts to remove by ICA or by other means is known to be a critical task because the separate methods yield different corrected data [2]. In the future, we believe that variability estimation of the ICs will be an important addition to the preprocessing pipelines. It will inform the user about the expected accuracy of the data correction before eliminating a component, allowing for the analyzer to choose another cleaning strategy if necessary. Overall, this precaution potentially makes the data interpretations of the future studies more reliable.

4.4. Limitations

While in this study, we focused on the variability properties of the artifacts, we also need to keep in mind that there can be other critical features of components or the recorded data that are equally important for successful ICA-based separation and data cleaning. ICA is a statistical method, which needs a sufficient number of data samples for each estimated component. Moreover, to get a representative collection of samples, a sufficiently large set of independent samples is needed. Taken together, these requirements mean that if the artifact is very short-lived or its samples have high auto-correlation values (predominantly slow frequencies in the time-course), ICA may be prone to failure. Such aspects of the artifacts were out of the scope of this work. Moreover, we did not take into account that the artifact topographies may be relevant for ICA: If the artifact and the neural topographies appear similar, their separation can be difficult due to numerical problems. Here, we chose the topography such that it clearly had a different pattern as compared to the neural EEG. This decision was deliberately made to focus on component waveform properties, defining the independence aspect, and to keep the interpretation clear from confounding factors.

The chosen waveform shapes were rather simple, and one could argue that testing with more complex or diverse patterns could yield more realistic cleaning outcomes. We chose the waveform shape such that we could easily control the desired properties by tuning the function parameters. As a result, we were able to pinpoint the problems of the ICA-based cleaning very clearly. In fact, our results can be generalized over a wide selection of waveforms because the chosen waveform is actually a form of wavelet (Bartlett) whose weighted summations can be used to compose all discrete signals.

CRediT authorship contribution statement

Iiris Atti: Formal analysis, Visualization, Writing – original draft. **Paolo Belardinelli:** Data curation, Validation, Writing – original draft. **Risto J. Ilmoniemi:** Conceptualization, Funding acquisition, Supervision. **Johanna Metsomaa:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: R.J.I. is an advisor and a minority shareholder of Nexstim Plc. The other authors declare no competing interests.

Data availability

The code we used to simulate and clean the data, as well as to measure the variability of the estimated components, are available on GitHub (https://github.com/jmetsoma/ICA-accuracy-variability).

Acknowledgements

This work has been supported by the European Research Council (ERC Synergy) under the European Union's Horizon 2020 research and innovation programme (ConnectToBrain; grant agreement No 810377).

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.brs.2023.12.001.

References

[1] Baumer FM, Pfeifer K, Fogarty A, Pena-Solorzano D, Rolle CE, Wallace JL, et al. Cortical excitability, synaptic plasticity & cognition in benign epilepsy with centrotemporal spikes: a pilot tms-emg-eeg study. J Clin Neurophysiol, Off Publ Am Electroencephalogr Soc 2020;37(2):170.

[2] Bertazzoli G, Esposito R, Mutanen TP, Ferrari C, Ilmoniemi RJ, Miniussi C, et al. The impact of artifact removal approaches on tms–eeg signal. NeuroImage 2021;239:118272.

[3] Biabani M, Fornito A, Mutanen TP, Morrow J, Rogasch NC. Characterizing and minimizing the contribution of sensory inputs to tms-evoked potentials. Brain Stimul 2019;12(6):1537–52.

[4] Conde V, Tomasevic L, Akopian I, Stanek K, Saturnino GB, Thielscher A, et al. The non-transcranial tms-evoked potential is an inherent source of ambiguity in tms-eeg studies. NeuroImage 2019;185:300–12.

[5] Delorme A, Makeig S. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. J Neurosci Methods 2004;134(1):9–21.

[6] Groppa S, Oliviero A, Eisen A, Quartarone A, Cohen L, Mall V, et al. A practical guide to diagnostic transcranial magnetic stimulation: report of an ifcn committee. Clin Neurophysiol 2012;123(5):858–82.

[7] Hernandez-Pavon JC, Kugiumtzis D, Zrenner C, Kimiskidis VK, Metsomaa J. Removing artifacts from tms-evoked eeg: a methods review and a unifying theoretical framework. J Neurosci Methods 2022:109591.

[8] Hyvarinen A. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans Neural Netw 1999;10(3):626–34.

[9] Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. Neural Netw 2000;13(4–5):411–30.

[10] Korhonen RJ, Hernandez-Pavon JC, Metsomaa J, Mäki H, Ilmoniemi RJ, Sarvas J. Removal of large muscle artifacts from transcranial magnetic stimulation-evoked eeg by independent component analysis. Med Biol Eng Comput 2011;49:397–407.

[11] Lyzhko E, Hamid L, Makhortykh S, Moliadze V, Siniatchkin M. Comparison of three ica algorithms for ocular artifact removal from tms-eeg recordings. In: 2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC). IEEE; 2015. p. 1926–9.

[12] Makeig S, Debener S, Onton J, Delorme A. Mining event-related brain dynamics. Trends Cogn Sci 2004;8(5):204–10.

[13] Massimini M, Ferrarelli F, Huber R, Esser SK, Singh H, Tononi G. Breakdown of cortical effective connectivity during sleep. Science 2005;309(5744):2228–32.

[14] McMenamin BW, Shackman AJ, Maxwell JS, Bachhuber DR, Koppenhaver AM, Greischar LL, et al. Validation of ica-based myogenic artifact correction for scalp and source-localized eeg. NeuroImage 2010;49(3):2416–32.

[15] Metsomaa J, Belardinelli P, Ermolova M, Ziemann U, Zrenner C. Causal decoding of individual cortical excitability states. NeuroImage 2021;245:118652.

[16] Metsomaa J, Sarvas J, Ilmoniemi RJ. Multi-trial evoked eeg and independent component analysis. J Neurosci Methods 2014;228:15–26.

[17] Metsomaa J, Sarvas J, Ilmoniemi RJ. Blind source separation of event-related eeg/meg. IEEE Trans Biomed Eng 2016;64(9):2054–64.

[18] Mutanen T, Mäki H, Ilmoniemi RJ. The effect of stimulus parameters on tms–eeg muscle artifacts. Brain Stimul 2013;6(3):371–6.

[19] Mutanen TP, Kukkonen M, Nieminen JO, Stenroos M, Sarvas J, Ilmoniemi RJ. Recovering tms-evoked eeg responses masked by muscle artifacts. NeuroImage 2016;139:157–66.

[20] Mutanen TP, Metsomaa J, Liljander S, Ilmoniemi RJ. Automatic and robust noise suppression in eeg and meg: the sound algorithm. NeuroImage 2018;166:135–51.

[21] Rogasch NC, Thomson RH, Daskalakis ZJ, Fitzgerald PB. Short-latency artifacts associated with concurrent tms–eeg. Brain Stimul 2013;6(6):868–76.

[22] Rogasch NC, Thomson RH, Farzan F, Fitzgibbon BM, Bailey NW, Hernandez-Pavon JC, et al. Removing artefacts from tms-eeg recordings using independent component analysis: importance for assessing prefrontal and motor cortex network properties. NeuroImage 2014;101:425–39.

[23] Rossini PM, Barker A, Berardelli A, Caramia M, Caruso G, Cracco R, et al. Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an ifcn committee. Electroencephalogr Clin Neurophysiol 1994;91(2):79–92.

[24] Vigário RN. Extraction of ocular artefacts from eeg using independent component analysis. Electroencephalogr Clin Neurophysiol 1997;103(3):395–404.