NeuroImage 301 (2024) 120874

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

# Towards the definition of a standard in TMS-EEG data preprocessing

A. Brancaccio [a,*], D. Tabarelli [a], A. Zazio [b], G. Bertazzoli [b,1,2,3], J. Metsomaa [c,d], U. Ziemann [e,f], M. Bortoletto [b], P. Belardinelli [a,e,f]

[a] Center for Mind/Brain Sciences—CIMeC, University of Trento, I-38123 Trento, Italy
[b] Neurophysiology lab, IRCCS Instituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy
[c] Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland
[d] BioMag Laboratory, HUS Medical Imaging Center, Helsinki University Hospital, Helsinki University and Aalto University School of Science, Helsinki, Finland
[e] Department of Neurology & Stroke, University of Tübingen, Germany
[f] Hertie-Institute for Clinical Brain Research, University of Tübingen, Germany

ARTICLE INFO

*Keywords:*
TMS-EEG
Preprocessing
Reproducible research
Recommendations
Artist
TESA
SOUND/SSP-SIR

ABSTRACT

Combining Non-Invasive Brain Stimulation (NIBS) techniques with the recording of brain electrophysiological activity is an increasingly widespread approach in neuroscience. Particularly successful has been the simultaneous combination of Transcranial Magnetic Stimulation (TMS) and Electroencephalography (EEG). Unfortunately, the strong magnetic pulse required to effectively interact with brain activity inevitably induces artifacts in the concurrent EEG acquisition. Therefore, a careful but aggressive pre-processing is required to efficiently remove artifacts. Unfortunately, as already reported in the literature, different preprocessing approaches can introduce variability in the results. Here we aim at characterizing the three main TMS-EEG preprocessing pipelines currently available, namely ARTIST (Wu et al., 2018), TESA (Rogasch et al., 2017) and SOUND/SSP-SIR (Mutanen et al., 2018, 2016), providing an insight to researchers who need to choose between different approaches. Differently from previous works, we tested the pipelines using a synthetic TMS-EEG signal with a known *ground-truth* (the artifacts-free to-be-reconstructed signal). In this way, it was possible to assess the reliability of each pipeline precisely and quantitatively, providing a more robust reference for future research. In summary, we found that all pipelines performed well, but with differences in terms of the spatio-temporal precision of the ground-truth reconstruction. Crucially, the three pipelines impacted differently on the inter-trial variability, with ARTIST introducing inter-trial variability not already intrinsic to the ground-truth signal.

## 1. Introduction

Variability in the results of Non-Invasive Brain Stimulation (NIBS) experiments has been largely discussed in previous literature (e.g., Lopèz-Alonso et al., 2014; Ziemann and Siebner, 2015; Guerra et al., 2020). Inconsistency of the results has been ascribed to inter- and intra-individual factors. Regarding the former, one of the main elements contributing to inter-subject variability has been recognized in the individual structural and functional properties of the brain, which affect particularly the results of TMS-EEG studies. In this regard, classic metrics investigated in TMS-EEG studies, like TMS-evoked potentials (TEPs) amplitude and latency, have been found to depend on the individual structural properties of the brain investigated with Diffusion Tensor Imaging (DTI) techniques (e.g., Esposito et al., 2022; Bortoletto et al., 2021). One further source of variability has been ascribed to fluctuating brain-state dynamics, which have been found to modulate the impact of TMS on both electrophysiological and behavioral metrics. For example, it has been reported that both TEP amplitudes elicited by single-pulse TMS (spTMS) and the modulatory effects of repetitive TMS (rTMS) on TEP amplitudes depend on the cortical excitability state at the moment of stimulation (e.g., Zrenner et al., 2018; Baur et al., 2020; Desideri et al., 2019; Stefanou et al., 2019). Therefore, the aforementioned intra- and inter-subject variability factors should be controlled in order to reduce mixed results across studies.

* Corresponding author.
*E-mail address:* arianna.brancaccio@unitn.it (A. Brancaccio).
[1] Current affiliation: Neurology department, Brigham and Women's Hospital, Boston, MA, USA
[2] Current affiliation: Neurology department, Massachusetts General Hospital, Boston, MA, USA
[3] Current affiliation: Neurology department, Harvard Medical School, Boston, MA, USA

https://doi.org/10.1016/j.neuroimage.2024.120874
Received 13 May 2024; Received in revised form 25 September 2024; Accepted 26 September 2024
Available online 26 September 2024

Different techniques have been suggested in order to reduce inter- and intra-subject variability already at the TMS-EEG recording stage. For example, Casarotto et al. (2022) developed a toolbox for the real-time monitoring of the quality of recorded TEPs and consistency of artifactual components. On the other hand, other techniques can be used, such as complementing TMS-EEG recording with structural neuronavigation, as suggested by Lioumis and Rosanova (2022).

However, there is another residual remarkable source of variability not depending on intra- and inter-subject differences, which is due to the different preprocessing pipelines employed to clean TMS-EEG data from artifacts. As a matter of fact, TMS introduces both cortical responses as well as electromagnetic and physiological artifacts in EEG traces. The TMS pulse interacts electromagnetically with the conductive EEG electrodes and wires, exploiting their inductive and capacitive effects, and physiologically with the neuro-muscular system of the scalp. By magnetic induction, the magnetic field induces currents resulting in artefactual potentials with an amplitude that may saturate the amplifier's electronics. The strength of the magnetic gradient couples with the tiny electrical capacity of the interface between the electrodes and the skin, resulting in large ripples and decay artifacts in the recordings. The magnetic gradient interacts also with the scalp muscles, generating muscular artifacts. All this artifactual activity masks the genuine EEG response stemming from the neural processes. Therefore, it is impossible to look at genuine TMS-evoked activity (e.g., TEPs) without an accurate but aggressive preprocessing (Ilmoniemi and Kičić, 2010; Rogasch et al., 2014). For this reason, different preprocessing pipelines have been developed over the years in order to deal with artifacts introduced by TMS in the EEG signal. The most popular ones are based on Independent component analysis (ICA) and manual identification of artifacts as in TESA (Rogasch et al., 2017). Alternative proposed approaches include a fully automated ICA-based preprocessing pipeline (ARTIST, Wu et al., 2018), and a non-ICA based pipeline employing SOUND (Source-Estimate-Utilizing Noise-Discarding algorithm) and SSP–SIR (Signal-Space Projection–Source-Informed Reconstruction) (Mutanen et al., 2018, 2016) for TMS-related artifact correction.

Recent works have attempted at defining the impact of the preprocessing pipeline choice on the reconstructed TMS-EEG signal, and, in particular, on its variability. For example, Bertazzoli and colleagues (2021) have tested the impact of four different published pipelines (i.e., ARTIST, TESA, SOUND/SSP-SIR and TMSEEG) applied on the same TMS-EEG dataset. The results showed that the obtained preprocessed TEPs were significantly impacted by the chosen pipeline. Specifically, amplitude of TEPs and Global Mean Field Power (GMFP) varied across pipelines, and potential topographies over the scalp, mainly in the early responses, showed variable correlations ranging between poor and substantial. Moreover, the test-retest reliability of TEPs obtained in two separate sessions extensively varied across pipelines. These findings are supported by one further study applying a similar approach, highlighting how even small changes in the same cleaning pipeline can lead to different results in the reconstructed TMS-evoked activity, both in terms of amplitude and spatial topography (Rogasch et al., 2022). Taken together, these studies demonstrated that the methods used to clean the data highly influences the resulting TMS-EEG signal. However, they did not establish directly how accurate each pipeline is, i.e., how effectively the pipeline removes the unwanted artifacts while preserving the cortical responses evoked by the electromagnetic field. This is because they cross-compared the results of each pipeline in terms of TEPs and GMFP, without knowing *a-priori* the real signal that has to be recovered out of TMS induced artifacts.

More recently, some works addressed the efficacy of different methods in removing TMS-related artifacts by superimposing simulated TMS artifacts to a known ground-truth EEG signal. For example, Atti et al. (2024) tested the success of Independent Component Analysis (ICA) in removing various simulated TMS-EEG artifacts, while Mutanen et al. ((2024) compared ICA and SSP-SIR in removing muscle artifacts induced by the TMS pulse. The latter study reported that SSP-SIR performs better for cleaning artefacts when there is a substantial difference between the topography of the artefact and the topography of the signal of interest. Moreover, both studies highlighted that ICA-based cleaning is prone to error if the variability of an artefact is small. While these works have the advantage of introducing, at least partially, an *a-priori* known ground-truth signal, they focused on specific artifacts or analysis steps without considering the impact of the cleaning pipeline as a whole, which is what is commonly applied on TMS-EEG data.

All these considerations and results suggest, in principle, that each TMS-EEG experiment should be carefully designed, from data acquisition to preprocessing, according to the specific experimental hypothesis. While standardized pipelines are essential for ensuring consistency and reliability in TMS-EEG preprocessing, it is equally crucial to establish robust procedures for data acquisition that minimize experimenter dependency and ensure the recorded activity accurately reflects brain stimulation rather than artefacts. However, considering the research context and clinical practice, the aim of the published and standardized pipelines is usually to provide a reasonable standard approach including several preprocessing steps, each one differently influencing the others. For this reason, here we tested the performance of three TMS-EEG cleaning pipelines (ARTIST, TESA, SOUND/SSP-SIR) that well represent the state-of-the-art of preprocessing procedures, by combining the advantages of previous approaches. We did this by 1) using realistic TMS-EEG artifacts superimposed to a real ground-truth EEG signal and 2) testing the pipelines as a whole. In particular, to obtain a quantitative assessment of the level to which different preprocessing approaches affect TMS-EEG data, we compared the accuracy of each pipeline in cleaning the EEG signal from TMS-induced artifacts. To do this, we used a real sensory evoked potentials' EEG signal as ground-truth (therefore a known signal not containing any TMS artifacts), on which we superimposed typical and realistic TMS artifactual activity, extracted by means of ICA, from a real TMS-EEG dataset. Based on this approach, the original signal masked by the TMS artifacts is perfectly known. Hence, the reliability of the different pipelines in recovering the "ground-truth" can be optimally assessed using different metrics.

The primary goal of this study is to characterize the preprocessing accuracy of the ARTIST, TESA and SOUND/SSPIR pipelines in cleaning TMS artifacts from the EEG signal. We aim at understanding which is the overall best performing pipeline, but also at identifying in depth the most critical points raised by their use. This will provide valuable information for researchers performing TMS-EEG experiments, providing them also with additional information about how well the tested pipelines fit their data or, on the contrary, if the development of custom preprocessing code is needed.

We think that our work is a step towards defining a "gold standard" in cleaning contaminated TMS-EEG data, which is essential both in basic and clinical research. In basic research, this is a prerequisite for facilitating more comparable results across studies and research groups. In clinical research, where TMS-EEG has been increasingly applied to investigate the neurophysiological bases of psychiatric and neurological disorders, it is crucial for biomarker development by ensuring reproducible measures across centers (Julkunen et al., 2022; Tremblay et al., 2019).

## 2. Methods

We generated a test EEG signal for the evaluation of the three preprocessing pipelines (henceforth denoted as "test signal"). We superimposed an "artifactual signal", from a real TMS-EEG experiment with stimulation over the primary motor cortex (M1), to a "ground-truth" signal from one subject belonging to a dataset with peripheral stimulation generating sensory-evoked potentials (SEPs) in the EEG signals (Zazio et al., 2019). The test EEG signal has been then preprocessed separately using the three pipelines under evaluation: ARTIST, TESA and SOUND. After the complete preprocessing procedures, each of the three resulting cleaned EEG signals was compared with the ground-truth

one, in order to assess which preprocessing pipeline can better retrieve the ground-truth EEG data (see Fig. 1 for the methodological workflow). All the analyses were run using EEGLAB (Delorme and Makeig, 2004), FieldTrip (Oostenveld et al., 2011; version: 2022/02/06) and custom Python and MATLAB code (The Mathworks, Natick, MA, USA).

### 2.1. SEPs (ground-truth signal)

The rationale behind choosing the SEP signal as ground-truth data lies in the fact that SEPs reflect activations of neural structures along the somatosensory pathways, close to where the M1 TMS artefacts are mostly expected. We used EEG data (31 channels: Fp1, Fp2, C1, C2, CP3, F3, Fz, F4, CP4, FC5, FC1, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, PO7, PO8, O1, O2; 10–05 layout of the International EEG system; 5000 Hz sampling rate) from one subject belonging to a SEP dataset described in Zazio et al. (2019). This study was performed in accordance with the ethical standards of the Declaration of Helsinki and approved by the Ethics Committee of the IRCCS Instituto Centro San Giovanni di Dio Fatebenefratelli (Brescia, 19/2016). SEPs were assessed by means of 500 trials of electrical stimulation of the median nerve at the wrist of the left hand at 200% of individual perceptual threshold; stimulation frequency was set at 3.3 Hz. SEP data were epoched between −100 and 200 ms around stimulation. For further details about data collection and experimental setup, please refer to Zazio et al. (2019).

### 2.2. TMS-related artifactual signal generation

The "artifactual signal" was extracted from a real TMS-EEG dataset, with TMS over the left M1. Recruitment and experiments were conducted at the University of Tübingen, Germany, at the Department of Neurology and Stroke, and the Hertie Institute for Clinical Brain Research. All procedures were in accordance with the Declaration of Helsinki approved by the local ethics committee at the medical faculty of the University of Tübingen (810/2021BO2). All participants provided written informed consent to the experiment. Experiments were performed in accordance with the safety guidelines (Rossi et al., 2021). Participants were seated on a comfortable chair for the whole duration of the TMS-EEG experiment. 130-channels Ag/AgCl sintered ring electrode cap (EasyCap GmbH, Germany) was used for EEG recording. Electrodes were prepared by mild skin abrasion and filled by a conductive gel (Electrode Cream, GE Medical Systems, USA) until the desired impedance ($< 5$ kΩ) was attained. EEG and EMG were recorded simultaneously with a 24-bit biosignal amplifier (NeurOne Tesla with Digital Out Option, Bittium Biosignals Ltd., Finland) at a sampling rate of 5 kHz. Biphasic TMS pulses were delivered by a stimulator (MagPro R30, MagVenture, Denmark) connected to a 75 mm coil (MCF-B65, MagVenture, Denmark). The motor hotspot was identified as the coil position and orientation resulting in highest and most consistent motor evoked potential amplitudes in the contralateral first dorsal interosseus muscle (recorded by surface electrodes in a bipolar belly-tendon montage) and resting motor threshold was defined as the lowest stimulation intensity eliciting peak-to-peak motor evoked potential of $\geq$50 μV in at least 5 out of 10 trials (Rossini et al., 2015). Each participant received 800 TMS pulses at 115% of resting motor threshold with an interstimulus interval of 2.5 $s \pm 0.1$ s jitter.

We selected 4 healthy participants from this TMS-EEG study in order to extract a realistic artifactual signal, by means of an ICA approach. It has to be noticed that the EEG signal after the onset of TMS contains not only the artifactual EEG signal, but also the genuine brain potentials elicited by the stimulation: the TEPs. One of the goals of the preprocessing pipelines for TMS-EEG data is indeed to remove the artifacts induced by the pulse, while keeping the activity including the TEPs as intact as possible. For this reason, when extracting a "realistic artifactual signal" to be superimposed to the SEP ground-truth signal, we decided to use the data from 4 healthy participants rather than just one single subject. We did this because we can safely assume that the inter-subject variability in latency and signal shape of the TEPs across subjects is in general greater than the one of most of the TMS-induced artifacts (e.g., pulse and decay artifacts). For this reason, the ICA algorithm will be more efficient in separating the artifactual components from the TEPs when fed with data from different subjects. Moreover, when artifactual subject-dependent variability is expected, as for example for the muscular artifact component, using more than a single subject will result in a more generalized artifactual component.

In order to extract the TMS-related artifactual signal, we first epoched each of the 4 subjects' datasets between −100 and 200 ms around the TMS pulse. Bad trials were visually identified and discarded from each subject's data, resulting in 748, 725, 397 and 544 good trials, respectively, for the 4 subjects (total number of trials = 2414). Bad channels, for each subject, were also identified, removed and interpolated. Then, the artefact around the TMS pulse was cut out (- 2 to 5 ms around the pulse); this step is considered as necessary, since 1) removing the time around the artefact is a standard procedure across all preprocessing pipelines; 2) keeping the most relevant part of the pulse artifact would lead the ICA algorithm to be completely dominated by its

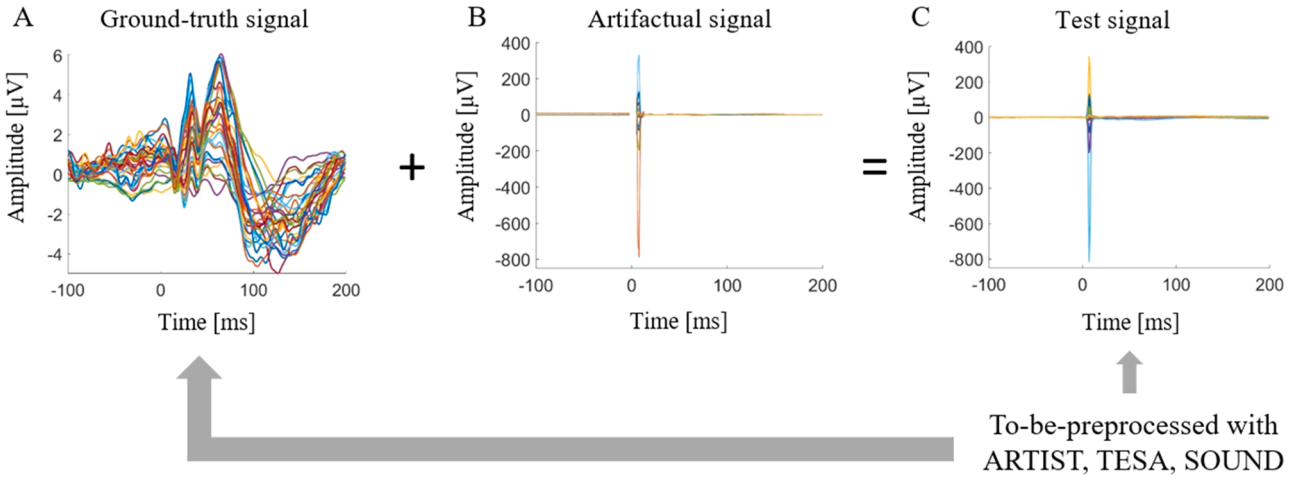


**Fig. 1.** Methodological workflow. A) The ground-truth clean signal from the SEP data; B) Artifactual signal generated from the 6 ICA components extracted from the TMS-EEG dataset and identified as artifactual; C) The superposition of the ground-truth signal and artifactual signal generates the test signal. The test signal was then preprocessed, separately, with the three preprocessing pipelines (ARTIST, TESA, SOUND) and each resulting cleaned EEG signal compared with the ground-truth signal to determine the most reliable pipeline.

very large amplitude (Hyvärinen et al., 2001). The TMS-EEG recordings of the 4 subjects were then concatenated and fed into a FASTICA algorithm (number of components: 30; epsilon: 1e-9; hyperbolic arc-tangent nonlinearity). ICA components extracted through FASTICA were visually inspected and six of them identified as "artifactual only", meaning they were carrying either only TMS-related artifactual signal or ocular artifacts (see Suppl. Fig. 1, for the components' waveforms and topographies). By projecting these six artifactual components back into sensor space, we obtained a full (130-channels; 2414 trials) EEG artifactual signal. Afterwards, only the 31 channels also present in the ground-truth signal were kept from the artifactual signal and only 500 epochs (same number of epochs present in the SEP signal) were randomly selected. At the end of this process, the artifactual signal (Fig. 1B) contains only TMS-related artifactual and ocular components and can be superimposed to the ground-truth signal for the generation of the "test-signal", as described in the next section.

### 2.3. Test-signal generation

Before combining the ground-truth signal with the artifactual signal to obtain our final "test-signal", we had to implement an aligning signal procedure in order to make 1) the SEP data comparable in amplitude with standard TEPs and 2) the ground-truth and the artifactual signals more homogeneous.

Therefore, the following adjustments were implemented on the ground-truth signal: first, the noise introduced by the median nerve stimulation between −1 and 3 ms around stimulation was removed and interpolated. Afterwards, a high-pass filter (0.1 Hz) was applied only for running FASTICA in order to discard ocular components from the ground-truth data. This was done because the ground-truth data would have to be later compared with the three preprocessed signals (ARTIST, TESA and SOUND) from which ocular components would have been discarded by preprocessing. However, as described in Section 2.2, one of the artifactual components extracted from the TMS-EEG data and superimposed with the clean signal already carries ocular movements. Keeping ocular artifacts from both the TMS-EEG and SEP dataset could have resulted in confounding the preprocessing pipelines. In this way, the goodness of the three tested preprocessing pipelines could be robustly tested not only in terms of how well each of it deals with TMS-related artifacts removal, but also with the removal of a realistic ocular-related signal. Furthermore, the SEPs in the ground-truth signal had been evoked over the right hemisphere via stimulation of the left hand and, hence, ground-truth data were flipped between hemispheres in order to have the SEP activity on in the same hemisphere as the TMS-related artifactual signal extracted from the dataset involving TMS over the left M1. Moreover, the comparison between the baseline activity of the ground-truth signal and the artifactual signal extracted from the TMS-EEG data revealed an offset between the two signals. For this reason, before overlapping the ground-truth signal with the artifactual signal, the offset was corrected with a trial-by-trial baseline correction implemented separately on the ground-truth and artifactual signal. This procedure homogenized the two signals. Finally, given that SEP deflections usually have a lower amplitude compared to TEPs, in order to make our ground-truth signal comparable with classic TMS evoked activity, SEPs amplitude was rescaled by multiplying it by a factor 2.

At this point, the ground-truth signal could be overlapped with the artifactual signal, generating the final test-signal for preprocessing with the three pipelines as described in the next section.

### 2.4. Preprocessing pipelines applied to test-signal

#### 2.4.1. ARTIST

The fully automated ARTIST pipeline, in principle, has the advantage to reduce individual variability due to the choices of the user by automatizing all preprocessing steps. For the preprocessing of the test signal, ARTIST discarded two bad channels (T7 and T8). Furthermore, it discarded 72 out of 500 trials as bad. Regarding the first round of ICA, which ARTIST uses to eliminate the TMS-related decay artifactual component, ARTIST found one component carrying this artifact, therefore eliminating one artifactual component in the first ICA round. In the second round of ICA, which aims at eliminating all other artifactual components except for the decay one, ARTIST discarded 7 artifactual components out of 25 total components extracted from the signal. A summary of bad channels, trials and components removed by the pipeline can be found in Table 1.

#### 2.4.2. TESA

The TESA semi-automated pipeline marked 3 channels for rejection. Furthermore, TESA marked 16 trials as artifactual epochs. As for the first ICA round, aimed solely at discarding component carrying TMS-related muscular artefact, TESA discarded 1 out of 28 components. The second ICA round found 1 bad component related to ocular artifacts and 7 other bad components related to electrodes noise out of 27 components, for a total of 8 bad components discarded from the signal in the second ICA round. A summary of bad channels, trials and components removed by the pipeline can be found in Table 1.

#### 2.4.3. SOUND/SSP-SIR

The SOUND/SSP-SIR is not an automated pipeline as it requires the user to mark bad channels (if the user chooses not to delegate this step completely to SOUND), bad trials, the ocular artifactual components extracted from the only round of ICA present in this pipeline and the independent components extracted with SSP-SIR carrying TMS-related muscle artefacts. In this application of the SOUND cleaning pipeline, no channels were marked for rejection by the user. Furthermore, 39 trials were marked as bad. Regarding the ICA round for identification of components carrying ocular artifacts, 2 out of 30 components were marked as ocular artifacts. Finally, the first 2 independent components identified by SSP-SIR were marked for rejection for carrying muscle-related artifacts. A summary of bad channels, trials and components removed by the pipeline can be found in Table 1.

### 2.5. Preprocessing quality evaluation

In order to evaluate the reconstruction quality of each preprocessing pipeline, we investigated 1) if the ground-truth and reconstructed signal statistically differ; 2) how the preprocessing impacts on the inter-trial variability; 3) to which extent the ground-truth and reconstructed signal linearly depend on each other.

We mainly focused our analysis on a list of channels of interests (P3, CP3, CP5, CP1, C3), where the most prominent effect of both the SEP and the TMS artifact is expected (Zrenner et al., 2022), and on three time-windows of interests (TOIs), defined on the basis of the key deflections observed in the ground-truth data (TOI 1: 22 – 42 ms; TOI 2: 52 – 82 ms; TOI 3: 100 - 150 ms). However, also results for the whole time-window of interest (5 – 200 ms) and for all channels are reported and discussed. All the subsequent analyses were performed using custom Python code, based on NumPy (Harris et al., 2020; version: 1.26.4) and SciPy (Virtanen et al., 2020; version: 1.14.0) and custom MATLAB code, partially based on FieldTrip (Oostenveld et al., 2011; version: 2022/02/06).

Before any subsequent analyses, both ground-truth and preprocessed

**Table 1**
Summary of trials, channels and components marked as bad and rejected by each preprocessing pipeline.

| | ARTIST | TESA | SOUND/SSP-SIR |
|---|---|---|---|
| Bad channels | FC5, T7, T8 | FP2, FC6, C4 | FC6, C4 |
| Number of bad trials | 29 | 16 | 30 |
| Number of bad components | 1st Round: 1<br>2nd Round: 7 | 1st Round: 8<br>2nd Round: 1 | Only Round: 4 |

EEG data were re-referenced to Cz and baseline corrected (baseline period from −100 to −15 ms). The three pipelines use the common average for re-referencing, even if at different steps (see Fig. 2). Thus, we further chose the central electrode Cz for a final homogeneous re-referencing, given also that a common average might potentially mask global effects.

We computed, for each pipeline and its own ground-truth, the mean across trials 1) separately for each channel, 2) pooling together the channels of interest (P3, CP3, CP5, CP1, C3) and 3) pooling together all channels. On each of these signals, we performed a non-parametric statistical comparison with the null-hypothesis of no differences between the reconstructed and original signal. We used a cluster-based permutation test (Maris and Oostenveld, 2007), exploiting only the time dimension as cluster forming feature for multiple comparison correction, because the low number of channels and their non-homogeneous distance does not allow for a robust spatial clustering. The test parameters were the following: 5000 permutations; cluster forming threshold = 0.05; cluster significance threshold = 0.05; two-tailed test with Bonferroni correction; time of interest = 5 – 200 ms. In addition to the signal averages, also the global mean field power, as in Esser et al. (2006), was computed for all preprocessed and ground-truth signals.

We evaluated the influence of the pipeline on inter-trial variability by defining, for each channel and computed averages, the following ratio:

$$\Delta(\mathrm{t}) = \frac{\sum_i^R [y_i(t) - \overline{y}(t)]^2}{\sum_i^R [x_i(t) - \overline{x}(t)]^2}$$

where $x_i(t)$ and $y_i(t)$ denote the time series of the trial $i$ of the ground-truth and preprocessed data, respectively, and $\overline{x}$ and $\overline{y}$ the correspondent averages across trials. The summation extends over all trials $R$. It can be noticed how this quantity corresponds to the ratio of the signal variance across trials, a widely used measure of inter-trial variability. We further averaged $\Delta(t)$ over time, using both the whole time-window of interest and each TOI. We expect $\Delta(t) \leq 1$ if the preprocessing does not change, or reduce, the inter-trial variability. On the contrary, values of $\Delta(t) > 1$ will indicate that the pipeline increased the original inter-trial variability.

Furthermore, to assess how much the linear dependence between the reference signal and the reconstructed one is preserved, we computed the temporal Pearson's correlation coefficient, for each channel and for the data averaged over selected and all channels, as follows:

$$\rho_c = \frac{\sum_i^N \left(x_{c,i} - \overline{x_c}\right)\left(y_{c,i} - \overline{y_c}\right)}{\sqrt{\sum_i^N \left(x_{c,i} - \overline{x_c}\right)^2}\sqrt{\sum_i^N \left(y_{c,i} - \overline{y_c}\right)^2}}$$

Here $x_{c,i}$ and $y_{c,i}$ refers to signal at time $i$ of the selected channel, or average, $c$ for the ground-truth and preprocessed data, respectively, with correspondent averages over time $\overline{x}_c$ and $\overline{y}_c$. The summation extends over all time points in the time of interest, namely $N$. It is worth to mention that the Pearson's correlation coefficient $\rho_c$ is a measure of the linear dependence between two signals not influenced by constant scaling or offsets. Thus, large values of $\rho_c$ would indicate that the shape of the original signal is well preserved in the reconstruction, even if scaling by a constant value or an offset are introduced.

Finally, for each TOI, we computed a Pearson's spatial correlation over time as follows:

$$\sigma(t) = \frac{\sum_c^C [x_c(t) - \overline{x}(t)][y_c(t) - \overline{y}(t)]}{\sqrt{\sum_c^C [x_c(t) - \overline{x}(t)]^2}\sqrt{\sum_c^C [y_c(t) - \overline{y}(t)]^2}}$$

where here $x_c$(t) and $y_c$(t) denote the time series of the channel $c$ at the time $t$ of the ground-truth and preprocessed data averaged across trials, respectively, and $\overline{x}(t)$ and $\overline{y}(t)$ denote the data further averaged across all channels $C$. The spatial correlation was then averaged over the whole time window of interest and for each TOI. The resulting values indicate how much the topographic structure is preserved by the preprocessing, apart from constant scaling or offsets.

## 3. Results

In Fig. 3, the EEG preprocessed and ground-truth signal averaged across all channels (panel A) and selected channels (panel C) are shown for the three pipelines. Moreover, panel B) shows the GMFP for the three preprocessed signals plotted against the correspondent GMFP of the

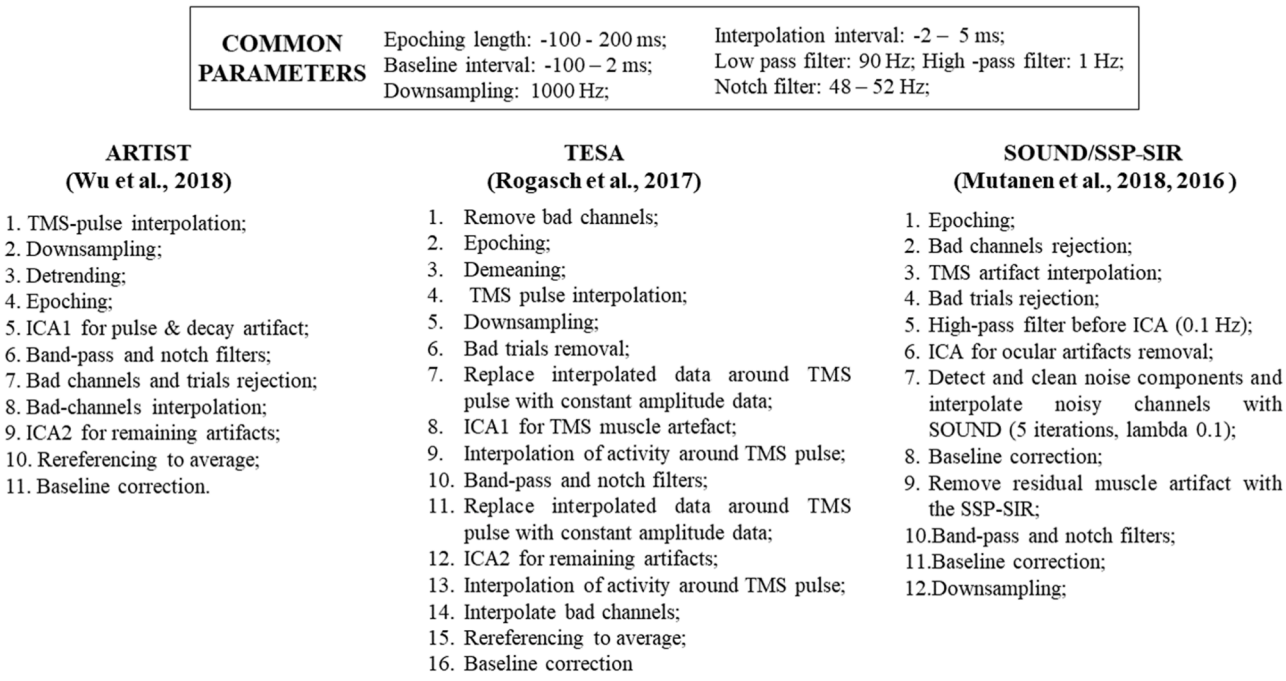


**Fig. 2.** Steps order in each preprocessing pipeline. Common parameters across pipelines are reported.

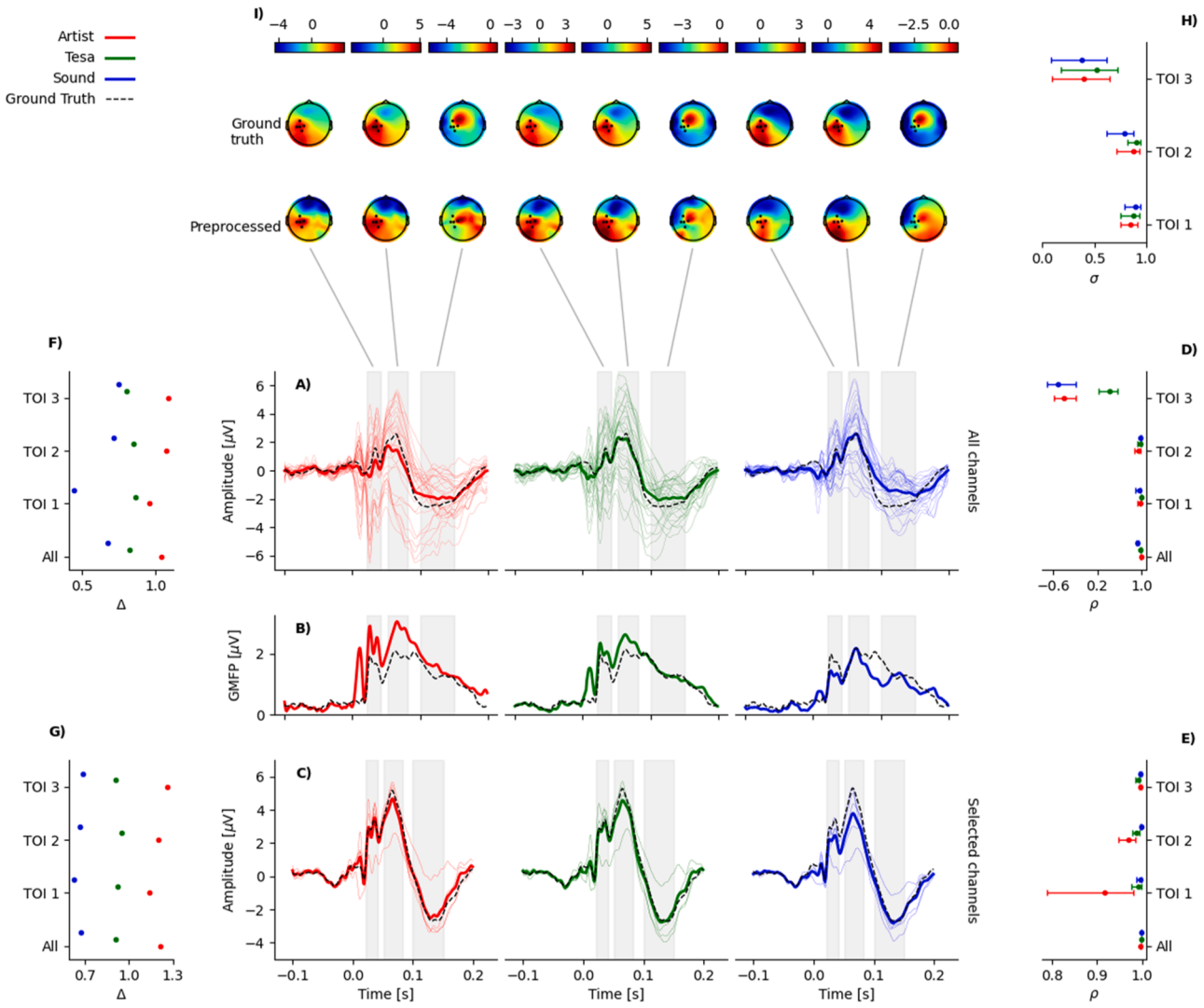


**Fig. 3.** Summary of pipeline quality evaluation results. The figure is conceptually organized in rows and columns. The first row refers to spatial information, while the last three rows refer to temporal information of all channels, global mean field power and selected channels, respectively. Columns refer to inter trial variability ratio (first column), average EEG activity (central column), Pearson's correlations (last column). Moreover, plot content is encoded in colors for each pipeline (red: ARTIST, green: TESA, blue: SOUND). EEG activity after preprocessing (colored solid lines) averaged over all channels (A) and channels of interests (C), is shown; channels of interests are P3, CP3, CP5, CP1, C3. Thick lines show single channel activity, while black dashed line show ground-truth activity. Gray shaded areas indicate time windows of interests (TOIs). B) GMFP computed from preprocessed (solid colored line) and original (dashed black line) data. Temporal correlations $\rho$ between reconstructed and original signal, averaged across TOIs and all (D) or selected (E) channels. Bars indicate 95 % confidence intervals. F) and G) show the ratio of inter-trial variability Δ, averaged across TOIs and all or selected channels, respectively. In I) topographic maps of activity averaged on each TOI is shown, for the ground-truth (top row) and for the preprocessed data (bottom row). Channels of interests are highlighted with black dots. The correspondent spatial correlations $\sigma$, for each TOI and pipeline, are shown in H), with 95 % confidence intervals.

ground-truth.

Even if differences can be visually observed in how the reconstructed signal follows the ground-truth, the statistical permutation test gave no significantly different time clusters, for all the pipelines and in both averaging modes. The high temporal correlations for the whole time-window, shown in insets D) and E) of Fig. 3 and reported in Table 2, confirm this result. From all values of $\rho$, further reported in Table 2, we can appreciate how the linear dependence between original and

**Table 2**
Values of temporal Pearson's correlation coefficients $\rho$ between the ground-truth and the preprocessed signal for the three pipelines and time windows of interest (TOIs), considering data averaged across all channels and channels of interest (P3, CP3, CP5, CP1, C3.). In brackets bootstrap confidence intervals at 95 % level.

| | | Temporal correlation $\rho$ | | | |
|---|---|---|---|---|---|
| | | All time | TOI 1 | TOI 2 | TOI 3 |
| All channels | **ARTIST** | 0.9856 [0.9827, 0.9881] | 0.9666 [0.9244, 0.9834] | 0.9424 [0.8697, 0.9756] | −0.4009 [−0.5767, −0.2059] |
| | **TESA** | 0.9796 [0.9712, 0.9848] | 0.9950 [0.9886, 0.9978] | 0.9676 [0.9276, 0.9843] | 0.4186 [0.2122, 0.5708] |
| | **SOUND** | 0.9208 [0.8999, 0.9371] | 0.9489 [0.8891, 0.9756] | 0.9830 [0.9709, 0.9892] | −0.5099 [−0.7037, −0.1693] |
| Selected channels | **ARTIST** | 0.9969 [0.9957, 0.9977] | 0.9175 [0.7921, 0.9807] | 0.9705 [0.9488, 0.9864] | 0.9965 [0.9938, 0.9979] |
| | **TESA** | 0.9982 [0.9978, 0.9985] | 0.9921 [0.9775, 0.9977] | 0.9882 [0.9781, 0.9938] | 0.9914 [0.9855, 0.9949] |
| | **SOUND** | 0.9985 [0.9977, 0.9990] | 0.9963 [0.9874, 0.9992] | 0.9978 [0.9962, 0.9985] | 0.9968 [0.9949, 0.9980] |

reconstructed signal is high for all the pipelines and TOIs, where all temporal correlations are >0.9, except for the third TOI, in the case of the average of all channels, where temporal correlations are < 0.5 and in some cases even negative.

Insets F) and G) in Fig. 3 show the values of inter-trial variability ratio Δ for the time of interests and for the three pipelines. All Δ values are also reported in Table 3. From these data it can be appreciated how the SOUND preprocessing always reduces the inter-trial variability of the considered sample, while ARTIST increases it. The TESA pipeline application, instead, does not affect much the inter-trial variability.

Finally, topographic maps of activity for each TOI and for the reconstructed and original signal are shown in Fig. 3 (panel I), together with the correspondent values of spatial correlations (Fig. 3, panel H). Values of spatial correlations $\sigma$ are further listed in Table 4, together with the 95 % confidence intervals. All the pipelines maintain quite well the spatial structure, in term of Pearson's correlations, in all TOIs except for the third one, where $\sigma$ values are close to or below 0.5 for all the three pipelines.

Results and statistical comparisons between ground-truth and reconstructed signal, computed at the single channel level, are further reported in Supplementary Fig 2.

The topographic plots of inter-trial variability ratio, for each pipeline and TOI, are shown in Fig. 4. It can be appreciated clearly that, while ARTIST systematically increases Δ on all channels, SOUND decreases it, and TESA keeps it quite stable, except for channel C1 where $\Delta = 2.15$. Insets in B) report the Δ averaged across channels and TOIs of interests, which confirm the same trend.

The topographic plots of temporal correlations, for each pipeline and TOI, are further shown in Fig. 5. For the sake of clarity, colormaps are low cut at $\rho \leq 0$. It can be noticed how all pipelines linearly correlate well with the original signal for all TOIs in the channels of interest (see Table 2 for values). This can be appreciated also in inset B), where the Pearson's temporal correlation is reported as averaged on the channels of interest. However, topographies also show how, for channels and TOIs far from the spatio-temporal point of stimulation, that the linear correlations between original and reconstructed signals decline.

## 4. Discussion

In this work we compared the accuracy of three representative pipelines for preprocessing of TMS-EEG datasets, namely ARTIST (Wu et al., 2018), TESA (Rogasch et al., 2017) and SOUND/SSP-SIR (Mutanen et al., 2018, 2016). Crucially and differently from previous works (e.g., Bertazzoli et al., 2021; Rogasch et al. 2022; Mutanen et al., 2022; Atti et al., 2024, 2024), we used a *ground-truth-recovery* approach, in which we synthesized a *to-be-cleaned* dataset starting from real EEG data and realistic TMS-EEG artifacts. Knowing exactly the signal to be recovered, we were able to feed the synthetic data to the pipelines and assess, both qualitatively and quantitatively, the degree of reliability in extracting the ground-truth EEG signal out of TMS induced artifacts. Moreover, we did that by considering the pipeline as a whole, as in a real experimental analysis, thus addressing the effects of the order and combination of the preprocessing steps all at once.

**Table 3**
Values of inter-trial variability ratio Δ for the pipelines and time windows of interest (TOIs), considering data averaged across all channels and channels of interest (P3, CP3, CP5, CP1, C).

| | | Inter-trial variability ratio **Δ** | | | |
|---|---|---|---|---|---|
| | | All time | TOI 1 | TOI 2 | TOI 3 |
| All channels | **ARTIST** | 1,04 | 0,96 | 1,07 | 1,09 |
| | **TESA** | 0,83 | 0,86 | 0,85 | 0,8 |
| | **SOUND** | 0,67 | 0,44 | 0,71 | 0,75 |
| Selected channels | **ARTIST** | 1,21 | 1,14 | 1,2 | 1,26 |
| | **TESA** | 0,9 | 0,92 | 0,95 | 0,9 |
| | **SOUND** | 0,67 | 0,62 | 0,66 | 0,68 |

**Table 4**
Values of spatial Pearson's correlation coefficients $\sigma$ between the ground-truth and the preprocessed signal of each pipeline averaged across each time window of interest (TOI). In brackets bootstrap confidence intervals at 95 % level.

| | Spatial correlation $\sigma$ | | | |
|---|---|---|---|---|
| | All time | TOI 1 | TOI 2 | TOI 3 |
| ARTIST | 0.65 [0.41, 0.81] | 0.85 [0.75, 0.92] | 0.87 [0.72, 0.93] | 0.40 [0.09, 0.64] |
| TESA | 0.66 [0.47, 0.81] | 0.87 [0.76, 0.93] | 0.90 [0.82, 0.95] | 0.52 [0.19, 0.74] |
| SOUND | 0.51 [0.28, 0.69] | 0.89 [0.80, 0.94] | 0.79 [0.62, 0.88] | 0.38 [0.09, 0.63] |

In general, all the pipelines demonstrated good performance in retrieving the ground-truth signal, as indicated by the absence of significant differences between each preprocessed and ground-truth signal as revealed by the permutation cluster statistics, on both all channels and selected channel averages. This result is confirmed by the high temporal correlation values observed across the entire time window of interest (5 – 200 ms post-TMS) on all channels and on channels of interests (See Fig. 3D-E, Fig. 5 and Table 2). All the preprocessing pipelines maintain robust linear correlation with their ground-truth, supporting their overall efficacy.

We also examined three distinct time windows of interest (TOIs), selected on the main deflections in the ground-truth data and a subset of channels expected to carry most of both the artifactual and signal of interest ('P3', 'CP3', 'CP5', 'CP1', and 'C3'). These analyses allowed to characterize the pipelines performance more in detail and revealed some crucial differences.

As it can be appreciated from temporal correlations in the different TOIs reported in Fig. 3D-E and Fig. 5, preprocessing robustness is high mostly for time points closest in time to stimulation (TOI 1). This holds true especially for TESA and SOUND/SSP-SIR, while ARTIST preprocessed signal shows slightly larger confidence intervals already in TOI 1 for the channels of interest, pointing to a less precise linear correlation. Nevertheless, all three preprocessed signals show high correlation values in all TOIs and particularly around the channels with the strongest effect of interest (see Fig. 5). Finally, the same trend can be observed in spatial correlations (Fig. 3H) where values of $\sigma$ dramatically decrease for all pipelines in the third TOI. This temporal and spatial congruence indicates that the preprocessing is in general particularly effective at preserving the neural effects close to the stimulation hotspot/time, which potentially carry the strongest TMS-evoked neural signal. On the contrary, researchers have to be particularly careful in interpreting preprocessed data far away from the site/time of stimulation, since results might strongly depend on the chosen pipeline.

As regards this preserved temporal congruence in early components, it has to be considered that, in our characterization analysis, we cut data and interpolated around the stimulus between −1 and 5 ms. In most studies, the TMS artifact is cut and interpolated for a longer period of time i.e., 5 to 15 ms across studies, even though the TMS artifact can be less than 5–6 ms with appropriate amplifier settings (Pavon et al., 2023; Veniero et al. 2009; Freche et al. 2018). This difference turns out in favor to the fitness of our procedure because it allows to evaluate the pipelines more effectively in correspondence of early components (5–70 ms). Assessing pipeline reliability on these early deflections is very important, since they are considered genuine responses of the brain (Belardinelli et al., 2019; Gordon et al., 2021) and believed to reflect fast activations from areas connected with the motor cortex (Bortoletto et al., 2021; Zazio et al., 2022). In particular, it can be noticed that SOUND/SSP-SIR and TESA performed slightly better than ARTIST in the accuracy of the signal reconstruction for the deflection investigated in TOI 1, especially when only channels of interests are considered. This evoked potential mostly resembles the TMS-related N45 deflection. In

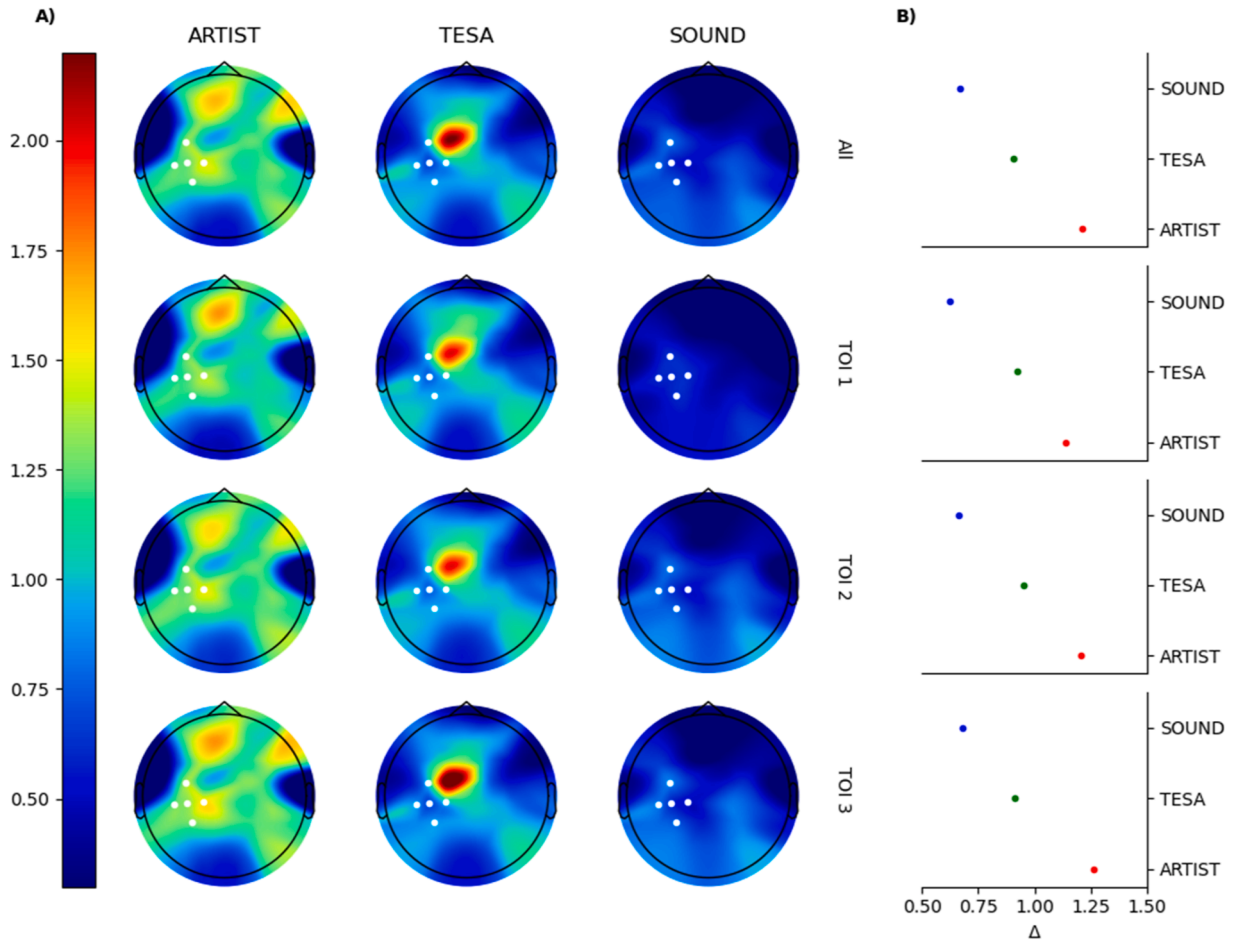


**Fig. 4.** A) Topographic plots of inter-trial variability ratio Δ for each pipeline and time window of interest (TOI). All topoplots are set on the same color scale for comparison. B) Inter-trial variability ratio averaged across qqqqqchannels of interest (P3, CP3, CP5, CP1, C) for each TOI and for the whole time. Results for each pipeline are encoded in color (red: ARTIST, green: TESA, blue: SOUND).

the last few years, this TEP has been demonstrated to be a metric of the Excitation/Inhibition balance in TMS-EEG experiments. In fact, positive allosteric modulators at the GABAAR like diazepam and alprazolam increase the N45 (Premoli et al., 2014; Gordon et al., 2023). Consistently, the NMDAR antagonist dextromethorphan also increases the N45 (Belardinelli et al., 2021). Furthermore, Darmani et al. (2016) found that the selective α5-GABAAR antagonist S44819 decreases the amplitude of the N45. Taken together, current evidence suggests that N45 could be a potential biomarker of the Excitation-Inhibition balance in TMS-EEG experiments. Therefore, it is crucial to reliably access its variations in amplitude in a precise and robust fashion and, thus, to characterize the reliability of the different TMS-EEG preprocessing pipelines in this time window of interest.

One further key point in the preprocessing quality evaluation is represented by the effect of the pipeline on the inter-trial variability. This analysis was carried out with the assumption that a preprocessing pipeline should never increase the inter-trial variability already intrinsic to the original signal. However, as it can be appreciated from Fig. 4 and Table 3, this is not always the case. We found ARTIST always increasing the original inter-trial variability of the ground-truth ($\Delta \geq 1$); crucially, this increase is even more pronounced when considering the channels closest to the stimulation site (see Table 3). Differently, TESA keeps inter-trial variability quite stable with respect to its ground-truth (Δ mostly close to 1), while SOUND/SSP-SIR even reduces it ($\Delta < 1$). In a nutshell, while TESA and SOUND/SSP-SIR do not introduce variability not already contained in the original ground-truth data, the ARTIST pipeline does. This spurious extra-variability introduced by ARTIST in the preprocessed signal is a crucial information for researchers interested in comparing experimental conditions from TMS-EEG dataset preprocessed with ARTIST. In fact, an increase in inter-trial variability might strongly impact especially the very common situation in which EEG data of different experimental conditions are compared by employing permutation-based statistics. In this case, the greater variability will have an impact on the statistical empirical distribution as computed by permutation, potentially masking genuine neural effects.

In interpreting these results, some limitations of our study have to be pointed out, especially about how the generation of the synthetic test signal might interfere with the performance evaluation of the pipelines, given their specific approaches for artifacts and noise identification and removal. As a matter of fact, we isolated TMS-EEG artifacts by means of an ICA decomposition, choosing this approach as the best possible compromise between time and accuracy for extracting realistic artifactual components. Indeed, it must be pointed out that a more precise but realistic generation of the synthetic signal is in principle possible by exploiting the physics of the involved processes. EEG clean ground-truth signal can be generated using an accurate forward model computed from high resolution magnetic resonance anatomical scans and simulating physiologically realistic EEG activity at sensor level (Neymotin et al., 2020). On the other hand, in principle, TMS-induced artifacts can be generated by considering the capacitive and inductive effects of electrodes and using a previously obtained forward model to re-create "tissue effects" on the neuromuscular sector. However, this would require a huge effort in exploiting the correct biophysical modeling that is out of the scope of the current work, but might be the topic of further

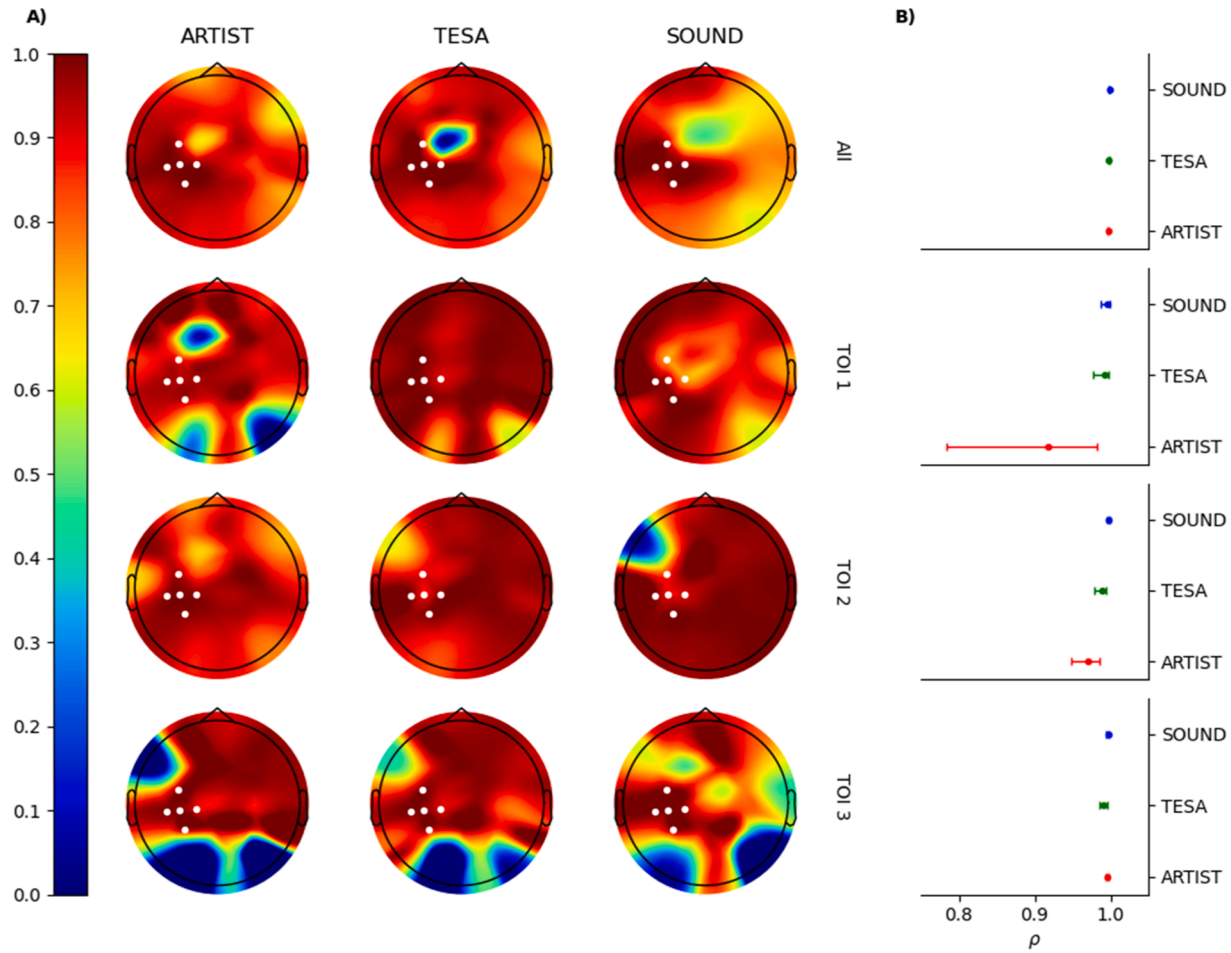


**Fig. 5.** A) Topographic plots of temporal correlations $\rho$ for each pipeline and time window of interest (TOI). All topoplots are set on the same color scale for comparison, where all values of $\rho \leq 0$ are encoded in dark blue, for clarity. B) temporal Pearson's correlation averaged across channels of interest (P3, CP3, CP5, CP1, C) for each TOI and for the whole time. Bars indicates 95 % confidence intervals. Results for each pipeline are encoded in color (red: ARTIST, green: TESA, blue: SOUND).

investigations.

In this context, it has to be considered that both ARTIST and TESA use the same ICA decomposition approach to remove TMS and other noise artifacts. On the contrary, SOUND/SSP-SIR is based on a mixture of recursive Inverse Modeling (SOUND) and Signal Source Projection (SSP). ICA algorithms (Hyvärinen et al., 2001), rely on creating a multivariate histogram (or distribution) of the channel data after collapsing the time dimension, and apply transformations to the multivariate channel dataset in order to find a "statistically independent" representation of the data. Statistical independence is then achieved by finding a transformation under which the multivariate histogram of transformed channels, namely the components, resemble as less as possible a product of normal distributions. In this sense, a single component histogram has already been transformed to be approximatively a maximally non-normal distribution and, in principle, this might favor ICA-based approaches like ARTIST and TESA. Despite this limitation introducing a potential negative bias for SOUND/SSP-SIR, this pipeline performed qualitatively and quantitatively very well, especially close to the site of stimulation and in the early components which are usually the most affected by the TMS-induced artifacts.

The different ways the pipelines deal with noise removal could also explain the differences found in the inter-trial variability ratio. As a matter of fact, ICA has been employed also for cleaning the ground-truth data, while SOUND/SSP-SIR exploits the power of methods based on Signal Source Projection and recursive Inverse Modeling. Therefore, the reduction in inter-trial variability for SOUND/SSP-SIR preprocessing could be ascribed to its different way of dealing with noise components, potentially cleaning noise still present in the ground-truth that was not further separated by standard ICA. While results obtained with TESA are, in principle, the expected ones, ARTIST, even if based on the same ICA decomposition approach, widely introduces variability that was not already intrinsic to the original signal, an adverse effect that must be taken into account, especially when efforts have been put to minimize variability at the recording stage, using different techniques, as mentioned in the Introduction.

A further limitation of our approach is that our pipelines did not process auditory artifacts due to the TMS click since the auditory component was already effectively masked in the original TMS-EEG dataset. However, given recently developed effective methods that mask this artifact at its origin (Russo et al., 2022), this should not be considered a major concern anymore. In fact, none of the here evaluated pipelines provides any specific correction for auditory artifacts. Moreover, our characterization design is optimal in evaluating pipeline performance for TMS-EEG of the primary motor cortex, as our ground-truth signal is located in that region.

Finally, we have to consider as a limitation that the way we generate the test signal, by linear superposition of ICA extracted noise components on a clean ground truth, might have led to a signal with a less complex noise structure than a real TMS-EEG dataset. As mentioned before, the synthesis of a more realistic TMS-EEG test dataset would require an extensive modeling work that is out of the scope of the present work. However, we think our results are valid and generalizable,

since in principle this limitation should have fostered the performance of ICA-based pipelines as ARTIST. But this was not the case as the ICA-based pipelines behaved worst. Conversely, ARTIST inter-performs, especially for what regards inter-trial variability.

To conclude, as far as it has been characterized in this work, all the preprocessing pipelines can be considered robust in reconstructing the original signal in the spatio-temporal proximity of the stimulation site, even if TESA and SOUND/SSP-SIR show slightly less variability than ARTIST in early components. On the other hand, for all the pipelines, results far from the time and site of stimulation have to be interpreted with caution. Crucially, the most significant result emerged from the analyses on the inter-trial variability, which clearly shows that TESA and SOUND/SSP-SIR must be preferred to ARTIST especially when the pre-processed data will be used in a scenario of permutation testing, given that the greater inter-trial variability introduced by ARTIST would potentially end up masking differences between experimental conditions. We think that the information provided in this paper can be valuable for researchers both in basic research and clinical settings. Moreover, we think our results establish a further step in the crucial task of defining a "gold standard" for TMS-EEG experiments.

## CRediT authorship contribution statement

**A. Brancaccio:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Methodology, Formal analysis, Conceptualization. **D. Tabarelli:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis. **A. Zazio:** Writing – review & editing, Investigation. **G. Bertazzoli:** Writing – review & editing. **J. Metsomaa:** Writing – review & editing, Methodology. **U. Ziemann:** Writing – review & editing, Investigation. **M. Bortoletto:** Writing – review & editing, Methodology, Investigation. **P. Belardinelli:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization.

## Declaration of competing interest

UZ received grants from the European Research Council (ERC), German Ministry of Education and Research (BMBF), German Research Foundation (DFG), and consulting fees from CorTec GmbH. All the other authors declare no competing interests.

## Data availability

Public sharing of raw data is not possible due to the data protection agreement with the participants. The data can be shared individually upon appropriate request to the corresponding author.

## Acknowledgments

MB has received funding from the Ministry of Health – Ricerca Corrente. UZ has received funding from the European Research Council (ERC Synergy) under the European Union's Horizon 2020 research and innovation programme (ConnectToBrain; grant agreement No 810377).

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2024.120874.

## References

Atti, I., Belardinelli, P., Ilmoniemi, R.J., Metsomaa, J., 2024. Measuring the accuracy of ICA-based artifact removal from TMS-evoked potentials. Brain Stimul. 17 (1), 10–18.

Baur, D., Galevska, D., Hussain, S., Cohen, L.G., Ziemann, U., Zrenner, C., 2020. Induction of LTD-like corticospinal plasticity by low-frequency rTMS depends on pre-stimulus phase of sensorimotor μ-rhythm. Brain Stimul. 13 (6), 1580–1587.

Belardinelli, P., Biabani, M., Blumberger, D.M., Bortoletto, M., Casarotto, S., David, O., Desideri, D., Etkin, A., Ferrarelli, F., Fitzgerald, P.B., Fornito, A., Gordon, P.C., Gosseries, O., Harquel, S., Julkunen, P., Keller, C.J., Kimiskidis, V.K., Lioumis, P., Miniussi, C., Rosanova, M., Rossi, S., Sarasso, S., Wu, W., Zrenner, C., Daskalakis, Z. J., Rogasch, N.C., Massimini, M., Ziemann, U., Ilmoniemi, R.J., 2019. Reproducibility in TMS–EEG studies: a call for data sharing, standard procedures and effective experi- mental control. Brain Stimul. https://doi.org/10.1016/j.brs.2019.01.010.

Belardinelli, P., König, F., Liang, C., Premoli, I., Desideri, D., Müller-Dahlhaus, F., Gordon, P.C., Zipser, C., Zrenner, C., Ziemann, U, 2021. TMS-EEG signatures of glutamatergic neurotransmission in human cortex. Sci. Rep. 11 (1), 8159.

Bertazzoli, G., Esposito, R., Mutanen, T.P., Ferrari, C., Ilmoniemi, R.J., Miniussi, C., Bortoletto, M., 2021. The impact of artifact removal approaches on TMS–EEG signal. Neuroimage 239, 118272.

Bortoletto, M., Bonzano, L., Zazio, A., Ferrari, C., Pedulla, L., Gasparotti, R., Miniussi, C., Bove, M., 2021. Asymmetric transcallosal conduction delay leads to finer bimanual coordination. Brain Stimul. 14 (2), 379–388.

Casarotto, S., Fecchio, M., Rosanova, M., Varone, G., D'Ambrosio, S., Sarasso, S., Pigorini, A., Russo, S., Comanducci, S., Ilmoniemi, R.J., Massimini, M, 2022. The rt-TEP tool: real-time visualization of TMS-Evoked Potentials to maximize cortical activation and minimize artifacts. J. Neurosci. Methods 370 (109486).

Darmani, G., Zipser, C.M., Böhmer, G.M., Deschet, K., Müller-Dahlhaus, F., Belardinelli, P., Schwab, M., Ziemann, U., 2016. Effects of the selective α5-GABAAR antagonist S44819 on excitability in the human brain: a TMS–EMG and TMS–EEG Phase I Study. J. Neurosci. 36 (49), 12312.

Delorme, A., Makeig, S., 2004. EEGLAB: an open-source toolbox for analysis of single-trial EEG dynamics. J. Neurosci. Methods 134, 9–21.

Desideri, D., Zrenner, C., Ziemann, U., Belardinelli, P., 2019. Phase of sensorimotor μ-oscillation modulates cortical responses to transcranial magnetic stimulation of the human motor cortex. J. Physiol. 597 (23), 5671–5686.

Esposito, R., Bortoletto, M., Zacà, D., Avesani, P., Miniussi, C., 2022. An integrated TMS-EEG and MRI approach to explore the interregional connectivity of the default mode network. Brain Struct. Function 227 (3), 1133–1144.

Esser, S.K., Huber, R., Massimini, M., Peterson, M.J., Ferrarelli, F., Tononi, G., 2006. A direct demonstration of cortical LTP in humans: a combined TMS/EEG study. Brain Res. Bull. 69 (1).

Freche, D., Naim-Feil, J., Peled, A., Levit-Binnun, N., Moses, E., 2018. A quantitative physical model of the TMS-induced discharge artifacts in EEG. PLoS. Comput. Biol. 14 (7), e1006177.

Gordon, P.C., Jovellar, D.B., Song, Y., Zrenner, C., Belardinelli, P., Siebner, H.R., Ziemann, U., 2021 Dec 15. Recording brain responses to TMS of primary motor cortex by EEG - utility of an optimized sham procedure. Neuroimage 245, 118708. https://doi.org/10.1016/j.neuroimage.2021.118708. Epub 2021 Nov 4. PMID: 34743050; PMCID: PMC8752966.

Gordon, P.C., Song, Y.F., Jovellar, D.B., Rostami, M., Belardinelli, P., Ziemann, U., 2023. Untangling TMS-EEG responses caused by TMS versus sensory input using optimized sham control and GABAergic challenge. J. Physiol. 601 (10), 1981–1998.

Guerra, A., López-Alonso, V., Cheeran, B., Suppa, A., 2020. Variability in non-invasive brain stimulation studies: reasons and results. Neurosci. Lett. 719, 133330.

Harris, C.R., Millman, K.J., van der Walt, S.J., et al., 2020. Array programming with NumPy. Nature 585, 357–362.

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent component analysis, adaptive and learning systems for signal processing, communications, and control. John Wiley & Sons, Inc, 1, 11–14.

Ilmoniemi, R.J., Kičić, D., 2010. Methodology for combined TMS and EEG. Brain Topogr. 22, 233–248.

Julkunen, P., Kimiskidis, V.K., Belardinelli, P., 2022. Bridging the gap: TMS-EEG from lab to clinic. J. Neurosci. Methods 369 (109482).

Lioumis, P., Rosanova, M., 2022. The role of neuronavigation in TMS-EEG studies: current applications and future perspectives. J. Neurosci. Methods 380 (109677).

López-Alonso, V., Cheeran, B., Río-Rodríguez, D., Fernández-del-Olmo, M., 2014. Inter-individual variability in response to non-invasive brain stimulation paradigms. Brain Stimul. 7 (3), 372–380.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164 (1), 177–190.

Mutanen, T.P., Kukkonen, M., Nieminen, J.O., Stenroos, M., Sarvas, J., Ilmoniemi, R.J., 2016. Recovering TMS-evoked EEG responses masked by muscle artifacts. Neuroimage 139, 157–166.

Mutanen, T.P., Metsomaa, J., Liljander, S., Ilmoniemi, R.J., 2018. Automatic and robust noise suppression in EEG and MEG: the SOUND algorithm. Neuroimage 166, 135–151.

Mutanen, T.P., Metsomaa, J., Makkonen, M., Varone, G., Marzetti, L., Ilmoniemi, R.J., 2022. Source-based artifact-rejection techniques for TMS–EEG. J. Neurosci. Methods 382, 109693.

Mutanen, T.P., Ilmoniemi, I., Atti, I., Metsomaa, J., & Ilmoniemi, R.J. A Simulation Study: comparative Analysis of ICA and SSP-SIR for rejecting TMS-Evoked-Muscle-Artifacts. Front. Hum. Neurosci., 18, 1324958.

Neymotin, S.A., Daniels, D.S., Caldwell, B., McDougal, R.A., Carnevale, N.T., Jas, M., Moore, C.I., Hines, M.L., Hämäläinen, M., Jones, S.R., 2020. Human neocortical neurosolver (HNN), a new software tool for interpreting the cellular and network origin of human MEG/EEG data. Elife 9. https://doi.org/10.7554/eLife.51214.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell. Neurosci. 2011, 1–9.

Hernandez-Pavon Julio, C., Domenica, Veniero, Til, Ole Bergmann, Paolo, Belardinelli, Marta, Bortoletto, Silvia, Casarotto, Casula, Elias P., Faranak, Farzan,

Matteo, Fecchio, Petro, Julkunen, Kallioniemi, Elisa, Pantelis, Lioumis, Metsomaa, Johanna, Carlo, Miniussi, Mutanen, Tuomas P., Lorenzo, Rocchi, Rogasch, Nigel C., Shafi, Mouhsin M., Siebner, Hartwig R., Thut, Gregor, Zrenner, Christoph, Ziemann, Ulf, Ilmoniemi, Risto J., 2023. TMS combined with EEG: recommendations and open issues for data collection and analysis. Brain Stimul. 16 (2), 567.

Premoli, I., Castellanos, N., Rivolta, D., Belardinelli, P., Bajo, R., Zipser, C., Espenhahn, S., Heidegger, T., Müller-Dahlhaus, F., Ziemann, U, 2014. TMS-EEG signatures of GABAergic neurotransmission in the human cortex. J. Neurosci. 34 (16), 5603–5612.

Rogasch, N.C., Biabani, M., Mutanen, T.P., 2022. Designing and comparing cleaning pipelines for TMS-EEG data: a theoretical overview and practical example. J. Neurosci. Methods 371, 109494.

Rogasch, N.C., Sullivan, C., Thomson, R.H., Rose, N.S., Bailey, N.W., Fitzgerald, P.B., Farzan, F., Hernandez-Pavon, J.C., 2017. Analysing concurrent transcranial magnetic stimulation and electroencephalographic data: a review and introduction to the open-source TESA software. Neuroimage 147, 934–951.

Rogasch, N.C., Thomson, R.H., Farzan, F., Fitzgibbon, B.M., Bailey, N.W., Hernandez-Pavon, J.C., Daskalakis, Z., Fitzgerald, P.B., 2014. Removing artefacts from TMS-EEG recordings using independent component analysis: importance for assessing prefrontal and motor cortex network properties. Neuroimage 101, 425–439.

Rossi, S., Antal, A., Bestmann, S., Bikson, M., Brewer, C., Brockmöller, J., Carpenter, L..L, Cincotta, M., Chen, R., Daskalakis, J.D., Di Lazzaro, V., Fox, M.D., George, M.S., Gilbert, D., Kimiskidis, V.K., Koch, G., Ilmoniemi, R.J., Lefaucheur, J.P., Leocani, L., Lisanby, S.H., Miniussi, C., Padberg, F., Pascual-Leone, A., Paulus, W., Peterchev, A. V., Quartarone, A., Rotenberg, A., Rothwell, J., Rossini, P.M., Santarnecchi, E., Shafi, M.M., Siebner, H.R., Ugawa, Y., Wassermann, E.M., Zangen, A., Ziemann, U., Hallett, M., basis of this article began with a Consensus Statement from the IFCN Workshop on "Present, Future of TMS: Safety, Ethical Guidelines", Siena, October 17-20, 2018, updating through April 2020, 2021. Safety and recommendations for TMS use in healthy subjects and patient populations, with updates on training, ethical and regulatory issues: expert Guidelines. Clin. Neurophysiol. 132 (1), 269–306.

Rossini, P.M., Burke, D., Chen, R., Cohen, L.G., Daskalakis, Z., Di Iorio, R., Di Lazzaro, V., Ferreri, F., Fitzgerald, P.B., George, M.S., Hallett, M., Lefaucheur, J.P., Langguth, B., Matsumoto, H., Miniussi, C., Nitsche, M.A., Pascual-Leone, A., Paulus, W., Rossi, S., Rothwell, J.C., Siebner, H.R., Ugawa, Y., Walsh, V., Ziemann, U., 2015. Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: basic principles and procedures for routine clinical and research application. An updated report from an I.F.C.N. Committee. Clin. Neurophysiol. 126 (6), 1071–1107.

Russo, S., Sarasso, S., Puglisi, S.G.E., Dal Palù, D., Pigorini, A., Casarotto, S., D'Ambrosio, S., Astolfi, A., Massimini, M., Rosanova, M., Fecchio, M, 2022. TAAC – TMS Adaptable Auditory Control: a universal tool to mask TMS clicks. J. Neurosci. Methods, 35101524.

Stefanou, M.I., Baur, D., Belardinelli, P., Bergmann, T.O., Blum, C., Gordon, P.C., Nieminen, J.O., Zrenner, B., Zrenner, C., 2019. Brain state-dependent brain stimulation with real-time electroencephalography-triggered transcranial magnetic stimulation. JoVE (J. Visualized Exp.) (150),, e59711.

Tremblay, S., Rogasch, N.S., Premoli, I., Blumberger, D.M., Casarotto, S., Chen, R., Di Lazzaro, V., Farzan, F., Ferrarelli, F., Fitzgerald, P.B., Hui, J., Ilmoniemi, R.J., Kimiskidis, V.K., Kugiumtzis, D., Lioumis, P., Pascual-Leone, A., Pellicciari, M.C., Rajji, T., Thut, G., Zomorrodi, R., Ziemann, U., Daskalakis, Z.J., 2019. Clinical utility and prospective of TMS–EEG. Clin. Neurophysiol. 150 (5), 802.

Veniero, D., Bortoletto, M., Miniussi, C., 2009. TMS-EEG co-registration: on TMS-induced artifact. Clin. Neurophysiol. 120 (7), 1392–1399.

Virtanen, P., et al., 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17 (3), 261–272.

Wu, W., Keller, C.J., Rogasch, N.C., Longwell, P., Shpigel, E., Rolle, C.E., Etkin, A., 2018. ARTIST: a fully automated artifact rejection algorithm for single-pulse TMS-EEG data. Hum. Brain Mapp. 39 (4), 1607–1625.

Zazio, A., Barchiesi, G., Ferrari, C., Marcantoni, E., Bortoletto, M., 2022. M1-P15 as a cortical marker for transcallosal inhibition: a preregistered TMS-EEG study. Front. Hum. Neurosci. 16, 937515.

Zazio, A., Guidali, G., Maddaluno, O., Miniussi, C., Bolognini, N., 2019. Hebbian associative plasticity in the visuo-tactile domain: a cross-modal paired associative stimulation protocol. Neuroimage 201, 116025.

Ziemann, U., Siebner, H.R., 2015. Inter-subject and inter-session variability of plasticity induction by non-invasive brain stimulation: boon or bane? Brain Stimulation: Basic, Translational, Clin. Res. Neuromodul. 8 (3), 662–663.

Zrenner, C., Belardinelli, P., Ermolova, M., Gordon, P.C., Stenroos, M., Zrenner, B., Ziemann, U., 2022. μ-rhythm phase from somatosensory but not motor cortex correlates with corticospinal excitability in EEG-triggered TMS. J. Neurosci. Methods 379, 109662.

Zrenner, C., Desideri, D., Belardinelli, P., Ziemann, U., 2018. Real-time EEG-defined excitability states determine efficacy of TMS-induced plasticity in human motor cortex. Brain Stimul. 11 (2), 374–389.