Contents lists available at ScienceDirect

NeuroImage

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

The impact of artifact removal approaches on TMS–EEG signal

Giacomo Bertazzoli [a,b], Romina Esposito [a], Tuomas P. Mutanen [c], Clarissa Ferrari [d],
Risto J. Ilmoniemi [c], Carlo Miniussi [a], Marta Bortoletto [b,*]

[a] Center for Mind/Brain Sciences - CIMeC, University of Trento, Rovereto, TN, Italy
[b] Neurophysiology lab, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Via Pilastroni 4, 25125 Brescia, Italy
[c] Department of Neuroscience and Biomedical Engineering, Aalto University, Espoo, Finland
[d] Service of Statistics, IRCCS Istituto Centro San Giovanni di Dio Fatebenefratelli, Brescia, Italy

ARTICLE INFO

ABSTRACT

Transcranial magnetic stimulation (TMS)-evoked potentials (TEPs) allow one to assess cortical excitability and effective connectivity in clinical and basic research. However, obtaining clean TEPs is challenging due to the various TMS-related artifacts that contaminate the electroencephalographic (EEG) signal when the TMS pulse is delivered. Different preprocessing approaches have been employed to remove the artifacts, but the degree of artifact reduction or signal distortion introduced in this phase of analysis is still unknown. Knowing and controlling this potential source of uncertainty will increase the inter-rater reliability of TEPs and improve the comparability between TMS–EEG studies. The goal of this study was to assess the variability in TEP waveforms due to of the use of different preprocessing pipelines. To accomplish this aim, we preprocessed the same TMS–EEG data with four different pipelines and compared the results. The dataset was obtained from 16 subjects in two identical recording sessions, each session consisting of both left dorsolateral prefrontal cortex and left inferior parietal lobule stimulation at 100% of the resting motor threshold. Considerable differences in TEP amplitudes and global mean field power (GMFP) were found between the preprocessing pipelines. Topographies of TEPs from the different pipelines were all highly correlated ($\rho > 0.8$) at latencies over 100 ms. By contrast, waveforms at latencies under 100 ms showed a variable level of correlation, with ρ ranging between 0.2 and 0.9. Moreover, the test–retest reliability of TEPs depended on the preprocessing pipeline. Taken together, these results take us to suggest that the choice of the preprocessing approach has a marked impact on the final TEP, and that further studies are needed to understand advantages and disadvantages of the different approaches.

1. Introduction

Over the last decade, transcranial magnetic stimulation (TMS) coupled with electroencephalography (EEG) has been an increasingly popular tool for investigating cortical excitability and connectivity. TMS-evoked potentials (TEPs), i.e., the cortical responses time-locked to the TMS pulse, provide information on the cortical excitability (Ilmoniemi and Kičić, 2010; Komssi and Kähkönen, 2006; Veniero et al., 2013) and effective connectivity, i.e., patterns of signal spread through the network in which the stimulated area is embedded (Bortoletto et al., 2015). TEPs can provide information about the status of the stimulated cortical location and about its causal relationships with connected areas, without the similar need for *a priori* assumptions as with functional magnetic resonance imaging (fMRI) or EEG alone. For these reasons, TEPs are used in clinical research to investigate neurophysiological alterations related to several psychiatric and neurological diseases (for review, see Tremblay et al., 2019), e.g., in disorders of con-

sciousness to distinguish between different clinical subtypes of patients (Rosanova et al., 2012) and in Alzheimer's disease as a marker of disease progression (Bagattini et al., 2019). Therefore, TEPs have been proposed as biomarkers that could improve diagnosis and monitor treatment-induced neuronal changes.

A crucial step for the development of a useful biomarker is demonstrating high reliability, usually equated with reproducibility, i.e., the degree to which a measurement is free from variable measurement error (Mokkink et al., 2010). High reliability allows one to reliably detect meaningful changes in the signals of interest. Reliability can be assessed in a population under stable conditions by taking the same measure with different instruments (internal consistency), by different raters (inter-rater reliability) and by the same rater over time (intra-rater reliability or test–retest reliability) (Beaulieu et al., 2017; Mokkink et al., 2010).

Up to now, only test–retest reliability of TEPs has been assessed in a few studies, which found overall high reliability and replicability for

TEP components (Casarotto et al., 2010; Lioumis et al., 2009), mainly at late latencies (Kerwin et al., 2018; Ozdemir et al., 2020).

However, these studies quantified the test–retest reliability using the same data-analysis pipeline, while reproducibility of TEPs may be affected by the preprocessing steps. Indeed, TMS–EEG signals are often heavily contaminated by various TMS-related artifacts, which can be an order of magnitude higher than the neuronal signal and time-locked with the TMS pulse, reducing the signal-to-noise ratio (SNR). To deal with this issue, several methodologies and algorithms have been developed to remove the artifacts while sparing the neuronal signal. On one side, the preprocessing phase allows the extraction of TEPs, but on the other side it could also affect their reproducibility. Specifically, the use of different approaches to reduce artefacts by different experimenters could affect the inter-rater reliability of TEPs. Moreover, the results of the test–retest reliability could vary across studies, due to the use of different approaches.

The goal of this study is to determine the amount of variability brought in the TEPs by the use of different preprocessing pipelines. To this end, we processed the same TMS–EEG dataset with four published pipelines: ARTIST (Wu et al., 2018), TMSEEG (Atluri et al., 2016), TESA (Rogasch et al., 2017), and SOUND/SSP–SIR (Mutanen et al., 2018, 2016), and compared the resulting TEPs. Although these pipelines share the common goal of removing artifacts while sparing the neuronal signal from TMS–EEG recordings, they employ different strategies. ARTIST is a fully automatic pipeline, developed to minimize the variability brought by the experimenter. Conversely, TMSEEG is operated via a fully manual graphical user interface (GUI) that guides the user through the steps needed to clean TMS–EEG signals. TESA, TMSEEG and ARTIST use two steps of independent component analysis (ICA) as a core function to isolate and remove artifacts, although employing different algorithms. SOUND-SSP–SIR employs ICA only for the removal of ocular artifacts, while its core functions are the source-estimate-utilizing noise-discarding algorithm (SOUND) (Mutanen et al., 2018) and signal-space projection–source-informed reconstruction (SSP–SIR) (Mutanen et al., 2016). Taken together, these processing differences may impact amplitude and topography of TEP components and ultimately their reproducibility.

2. Materials and methods

The dataset used in this study was acquired as a part of a larger TMS–EEG study to investigate brain connectivity. The study followed the Helsinki Declaration guidelines and was approved by the Human Research Ethical Committee of the University of Trento (protocol number 2017–014). BIDS-formatted data (Pernet et al., 2019) and code from this study are available on gin.g-node.org/CIMeC/TMS-EEG_brain_connectivity_BIDS.

2.1. Participants

Sixteen healthy volunteers (age 24.5 ± 2.8, 7 females) were enrolled in the experiment. Each participant gave a written informed consent for participating in this study and was screened for MRI and TMS compatibility (Rossi et al., 2009; Sammet, 2016). Each subject underwent three experimental sessions: in the first session, MRI and fMRI were acquired to obtain individualized structural and functional localization of the default-mode network (DMN); the other two identical sessions (test–retest) were 72.3 ± 35.8 days apart and involved TMS–EEG coregistration during DMN stimulation.

2.2. MRI data acquisition

High-resolution anatomical images were acquired with two T_1-weighted anatomical scans (MP-RAGE; $1 \times 1 \times 1$ mm^3; FOV, 256,224 mm^2; 176 slices; GRAPPA acquisition with an acceleration factor of 2; TR, 2700/2500 ms; TE, 4.18/3.37 ms; inversion time (TI),

1020/1200 ms; 7°/12° flip angle), through a 4T Siemens MedSpec Synco MR scanner and a birdcage transmit 8-channel receive head radiofrequency coil. Single shot T_2*-weighted gradient-recalled echo-planar imaging (EPI) sequence was used to acquire functional images. A 30-slice protocol was used, acquiring each slice in ascending inter-leaved order, within a repletion time (TR) of 2000 ms (voxel resolution, $3 \times 3 \times 3$ mm^3; echo time (TE), 28 ms; flip angle (FA), 73°; field of view (FOV), 192×192 mm^2). Each run consisted of 200 volumes.

2.3. TMS protocol and targeting

Subjects were seated in a comfortable armchair in a dimly illuminated room, while fixating a fixation cross in front of them and wearing earplugs.

Single-pulse TMS was delivered through a Magstim Rapid magnetic stimulator (Magstim, Whitland, UK) and a 70-mm figure-of-eight coil. In each session, the coil location corresponding to the hotspot for the right first dorsal interosseous muscle was identified and the resting motor threshold (rMT) was measured as the lowest intensity producing motor-evoked potentials of over 50 μV in a minimum of 5 out of 10 trials in the electromyogram recorded with bipolar montage (Rossini et al., 2015). Thereafter, the stimulator intensity was set at 100% of the rMT. Four nodes of the DMN extracted from functional maps (see supplementary materials for details) were stimulated: left and right dorsolateral pre-frontal cortex (DLPFC) and left and right inferior parietal lobule (IPL). The Talairach coordinates of the point of local maxima in each node and the structural MRI images were input to the SofTaxic Neuronavigation system (E.M.S., Bologna, Italy) to guide an accurate and consistent positioning of the TMS coil. The order of stimulation target was randomized across participants. On each site, 120 biphasic TMS pulses were delivered with an inter-stimulus interval of 2–10 s. The coil was placed by the experimenter in a position tangential to the scalp, with an angle of approximately 45° between the interhemispheric sulcus and the coil handle, with the handle pointing backwards, for the DLPFC and 10° for the IPL.

For the purposes of this study, we limited our analyses to the left DLPFC and the left IPL, from both the test (T1) and retest session (T2).

2.4. EEG data acquisition

EEG signals were acquired with a BrainAmp DC TMS-compatible system (BrainProducts GmbH, Germany), with a continuous recording from BrainCap 62 TMS-compatible Ag/AgCl-coated Multitrode electrodes (BrainProducts GmbH, Germany) positioned on the scalp according to the 10/10 International System. One electrode was placed to the left-eye temporal canthus to detect horizontal eye movements, and another electrode was placed beneath the left eye to detect vertical eye movements and blinks. TP9 was used as online reference. The ground electrode was placed at FPz. Signals from all channels were band-pass filtered at 0.1–1000 Hz and digitized at the sampling rate of 5 kHz. The skin–electrode impedance was kept below 5 kΩ. Trigger marks at the TMS pulse delivery were sent to the EEG system using a customized MATLAB script (MathWorks®, Inc., Massachusetts, USA).

2.5. EEG signal preprocessing

Preprocessing pipelines

Next, we give a brief introduction to the preprocessing pipelines employed in this study. For a detailed description, please refer to the respective original papers, i.e., ARTIST (Wu et al., 2018), TMSEEG (Atluri et al., 2016), TESA (Rogasch et al., 2017), SOUND/SSP–SIR (Mutanen et al., 2018, 2016).

2.5.1. TMSEEG v4.0

The TMSEEG (Atluri et al., 2016) pipeline is run via a GUI from which the user can call, in a fixed order, all the functions needed to

Table 1

Steps order in each processing pipeline. Exact values for each parameter are described in the "Preprocessing parameters and steps" paragraph.

	ARTIST	TMSEEG (4.0)	TESA (v1.0.1)	SOUND-SSP–SIR
1	EOGs removal	Epoching	EOGs removal	EOGs removal
2	TMS-pulse interpolation	Demeaning [#]	Channel rejection [+]	High-pass filter
3	Downsampling)	EOGs removal	Epoching	Epoching
4	Detrending	TMS-pulse removal[*]	Demeaning [#]	TMS-pulse interpolation
5	Epoching	Channels and trials rejection -1	TMS-pulse interpolation	Baseline correction
6	**ICA -1**	**ICA -1**	Downsampling	**SOUND**
7	Bandpass – Notch filters	Bandpass – Notch filters	Trials rejection [+]	Trials rejection
8	Channels and trials rejection [+]	**ICA -2**	ICA -1 [§]	ICA [@]
9	Bad-channels interpolation	Channels and trials rejection - 2	Bandpass – Notch filters	**SSP–SIR** [§]
10	**ICA -2**	Rereferencing	ICA -2 [§]	Bandpass – Notch filters
11	Rereferencing	Bad-channels interpolation	Bad-channels interpolation	Rereferencing
12	Baseline correction	TMS-pulse interpolation[*]	Rereferencing	Downsampling
13		Baseline correction [×]	Baseline correction	Baseline correction
14		Downsampling [×]		

[*] TMS-pulse artifact was removed, but not interpolated. Instead, the signal was cut and pasted. In the end of the pipeline, the removed TMS-pulse latencies where replaced with NaNs and then interpolated.

[#] Demeaning was applied when suggested to improve performance of the ICA.

[+] Automatic.

[§] Priori to this step, TMS-pulse interpolation was replaced with constant amplitude values. After this step, it was interpolated again.

[×] This step was added after the pipeline to allow comparability.

[@] Only for ocular artifacts.

process the TMS–EEG signal. It was designed to provide an easy-to-use streamline to guide the user (novice or advanced) through all the steps needed to clean raw TMS–EEG data. TMSEEG is an ICA-based pipeline (*fastICA* algorithm, Hyvärinen and Oja, 2000), with two ICA steps: the first aims at removing the large-amplitude artifacts, while the second removes the other TMS-related and common EEG artifacts. TMSEEG is fully manual: trials/channels and ICA-components' rejection are choices made by the user.

2.5.2. TMS–EEG signal analyser, tesa v1.0.1

The TESA (Rogasch et al., 2017) toolbox is a set of functions that works as an extension of the EEGlab software (Delorme and Makeig, 2004). Here the user can choose to run the functions from the EEG lab GUI or directly script the pipeline, calling the functions from MATLAB command window. The TESA toolbox was designed to provide multiple tools to the user that could be combined and used as preferred. For this reason, it has multiple core functions to remove artifacts from the signal (ICA, principal component analysis, enhanced deflation method, etc.); therefore, it does not have a unique pipeline. In this work, we chose the default pipeline suggested in (Rogasch et al., 2017). In this case, TESA is used to build an ICA-based pipeline (*fastICA* algorithm), with two ICA steps: the first aims to remove the large-amplitude artifacts (TMS-related muscle, decay and movement) and the second aims to remove the remaining artifacts. TESA can be used manually through the EEGlab GUI, semi-automatically (using automatic bad trials/channels rejection and supervised ICA-component rejection) or automatically (automatic bad trials/channels and ICA-component rejection), calling the functions through MATLAB command window. We employed the semi-automatic mode.

2.5.3. Automated aRTIfact rejection for Single-pulse TMS–EEG Data, artist

The idea behind ARTIST (Wu et al., 2018) is to remove the variability introduced by the user, by delegating most of the choices to an automatic algorithm. The user sets the necessary parameters for the cleaning (epoch length, TMS-pulse interpolation period etc.) and calls the main function from the MATLAB command window. The algorithm then performs all the necessary steps automatically. Like TMSEEG and TESA, it is an ICA-based pipeline (*Infomax* algorithm (Makeig et al., 1997)). It employs the first ICA run to remove the big decay artifact and the second

ICA run to remove the remaining artifactual data. Bad trials/channels and ICA-components are labelled and rejected by specific algorithms and trained classifiers.

2.5.4. Source-estimate-utilizing noise-discarding algorithm (SOUND) and signal-space projection–source-informed reconstruction (SSP–SIR)

The SOUND-SSP–SIR approach substitutes ICA largely with the SOUND (Mutanen et al., 2018) and SSP–SIR (Mutanen et al., 2016) algorithms as core functions to remove TMS–EEG artifacts. The SOUND algorithm reconstructs the signal in the source volume. Then, it attenuates artifactual segments of data in each sensor, by estimating how well that segment of signal is predicted by all other sensors. This process runs automatically and does not require the user to choose the signal to be removed. SSP–SIR, instead, is specifically designed to attenuate the TMS-related muscle artifacts, using their time-frequency features. The signal components that capture the TMS-related muscle activity can be rejected manually by the user, or automatically by setting an appropriate threshold. We chose to use SSP–SIR manually. The SOUND-SSP–SIR approach, applied here, is a combination of methods that have been unified in a coherent pipeline in previous studies (Bagattini et al., 2019; Bortoletto et al., 2020). Finally, an ICA step is also present in this pipeline (*Infomax* algorithm), but it is confined to removing ocular artifacts as these can be assumed to appear relatively independently from TMS-evoked brain signals. The SOUND-SSP–SIR pipeline does not have a GUI and it can be used by calling the functions through the MATLAB command window.

2.5.5. Preprocessing parameters and steps

Raw TMS–EEG signals were processed with the above-mentioned pipelines. Table 1 illustrates the workflow for each pipeline. Epoching was set to -/+1 s. Electrooculography (EOG) electrodes were removed using the built-in EEGlab function for channel removal. Given that TMS-pulse artifact lasts up to 5 ms with the employed recording settings (Veniero et al., 2009), data were removed and interpolated from –1 to 6 ms, which was sufficient to capture the full TMS-pulse artefact in our data. The recorded TMS-pulse artifact is depicted in Fig. 1S. Data were downsampled to 1000 Hz. When suggested, demeaning was applied using the whole epoch (TMSEEG and TESA). Baseline correction was set from –1000 to –2 ms. Channel and trial rejection was done either manually (TMSEEG) or automatically (ARTIST and TESA) employing built-in functions in EEGlab (Delorme and Makeig, 2004). The SOUND-SSP–SIR

pipeline requires only manual trial rejection, as bad channels are automatically corrected by the SOUND step. Data were bandpass filtered at 1–90 Hz and notch filtered at 48–52 Hz. When high-pass filtering was applied alone, the passband started at 1 Hz. Rereferencing was applied using the average of all electrodes. ICA and SSP–SIR component selection was performed either manually (TMSEEG, TESA, SOUND-SSP–SIR) or automatically (ARTIST). When run manually, ICA component selection was done independently by authors G.B. and R.E. and SSP–SIR component selection by G.B. and M.B. Components that were selected by just one author were discussed and a joint decision was made. For statistical analysis, the epoch was reduced to –100…+350 ms and baseline correction from –100 to –2 ms was performed.

2.6. Statistical analysis

First, we investigated the between-pipelines differences in trials, channels and ICA components removed during the preprocessing. Channels rejected and components removed in the ICA were summarized together using the data rank. The data rank represents the number of independent components that can be individuated in the data. The higher the rank, the "richer" are the data in terms of sources that compose the signal. In raw TMS–EEG signal, the rank is equal to the number of recording channels and it decreases by one unit during the preprocessing for each channel interpolated and ICA-component removed (minus one for the rereferencing). Depending on the preprocessing strategy, different pipelines interpolate a different number of bad channels and remove or attenuate a different number of components, hence impacting the rank in different ways. With methods like ICA or SSP, the decrease in rank corresponds to exactly the number of removed artifact components (signal dimensions). SOUND, however, weighs different signal dimensions based on the noise estimates. This means that while SOUND theoretically conserves the rank, it can effectively squash the noisiest dimensions close to zero. To allow a fair comparison between the different pipelines, we estimated the effective rank of the data as follows: We formed the singular-value decomposition of the data and defined the effective rank as the number of dimensions having singular value greater than 1% of the biggest singular value. Differences in data rank after the preprocessing were assessed with a separated one-way ANOVA per each condition (IPL session 1, IPL session 2, DLPFC session 1 and DLPFC session 2). Similarly, differences in trial removal were assessed with a Kruskal–Wallis test, to account for the non-normal distribution of the data. Post-hoc tests were applied when appropriate with alpha threshold 0.0042, for a series of six two-tailed tests with Bonferroni correction.

Reproducibility, i.e., the reliability of TEPs derived by different preprocessing pipelines, was assessed by testing for (1) differences between TEPs derived by different preprocessing pipelines, (2) similarities between TEPs derived by different preprocessing pipelines, and (3) comparing the test–retest reliability between TEP components in session 1 and session 2 across pipelines. Moreover, we took into account that some studies measure the TEPs traces over electrodes and some studies further calculate the global mean field power (GMFP). Therefore, the same three analyses described above were run on the GMFP as calculated in fieldtrip (Oostenveld et al., 2011).

Differences between TEPs derived by different preprocessing pipelines were investigated separately for the two areas (IPL, DLPFC) and sessions 1–2; in total, four separate one-way repeated measure ANOVAs with "preprocessing pipeline" as a four-level factor were used to test the hypothesis that there were differences in TEP amplitudes across preprocessing pipelines in some channels and time points after the TMS pulse (IPL session 1, DLPFC session 1, IPL session 2 and DLPFC session 2) . For controlling the familywise error rate, we used cluster-based permutation tests (Oostenveld et al., 2011). This strategy is commonly used when there is no a priori hypothesis on where or when the effect will be. The ANOVA with cluster-based correction was ran at each time point after the TMS-pulse interpolation (6 to 350 ms, SR: 1000 Hz) in each channel (63). Time points with significant test statistics (alpha

level was set at 0.05) were clustered together, constrained to those that involved at least 3 neighboring channels. Post-hoc tests were applied as a series of cluster-based two-tailed t-tests for paired samples between all possible combinations of pipelines, with alpha < 0.0042 (0.025 divided for six comparisons) to correct for multiple comparisons.

Similarities between TEPs derived by different preprocessing pipelines, were investigated through pairwise Spearman correlations in both temporal and spatial dimensions. The spatial correlation of TEPs for each pair of preprocessing pipelines was computed at each time point, by correlating voltage values of all channels in a specific pair of conditions, i.e., two preprocessing pipelines, for each individual. The resulting correlation values were then z-transformed and averaged across subjects. The z-coefficients were tested at each time point against zero using a two-tailed t-test, to assess the significance of the correlation. For each test, alpha level was set at 0.025 (two-tailed t-test). Since each contrast comprised a series of 344 t-tests (one for each time point) we controlled the False Discovery Rate (FDR) with the Benjamini–Hochberg correction (Benjamini and Hochberg, 1995; Martínez-Cagigal, 2020). The average z-coefficients were transformed back into correlation values using the inverse Fisher z-transformation. The temporal correlation between two TEPs derived by different preprocessing pipelines was computed at each channel with the same logic of the spatial correlation, i.e., by correlating the voltage values in a specific interval between two conditions. Intervals were the whole epoch (–100 +350 ms) and smaller time windows (6–80 ms, 81–150 ms and 151–350 ms). For interpretation of correlation values, we employed the scale of Shrout et al. 1998 (Shrout, 1998) (Table 5S).

Lastly, we assessed the test–retest reliability of the TEPs for each preprocessing pipeline with three analyses: First, we assessed the difference in TEPs between session 1 and 2 for each preprocessing pipeline using a series of paired-sample cluster-based t-tests. Alpha level was set at 0.0063 (two-tailed t-test Bonferroni corrected for four tests); Second, the similarities of TEPs between session 1 and 2 were investigated through pairwise Spearman correlations in both temporal and spatial dimensions and statistic of the correlations were computed as described for the between-pipelines tests; Third, we estimated the test–retest agreement between sessions 1 and 2, for each preprocessing pipeline, through the concordance correlation coefficient (CCC). CCC measures the agreement between two variables x and y, i.e., the test and retest measures, as distance from the line $x=y$ (Kerwin et al., 2018; King et al., 2007; Lin, 1989). Therefore, for this analysis, TEP components were individuated and their peak amplitudes and latencies were extracted following the collapsed localizer strategy suggested in Luck 2014 (Luck, 2014). To this end, TEPs of session 1 and 2 were averaged, followed by the choice of the time windows in which peaks were present in these grand-average signals. The identified time windows were then used separately in each condition to look for peak amplitudes and latencies. CCC was computed between sessions 1 and 2 for both peak-amplitude and peak-latency values. CCCs were considered significantly different from zero when their bootstrapped 95% confidence intervals (CI) did not include zero. Differences across preprocessing pipelines' CCCs were assessed using the percentile bootstrap method (Wilcox, 2009) within each peak. This method involves the calculation of the bootstrap distribution for the CCC difference between two conditions with its relative CI. If the CI of the CCC difference does not contain the zero, then the two conditions' CCCs are significantly different.

3. Results

In Fig. 1, TEPs with relative topographies are shown as obtained after being processed with the four pipelines, in both IPL (Fig. 1A) and DLPFC (Fig. 1B) for session 1. TEPs for session 2 are shown in supplementary results (Fig. 2S).

The motor threshold did not significantly differ between the two sessions (session 1: 66.3 ± 8.8%, session 2: 66.4 ± 8.2% of the maximum stimulator output). The Kruskal–Wallis test on of the number of

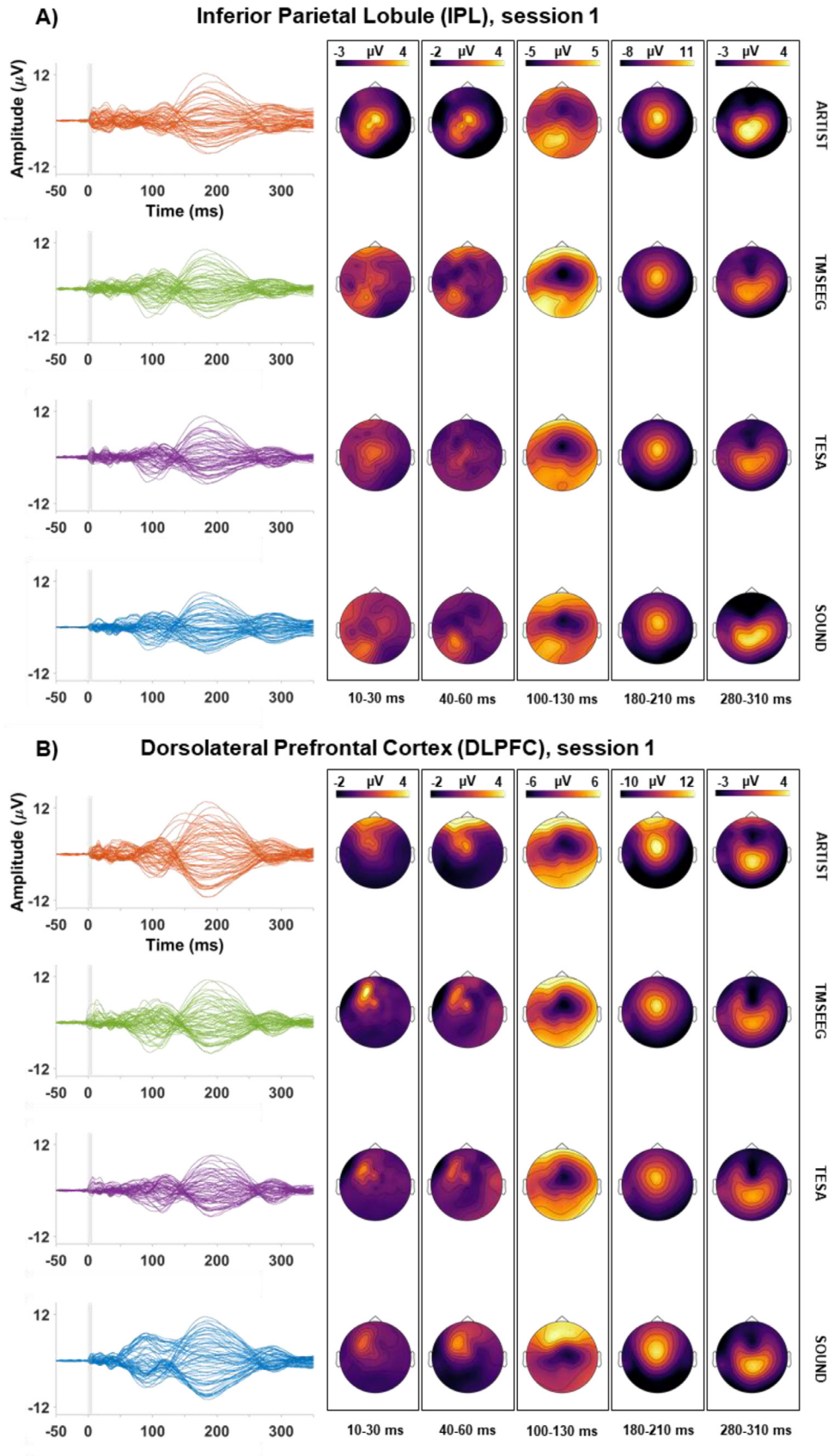**Fig. 1.** TEPs recorded after IPL (A) and DLPFC (B), processed with different pipelines (rows). The leftmost column depicts TEPs time-series, color-coded for each preprocessing pipeline. The other columns represent the scalp topographies of TEPs in five selected intervals. The topographies represent the mean voltage on the scalp in each interval. "SOUND" refers at the full SOUND-SSP–SIR pipeline.
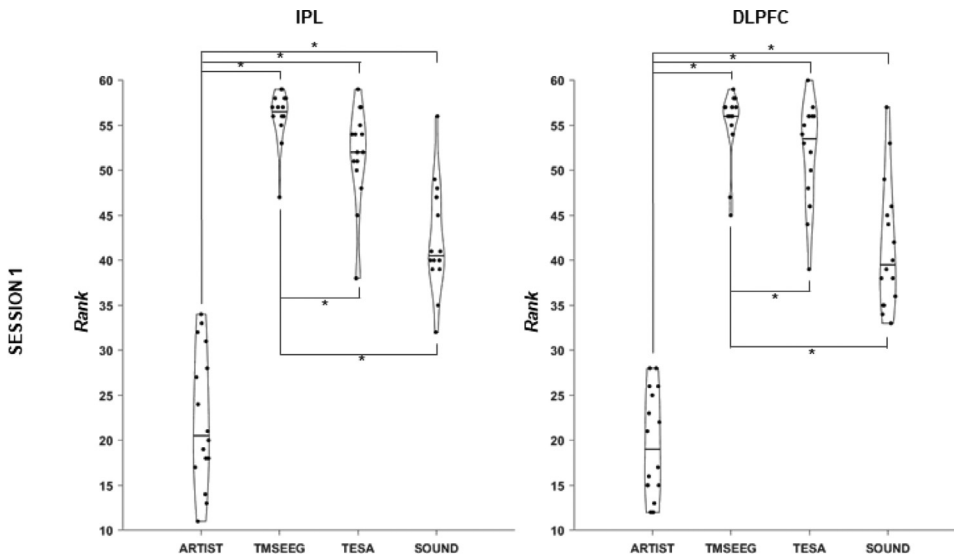
Fig. 2. Data rank after each preprocessing pipeline for IPL and for DLPFC in session 1. Each violin plot depicts the distribution of individual subject rank after the preprocessing. Horizontal black lines represent the median. Black lines, with starts asterisks, connects conditions significantly different in the post-hoc analysis (*p* threshold= 0.0042). "SOUND" refers at the full SOUND-SSP–SIR pipeline.

Table 2

IPL post-hoc paired t-tests, session 1. Cluster polarity reveals the direction of the significant cluster of differences. Cluster p report the specific probability of the cluster statistic. Cluster latency depict the extent in time of the significant cluster.

Comparison	Cluster Polarity	Cluster *p*	Cluster Latency (s)
ARTIST vs TMSEEG	+	<0.001	0.007 to 0.350
	–	0.002	0.015 to 0.350
ARTIST vs TESA	+	<0.001	0.009 to 0.350
	–	0.003	0.147 to 0.350
ARTIST vs SOUND–SSP–SIR	+	<0.001	0.006 to 0.350
	–	0.002	0.172 to 0.350
	–	0.002	0.006 to 0.149
TMSEEG vs TESA	+	0.002	0.010 to 0.343
TMSEEG vs SOUND-SSP–SIR	+	0.002	0.173 to 0.350
	–	<0.001	0.191 to 0.350
TESA vs SOUND–SSP–SIR	–	<0.001	0.134 to 0.350

Table 3

DLPFC post-hoc paired t-tests, session 1.

Comparison	Cluster Polarity	Cluster p	Cluster Latency (s)
ARTIST vs TMSEEG	+	0.004	0.199 to 0.350
ARTIST vs TESA	+	0.002	0.124 to 0.350
ARTIST vs SOUND-SSP–SIR	+	0.002	0.055 to 0.149
TMSEEG vs TESA	/		
TMSEEG vs SOUND-SSP–SIR	–	0.004	0.065 to 0.230
TESA vs SOUND-SSP–SIR	+	< 0.001	0.033 to 0.325
	–	0.003	0.045 to 0.304

differences in five contrasts out of six. The contrasts between TMSEEG vs TESA did not show significant differences (Table 3).

ANOVAs with related post-hocs between TEPs for session 2 are shown in the supplementary materials (Fig. 4S-5S, Tables 3S-4S).

3.2. TEP correlations

Despite the marked differences in the cluster-based *t*-tests, TEPs derived by different preprocessing pipelines were positively correlated in most of the time points of the epoch in both IPL and DLPFC (Fig. 5-6 A). As expected, spatial correlations in the baseline period were high for both areas (moderate-to-substantial, $\rho = 0.6$–0.9), since all preprocessing pipelines shared the same raw signal. This is different from previous studies where different conditions (raw signals) were correlated (e.g., Conde et al., 2019). In our case, we expected the baseline to be equal to one if not affected by the preprocessing steps, because the same dataset is used across pipelines. Spatial correlation dropped around the TMS pulse, and then slowly recovered. Correlation values at latencies under 100 ms were the most variable across comparisons ranging from approximately $\rho = 0.2$–0.9, while correlations at latencies over 100 ms were consistently moderate-to-substantial ($\rho > 0.6$) in all pairs.

Accordingly, whole-epoch temporal correlations (Fig. 5,6 B) were moderate-to-substantial ($\rho > 0.6$) for both areas. Correlations at window 6–80 ms were the most variable, showing slight-to-moderate correlation ($0.3 < \rho < 0.6$) over frontal and temporal electrodes. TEPs at 80–150 ms showed slight-to-moderate correlations over frontal electrodes in most comparisons ($0.2 < \rho < 0.9$), while the rest of the channels were substantially correlated ($\rho > 0.8$). At the 150–350-ms window, TEPs were moderately-to-substantially correlated in all pairs ($\rho > 0.6$).

Correlations between TEPs derived by different preprocessing pipelines for session 2 are shown in the supplementary materials (Fig. 6S-7S).

removed trials in session 1 (Table 1S) showed significant differences for IPL ($\chi^2 = 13.7499$, $p = 0.0033$) and DLPFC ($\chi^2 = 12.7422$, $p = 0.0052$). Post-hoc tests (*p* threshold = 0.0042) revealed significant differences in DLPFC where TMSEEG removed significantly more trials than SOUND-SSP–SIR ($t = 23.2813$, $p = 0.0022$). Data rank in session 1 was significantly different across pipelines for IPL ($F = 135.884$, $p < 0.001$) and for DLPFC ($F = 182.728$, $p < 0.001$). Post-hoc analyses showed that the rank for ARTIST was significantly smaller than the rank for the other pipelines, while TMSEEG rank was significantly higher than the others (Fig. 2 and Table 2S). Trial and rank analysis for session 2 are shown in the supplementary materials (Fig. 3S and Table 1-2S).

3.1. Differences in TEPs amplitude

In IPL, TEPs derived from the four preprocessing pipelines were significantly different in amplitude (IPL: *p* = 0.001, cluster window = 6–350 ms), as shown by the cluster-based ANOVA. In post-hoc tests (Fig. 3), differences were found in all contrasts and spanned for the whole epoch for all the comparisons with ARTIST and TMSEEG vs TESA. In TMSEEG vs SOUND-SSP–SIR and TESA vs SOUND-SSP–SIR the differences remained confined to late latencies (Table 2).

Similarly, TEPs for DLPFC stimulation were different across processing pipelines (DLPFC: *p* = 0.001, cluster window = 26–350 ms), as shown by the cluster-based ANOVA. Post-hoc tests (Fig. 4) showed significant
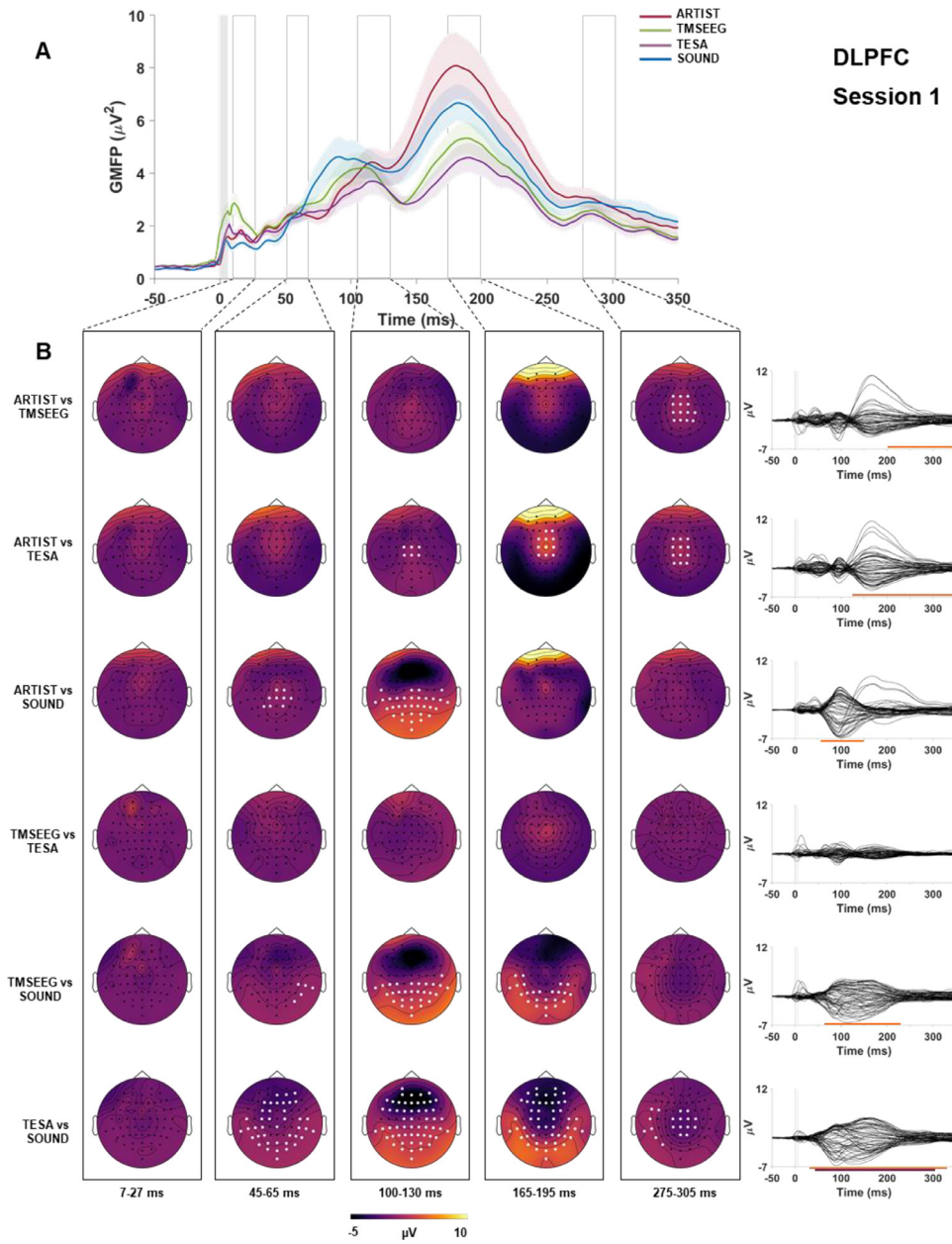
Fig. 3. A: Global Mean Field Power (GMFP, *y* axes) over time (ms, *x* axis) of TEPs resulting from IPL stimulation, cleaned with the four preprocessing pipelines (color-coded). Shaded area around each colored line represents SEM. Shaded grey column around zero represents the TMS-pulse interpolation interval. B: scalp topographies of the voltage differences (color-coded) for each condition contrast (rows) in five selected time windows (columns). White dots represent significant channels. The rightmost column represents the time-series of the voltage differences over time in each contrast. Colored bars at the bottom of each time-series represent the temporal extend of significant cluster(s). Orange positive clusters, purple negative cluster. "SOUND" refers at the full SOUND-SSP–SIR pipeline.

3.3. Test–retest reliability of preprocessing pipelines

Difference in TEP amplitude between session 1 and session 2 were found only for the TESA pipeline for DLPFC stimulation being significantly different in amplitude in the first session compared to the second session (Significant cluster: *p* =0.005, cluster window = 153–319 ms) (Fig. 8S-9S). No other significant differences were found for other pipelines in DLPFC and for IPL stimulation.

Analyses on TEPs spatial similarities revealed that TEPs across sessions were positively correlated, in most of the time points of the epoch, in both IPL and DLPFC (Fig. 10S-11S A). Spatial correlations showed an increasing trend from virtually-none-to-fair (0.1<ρ<0.6) around the TMS pulse toward 180 ms, in which it reaches moderate-to-substantial values (ρ>0.6). ARTIST in IPL was the only pipeline deviating from this trend, with slight correlation (0.2<ρ<0.4) around 80 ms.

Whole-epoch temporal correlations (Fig. 10S-11S S B) were slight-to-fair (0.2<ρ<0.6) for both areas, with a trend of lower values in frontal areas. Correlations at window 6–80 ms were the most variable, showing slight-to-moderate correlation (0.2<ρ<0.6) over frontal and temporal electrodes and substantial correlation (ρ>0.8) around Cz. Similarly,

TEPs at 80–150 and 150–350 ms showed slight-to-moderate correlations over frontal electrodes in most comparisons (0.2<ρ<0.6), while the rest of the channels were substantially correlated (ρ>0.8).

For the test–retest agreement estimation, we identified seven peaks on TEPs for the IPL (P15, P/N20, P50, N100, P120, P200 and P300) and five peaks for the DLPFC (P20, P50, N100, P200 and N300). Notably, the same peaks were not identifiable in all TEPs: P15 in IPL was found only for SOUND, P120 in IPL only for ARTIST, P50 in IPL was found for ARTIST and TMSEEG while P50 in DLPFC was found only for ARTIST. Fig. 7 and Tables 4,5 represent the amplitude and latency CCCs in session 1 vs session 2, for all preprocessing pipelines. CCCs were considered significant when the corresponding bootstrapped CIs did not include zero.

As in the above section, for the interpretation of CCC values we employed the scale in Shrout, 1998 (Shrout, 1998) (Table 5S).

For IPL, CCCs for amplitude were virtually-none-to-fair (-0.03 – 0.59) in early peaks (P15, P/N20, P50) and virtually-none-to-moderate (-0.10 – 0.79) in late peaks (N100, P120, P200 and P300). Notably, the CCC of N100 was significantly smaller for ARTIST compared to the other preprocessing pipelines (ARTIST vs TMSEEG p=0.002, ARTIST vs TESA

Fig. 4. A: Global Mean Field Power (GMFP, y axes) over time (ms, x axis) of TEPs resulting from DLPFC stimulation, cleaned with the four preprocessing pipelines (color-coded). Shaded area around each colored line represents SEM. Shaded grey column around zero represents the TMS-pulse interpolation interval. B: scalp topographies of the voltage differences (color-coded) for each condition contrast (rows) in five selected time windows (columns). White dots represent significant channels (p threshold = 0.0042, cluster based corrected). The rightmost column represents the time-series of the voltage differences over time in each contrast. Colored bars at the bottom of each time-series represent the temporal extend of significant cluster(s). Orange positive clusters, purple negative cluster. "SOUND" refers at the full SOUND-SSP–SIR pipeline.

Table 4
CCCs (CIs) for amplitude across pipelines in both IPL and DLPFC. Significant values in bold.

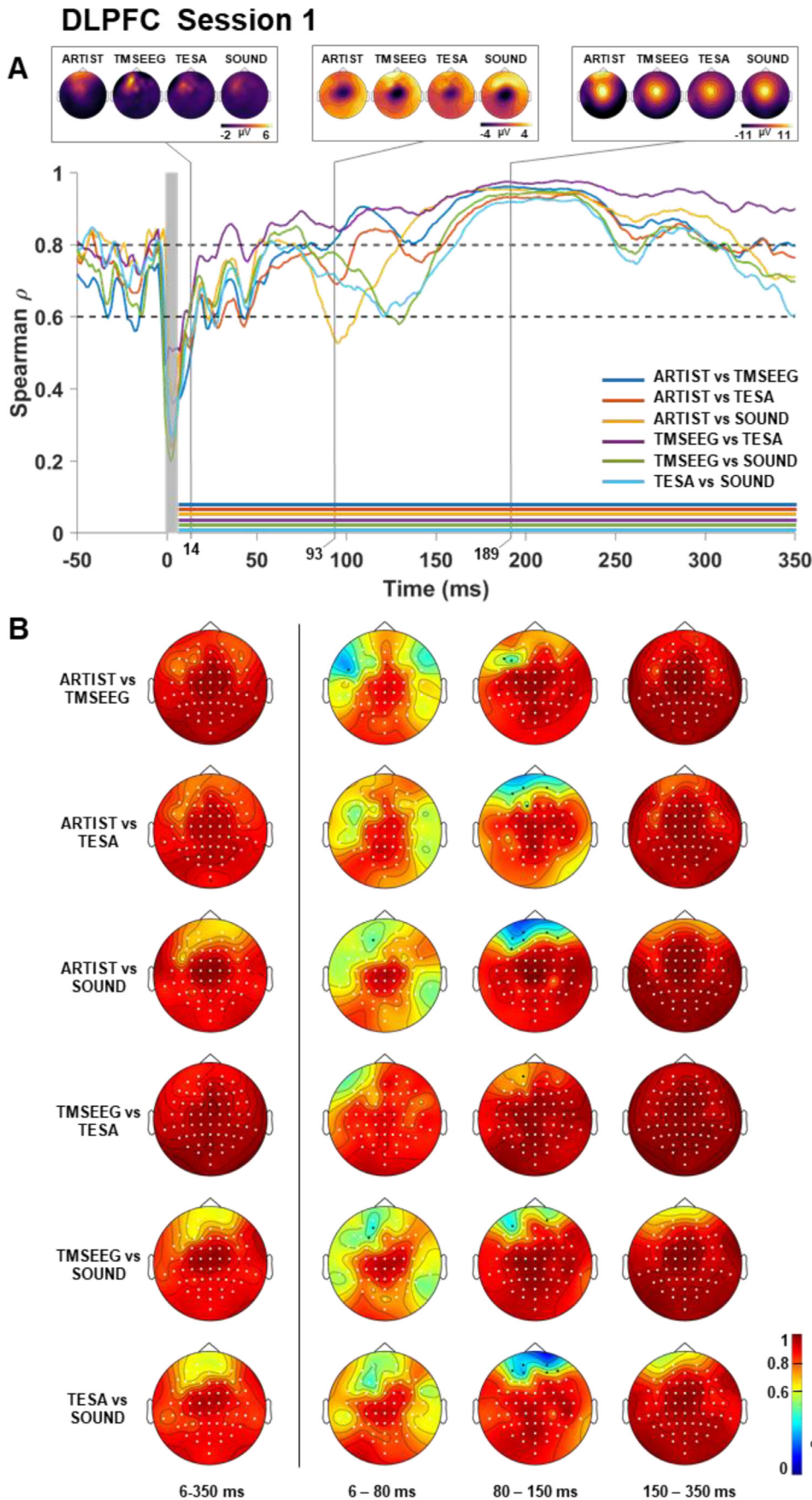| CCC		Amplitude					
	IPL						
	P15	P/N20	P50	N100	P120	P200	P300
ARTIST		0.46 (−0.01, 0.77)	0.37 (−0.12, 0.72)	−0.10 (−0.46, 0.28)	**0.64 (0.24, 0.86)**	**0.73 (0.41, 0.89)**	**0.55 (0.10, 0.82)**
TMSEEG		**0.59 (0.22, 0.81)**	**0.41 (0.03, 0.69)**	**0.68 (0.30, 0.87)**		**0.72 (0.39, 0.89)**	**0.59 (0.16, 0.83)**
TESA		0.46 (−0.02, 0.77)		**0.54 (0.09, 0.81)**		**0.77 (0.47, 0.91)**	**0.48 (0.05, 0.76)**
SOUND SSP–SIR	0.20 (−0.20, 0.55)	−0.03 (−0.47, 0.42)	0.37 (−0.03, 0.66)	**0.73 (0.40, 0.90)**		**0.69 (0.33, 0.88)**	0.47 (−0.01, 0.78)
	DLPFC						
	P20	P50	N100	P200	N300		
ARTIST	−0.02 (−0.48, 0.45)	0.43 (−0.04, 0.75)	**0.49 (0.04, 0.77)**	**0.85 (0.63, 0.94)**	**0.66 (0.27, 0.87)**		
TMSEEG	0.42 (−0.04, 0.73)		**0.48 (0.09, 0.74)**	**0.83 (0.59, 0.93)**	**0.80 (0.53, 0.92)**		
TESA	0.14 (−0.35, 0.53)		**0.64 (0.24, 0.86)**	**0.83 (0.61, 0.93)**	**0.77 (0.50, 0.90)**		
SOUND SSP–SIR	0.10 (−0.38, 0.53)		**0.87 (0.70, 0.95)**	**0.78 (0.50, 0.91)**	0.32 (−0.17, 0.68)		
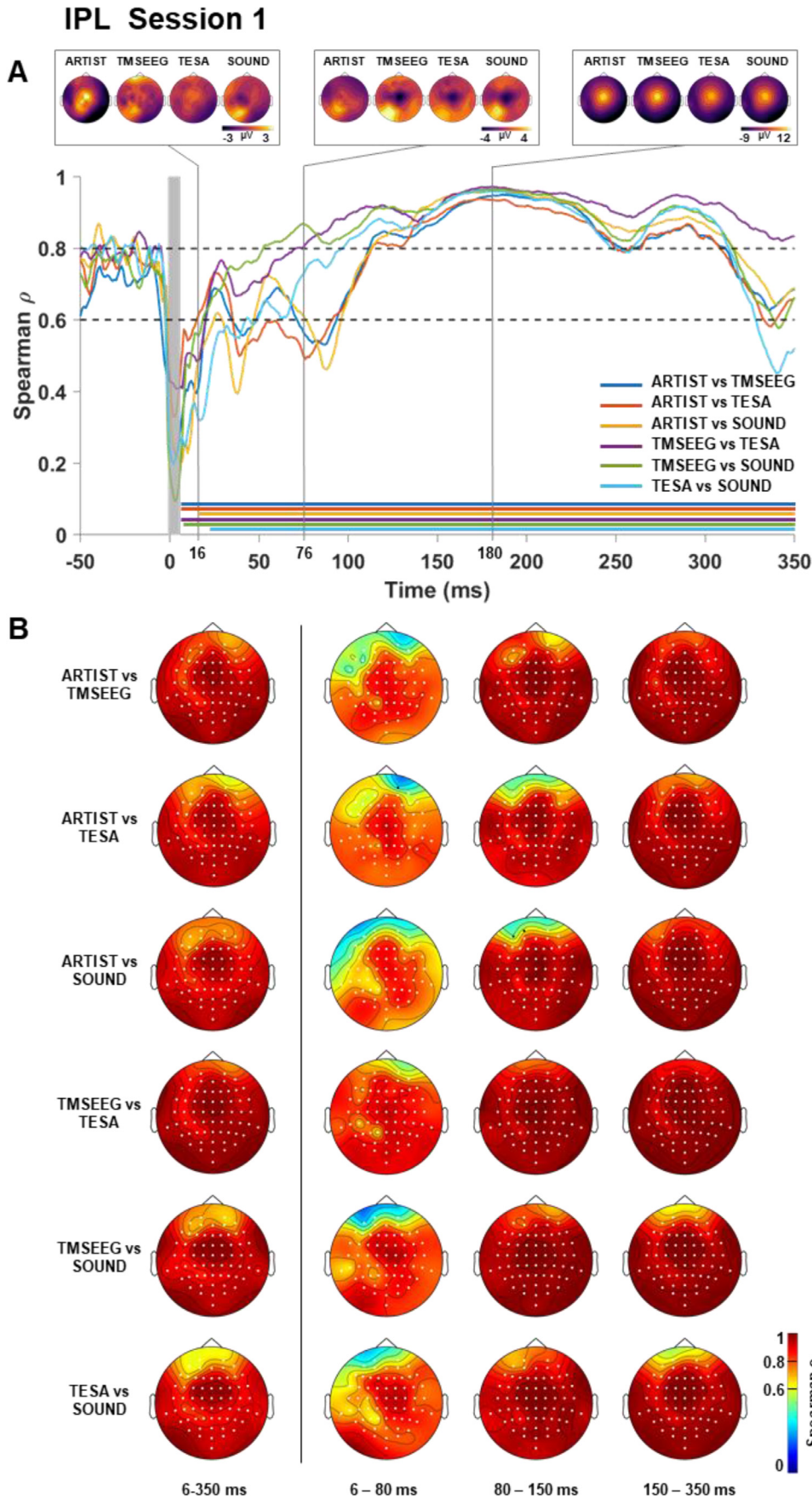
Fig. 5. IPL, TEPs spatial and temporal correlation. (A) Each colored line represents the spatial correlation (*y* axis) over time (*x* axis) of all condition contrasts. Shaded grey column around zero represent the TMS-pulse interpolation interval. Horizontal dotted line represent threshold for moderate (0.6) and substantial (0.8) correlation, according to Shrout et al. (1998). Horizontal colored lines represent instant in time in which the correlation resulted significantly different from zero (*p* threshold = 0.0042, FDR corrected). On the top, are depicted three representative instantaneous scalp topographies in the four conditions. Voltage on the scalp topographies is color-coded. (B) Temporal correlation of each contrast (rows) in four time intervals (columns). Correlation values are color-coded. Channels significantly different from zero are highlighted in white (*p* threshold = 0.0042, FDR corrected). "SOUND" refers at the full SOUND-SSP–SIR pipeline.

Fig. 6. DLPFC, TEPs spatial and temporal correlation. (A) Each colored line represents the spatial correlation (y axis) over time (x axis) of all condition contrasts. Shaded grey column around zero represent the TMS-pulse interpolation interval. Horizontal dotted line represent threshold for moderate (0.6) and substantial (0.8) correlation, according to Shrout et al. (1998). Horizontal colored lines represent instant in time in which the correlation resulted significantly different from zero (p threshold = 0.0042, FDR corrected). On the top, are depicted three representative instantaneous scalp topographies in the four conditions. Voltage on the scalp topographies is color-coded. (B) Temporal correlation of each contrast (rows) in four time intervals (columns). Correlation values are color-coded. Channels significantly different from zero are highlighted in white (p threshold = 0.0042, FDR corrected). "SOUND" refers at the full SOUND-SSP–SIR pipeline.

Table 5
CCCs for latency across pipelines in both IPL and DLPFC. Significant values in bold.

CCC	Latency					
IPL						
	P15	*P/N20*	*P50*	*N100*	*P120*	*P200*	*P300*
ARTIST		**0.48 (0.02, 0.77)**	**0.59 (0.16, 0.83)**	**0.77 (0.48, 0.91)**	0.18 (−0.29, 0.59)	**0.72 (0.38, 0.89)**	0.36 (−0.12, 0.70)
TMSEEG		**0.58 (0.14, 0.83)**	**0.52 (0.05, 0.80)**	**0.91 (077, 0.97)**		**0.77 (0.46, 0.91)**	0.33 (−0.17, 0.69)
TESA		0.09 (−0.36, 0.51)		**0.72 (0.37, 0.89)**		**0.74 (0.40, 0.90)**	0.26 (−0.23, 0.65)
SOUND SSP–SIR	0.39 (−0.02, 0.69)	0.39 (−0.07, 0.72)	**0.47 (0.05, 0.74)**	**0.76 (0.47, 0.90)**		**0.76 (0.44, 0.91)**	0.28 (−0.23, 0.67)
DLPFC						
	P20	*P50*	*N100*	*P200*	*N300*	
ARTIST	−0.06 (−0.49, 0.39)	0.26 (−0.21, 0.63)	**0.90 (0.74, 0.96)**	**0.83 (0.62, 0.93)**	0.44 (−0.03, 0.75)	
TMSEEG	0.16 (−0.31, 0.57)		**0.55 (0.11, 0.81)**	**0.87 (0.67, 0.95)**	**0.66 (0.26, 0.86)**	
TESA	0.25 (−0.19, 0.61)		**0.74 (0.46, 0.88)**	0.22 (−0.28, 0.62)	0.47 (−0.01, 0.78)	
SOUND SSP–SIR	0.09 (−0.29, 0.45)		**0.65 (029, 0.85)**	**0.70 (0.38, 0.87)**	−0.14 (−0.55, 0.32)	
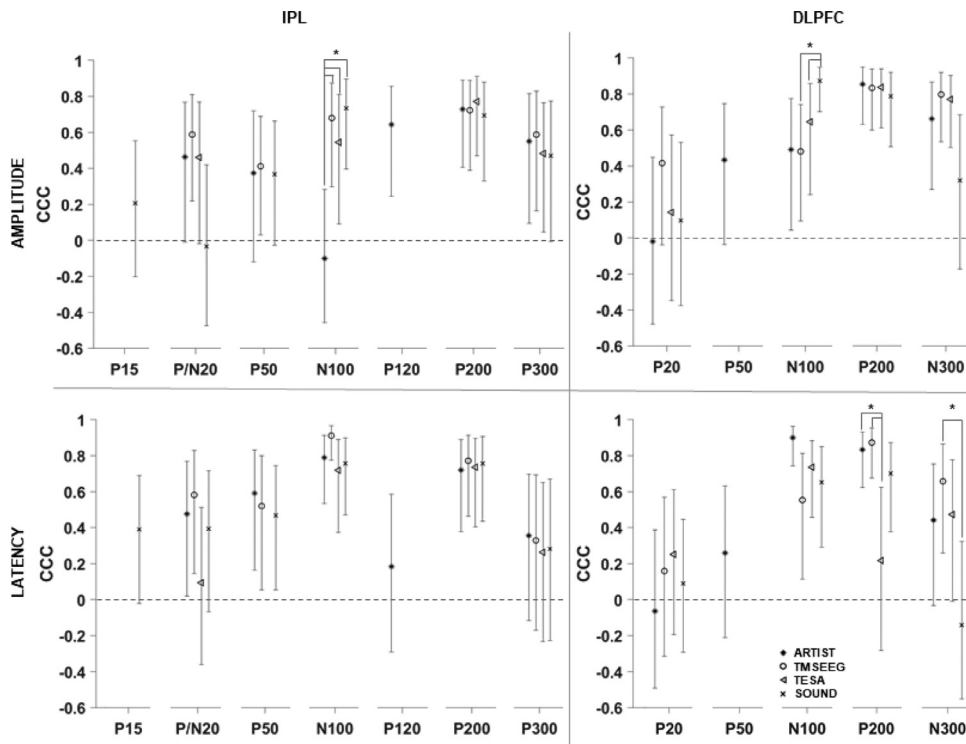
Fig. 7. Test-retest agreement of peaks amplitude and latency (rows) in IPL and DLPF (columns), computed on TEPs obtained from the four preprocessing pipelines (see legend). Note that peaks P15, P50 and P120 were not present after all preprocessing pipelines. Vertical error bars represent bootstrapped CIs. CCCs were considered significantly different from zero when the CIs did not include zero (*p* < 0.05). Asterisks represent significantly different CCCs within the same peak. "SOUND" refers at the full SOUND-SSP–SIR pipeline.

p=0.004 and ARTIST vs SOUND-SSP–SIR *p*<0.001). Latency CCCs followed a similar pattern (Fig. 7).

For DLPFC, CCC for amplitude was virtually-none-to-fair (-0.02 – 0.43) in early peaks (P20, P50) and increased to slight-to-substantial (0.32 – 0.88) in later peaks (N100, P200 and N300). SOUND-SSP–SIR CCC resulted significantly higher in N100 compared to TMSEEG (*p*<0.001) and TESA (*p*=0.022). Latency CCCs in early peaks (P20, P50) were relatively stable around virtually-none-to-slight concordance values (−0.06 – 0.26), while it varied considerably across preprocessing pipelines in later components (N100, P200 and N300), ranging from substantial for ARTIST (0.90) and virtually none for SOUND-SSP–SIR (-0.14). TESA CCC were significantly smaller in P200 compared to ARTIST (*p*=0.010) and TMSEEG (*p*=0.036) but not SOUND-SSP–SIR, while TMSEEG CCC resulted significantly higher in N300 compared only to SOUND-SSP–SIR (*p*=0.034).

3.4. Global mean field power (GMFP)

The analyses on the GMFP resulted in a pattern of effects substantially similar to the TEP traces. These results can be found in detail in the supplementary materials.

First, we found that the GMFPs derived from the four preprocessing methods were significantly different in amplitude both for IPL and DLPFC and for both session 1 and 2 (Fig. 12S:15S Tables 6S:9S); Second, the temporal correlations of the GMFP between pipelines were variable, with higher values at later latencies starting from 80 ms (Fig. 16S:19S) and so was the reliability of GMFP. Despite no differences were found in GMFP amplitude between sessions (Fig. 20S-21S), the temporal correlation between sessions for all pipelines varied between slight to substantial (Fig. 22S-23S) and the CCC varied in a range between virtually none to substantial (Fig. 24S; Table 10S-11S).

4. Discussion

We found significant differences in TEP amplitudes across preprocessing pipelines for both stimulated areas, i.e., IPL and DLPFC, in both session 1 and 2. Spatial and temporal correlations were generally moderate-to-substantial, although lower values were found at the beginning of the epoch and on frontal electrodes. Moreover, the test–retest agreement of TEPs was highly variable across components and also varied across preprocessing pipelines.

The same analyses on the GMFP showed a similar pattern of results. These findings suggest a marked impact of the choice of the preprocessing pipeline on the TEPs, mainly at early latencies, which may affect their reproducibility.

Variability brought by the choice of the preprocessing pipeline in the TEPs

The application of different processing pipelines produced TEPs that were significantly different in amplitude and were variably correlated, for both IPL and DLPFC.

Interestingly, TEPs at latencies below and above 100 ms showed different patterns.

Early-latency TEPs showed less differences in amplitude than late ones, but their correlations between pipelines and between sessions were variable and below moderate. The temporal correlations in the window 6–80 ms indicated that most of the low correlations were located in the fronto-temporal regions, where the muscular artifact is stronger. Finally, a general trend of low test–retest agreement was found in early peaks (P15, P20 and P50).

Early components are informative in the study of effective connectivity as they reflect fast activations of remote areas directly or indirectly connected with the stimulated region (Bortoletto et al., 2021). Unfortunately, they are superimposed to high-amplitude artifacts such as muscular and decay artifacts, that result in low SNR. Therefore, the preprocessing pipeline is likely to have a strong impact on the recovering of this part of the signal. To mitigate this issue, the signal around the TMS artefact is generally cut and lost for a longer time window that varies between 5 and 15 ms across studies. Nevertheless, according to the indication in the published preprocessing pipelines, the cut is meant to remove the TMS-pulse artifact, that lasts no longer than 5–6 ms with our settings (Veniero et al., 2009), while other steps are indicated for removal of high-amplitude artefacts, e.g., a first-round of ICA (Atluri et al., 2016; Rogasch et al., 2017; Wu et al., 2018). How the length of the interpolation window may reduce the variability that we have found is not known and should be investigated in further studies. However, a longer cut around the TMS pulse may not eliminate the detrimental effects of the initial artefacts. Previous studies that have assessed the test–retest agreement of TEPs outside the motor cortex reported contrasting results for early peaks (Casarotto et al., 2010; Kerwin et al., 2018; Lioumis et al., 2009; Ozdemir et al., 2020) . Specifically, two studies in which the TMS pulse was cut for more than 10 ms showed a reduced reliability in early peaks (Kerwin et al., 2018; Ozdemir et al., 2020). Differently, two studies that employed sample and hold amplifiers reported high reliability of early peaks (Casarotto et al., 2010; Lioumis et al., 2009).

A macroscopic consequence of preprocessing-related variability may be the inconsistent definition of TEP components across studies. For example, beside the N100–P200 complex, TEP components are not yet fully characterized outside the motor cortex. Most of the studies on DLPFC stimulation agree on deflections such as the N40, P60, N100 and P185 (Bagattini et al., 2019; Chung et al., 2019; Conde et al., 2019; Gordon et al., 2018b; Rogasch et al., 2014; Voineskos et al., 2019), but other peaks are often reported at ~30 ms, ~50 ms, or ~70 ms (Bagattini et al., 2019; Conde et al., 2019; Rogasch et al., 2014). For IPL, there is still little characterization of TEP components and the few studies that reported them have shown inconsistent results (Conde et al., 2019; Rogasch et al., 2020; Romero Lauro et al., 2014). It is already known that the peak characterization may be drastically altered should only one TMS-related artifact be left unhandled in the signal (Rogasch et al., 2014). Therefore, the variability in TEP components observed after the data have been preprocessed may be caused by a different artifact suppression in each preprocessing pipeline, especially for large-amplitude early-latency TMS-related artifacts, which are the most difficult to handle.

Late TEP components, i.e. above 100 ms, showed different amplitudes across pipelines for both areas but were highly correlated with each other in both space and time. This suggests that the functions used to reduce artifacts at those latencies yielded TEPs with similar unfolding of components, but with different amplitudes. Moreover, the test-retest reliability was between moderate to substantial for all late components except for the N/P300 latency that was highly variable, probably reflecting the slow response at the end of the epoch.

Late peaks, i.e., N100 and P200, are known as the most reliable and stable peaks of the TEPs (Biabani et al., 2019; Conde et al., 2019; Kerwin et al., 2018; Lioumis et al., 2009; Ozdemir et al., 2020). These studies are in line with our findings in that late latencies were the most resilient to the preprocessing pipeline in terms of shape, resulting in high correlations. However, it is likely that the earplugs employed in this study were not sufficient to suppress auditory response to the click emitted by the TMS pulse (Biabani et al., 2019; Conde et al., 2019; Nikouline et al., 1999; Nikulin et al., 2003; ter Braack et al., 2015; Tiitinen et al., 1999). Moreover, none of the pipelines provided correction for auditory artefacts. Therefore, these data should be interpreted with caution as the later peaks may be a mixture of genuine TEPs and afferent responses, which could itself be a reliable response. This issue could be solved with an appropriate sham control, although this issue continues to be debated in the TMS–EEG field (Belardinelli et al., 2019; Conde et al., 2019; Siebner et al., 2019).

Possible sources of variability between preprocessing pipelines

The differences observed between the TEP preprocessing pipelines could derive from multiple factors. The main structural differences between the pipelines used in this study were (1) the order of the steps in the pipeline, (2) the type of core functions to remove TMS-related artifacts and (3) the level of automatism. The order of the steps in processing TMS–EEG signal is crucial, because the interaction of some procedures, i.e., filtering and ICA, with the large-amplitude artifacts might introduce additional analysis-related artifacts to the signal (Rogasch et al., 2017).

Regarding the type of core functions to remove TMS-related artifacts, these methods employ different ICA algorithms [ARTIST uses *Infomax* (Makeig et al., 1997), TMSEEG and TESA use *fastICA* (Hyvärinen and Oja, 2000)] or other methodologies such as the SOUND-SSP–SIR (Mutanen et al., 2018, 2016). *Infomax* and *fastICA* were benchmarked in another study against another ICA algorithm called adaptive mixture ICA (AMICA) (Palmer et al., 2008), which retains the best performance in separating neuronal components so far (Delorme et al., 2012). Both *Infomax* and *fastICA* were highly correlated in terms of component scalp maps and spectral profile with AMICA components. This suggests that using different ICA algorithms in the TEPs should generate minimal variability, since the choice of artifactual components is based on the component scalp maps and spectral profile. However, the selection of ICA, which is done manually by the experimenter in some cases and by automatic algorithms in others, may have impacted the final result. A markedly different approach is the one from SOUND-SSP–SIR. As mentioned in the methods, SOUND cleans the signal in each sensor at the source level utilizing the information coming from nearby sensors, while SSP–SIR attenuates the TMS-induced muscular artifact detecting its time–frequency features at the source level. Another study already compared ICA and SSP–SIR (without SOUND) approaches in recovering the TMS–EEG neuronal signal (Biabani et al., 2019). There, ICA was found to reduce and distort the signal more than SSP–SIR. In contrast with our expectation, we found that SOUND-SSP–SIR' TEPs did not show marked statistical differences with all the ICA-based methodologies. Nonetheless, ICA might not be the ideal tool to separate brain signal from TMS-related artifact, because one of its core assumptions is that these two signals are independent from each other. Although this

assumption might hold in the EEG alone, in TMS–EEG many artifacts are time-locked to the TMS pulse, as it is the brain signal, breaking the assumption of independency (Metsomaa et al., 2014). This might be an issue for the ICA-based preprocessing pipelines.

As mentioned in Section 2, the level of automation was substantially different across pipelines. ARTIST was designed to be fully automatic, in order to minimize the error due to subjective choices of the experimenter. Therefore, trials, channels and ICA components were rejected without supervision. In our study, this feature led to a substantial difference in the number of channels and ICA components rejected from this methodology. This could explain why ARTIST TEPs resulted as the most different in terms of amplitude in both IPL and DLPFC and also why spatial correlations with ARTIST deviated from the others until almost 125 ms in the IPL. TMSEEG, TESA and SOUND-SSP–SIR also showed significant differences in rank after the preprocessing between each other, although remarkably less pronounced than with ARTIST. This is also reflected in the higher spatial and temporal correlations and in the reduced differences of the cluster-based ANOVA between those pipelines. It is interesting to note, that ARTIST shows the highest amplitude in TEPs despite having the lowest rank. This may be due to the removal of several low-signal-power independent components in ARTIST. Trial removals were different between pipelines. Specifically, SOUND-SSP–SIR pipeline consistently removed less trials than the other pipelines. This is not surprising, since SOUND-SSP–SIR approach is to clean the data before doing the trial rejection, in order to spare the signal as much as possible.

Lastly, it should be noticed that we investigated only four published preprocessing pipelines, while most of the TMS–EEG studies use custom preprocessing pipelines. Therefore, our results could barely represent the unaccounted variability possibly resulting from the various preprocessing pipelines within other TMS–EEG reports.

What to do to overcome the variability brought by the preprocessing approach?

Variability brought by differences in the preprocessing phase is a crucial emerging issue that we have reported here for TMS–EEG and that has been recently reported also for other techniques applied in the neuroscience field, such as fMRI (Botvinik-Nezer et al., 2020; Lindquist, 2020). This variability on TMS–EEG studies may be reduced when the results are based on the contrast between two conditions processed with the same pipeline, e.g., TEPs pre vs. post treatment or TEPs rest vs. task, and may be more critical in studies on resting-state TMS–EEG aimed to investigate effective connectivity, where there is no contrast to control the impact of a chosen pipeline. Nevertheless, in both cases accuracy of results would be affected if a pipeline does not reliably remove artefactual components and/or if it removes part of the evoked signal.

A first strategy to reduce this problem may be to improve the signal-to-noise ratio during TMS-EEG recording. TEP signal recorded following DLPFC or IPL stimulation is known to be particularly problematic, because there are no established procedures to set TMS parameters, e.g., TMS intensity and coil orientation, to obtain an optimal cortical response. Moreover, TMS pulses in those areas might interact with head and neck muscles, causing muscular twitches. In our study, the intensity of stimulation was set at 100% MT to minimize the discomfort when stimulating frontal regions, reducing muscular twitches. Despite this choice is in line with several TMS-EEG studies, it might have contributed to TEPs with low SNR. Other strategies to set TMS parameters include setting the TMS intensity based on a model the induced electric field in the target area and adjusting coil orientation to reduce muscle artefacts. However, it should be considered that both these strategies affect the signal and the noise at the same time: higher TMS intensities increase the signal as well as the artefacts, coil orientation modulates the amplitude of cortical responses and of muscular twitches at the same time. Therefore, an important point to be addressed in future studies concerns the strategies to increase SNR during recordings and whether a TMS–EEG dataset recorded in conditions of a higher SNR could yield more robust results, less susceptible to the choice of the preprocessing pipeline.

Nevertheless, the main obstacle for finding the best-performing pipeline is the lack of a ground truth, i.e., the knowledge of the underlying clean signal, that would allow one to assess the efficacy of each preprocessing pipeline. One way could be to run simulations of TMS–EEG data, where the ground truth is known. This approach presents some technical difficulties including a plausible reconstruction of EEG signals and an accurate modelling of TMS-induced artefacts. Regarding the former, there has been recent development in simulating physiologically realistic EEG signals (Neymotin et al., 2020). Regarding the latter, currently available realistic sham conditions have been performed by distancing the TMS coil from the scalp and delivering a pulse of electric stimulation (Conde et al., 2019; Duecker and Sack, 2015; Gordon et al., 2018a; Mennemeier et al., 2009). However, artefacts induced by electrical stimulation may be different from the ones induced by magnetic stimulation. A more realistic model of TMS-induced artefacts may be obtained with coils that are designed for a fast decrease of the magnetic field with distance, so that the surface of the head can be stimulated in a similar way as in the active TMS condition while less stimulation reaches the cortex (Deng et al., 2013).

Alternatively, validation of TEPs may be obtained with more empirical approaches including, validation through well-established physiological effects and other neuroimaging techniques. For example, recording TEPs from the same area in two conditions that are known to change cortical excitability, e.g., resting state vs. motor preparation (Nikulin et al., 2003) or before vs. after drug intake (Premoli et al., 2014). If the experimental conditioning should not affect the artifact sources, a well-performing analysis pipeline should uncover the expected TEP changes. Furthermore, validation of TEPs measures may be derived from testing the relationships with other already physiological and neuroimaging measures. A recent study brought evidence that an early component of TEPs obtained with the SOUND-SSP-SIR pipeline after M1 stimulation, the P15, reflects the activation of the contralateral motor cortex, as it has been shown to correlate with a peripheral measure of transcallosal inhibition and with microstructural measures of the corpus callosum connecting the motor cortices (Bortoletto et al., 2021). Despite a single study is not enough to validate a measure or the pipeline employed, an approach that combines multiple measures may allow to highlight physiological responses.

Already now, even when the ground truth is missing, some methods, such as linear spatial filtering methods like SOUND and SSP–SIR, allow a straightforward quantification of the over correction across all the possible cortical sources (Mutanen et al., 2016). Because the spatial filters are perfectly known, one can apply them to all lead-field topographies to determine which sources would be attenuated the most by, e.g., SOUND or SSP–SIR. This approach enables the assessment of potential distortions in the effective connectivity patterns of interest and the suitability of the chosen analysis pipeline for the research question i.e., the researcher could already quantify how much the applied linear spatial filters may attenuate the ground truth EEG signals assuming they are coming from a cortical region of interest.

In conclusion, we believe that the TMS-EEG community should focus more on the validation of these technical aspects in order to build a more solid base for all future TMS-EEG applications. The adoption of good practices similar to the ones proposed for fMRI by Botvinik-Nezer et al. (2020) and including the sharing of code and data, may promote validation of analysis pipelines by allowing more researchers to apply their own pipelines to others' data and, viceversa, to apply others' pipelines to their own data. Moreover, the adoption of preregistration is another practice that may improve the definition of the analysis pipeline: pre-registration requires to define the preprocessing steps based on current understanding of available methods and possibly in combination with the application of the preprocessing to pilot data, without possible bias induced by the final results. Despite pre-registration does not guarantee a well-functioning pipeline, as it does

not provide a technical validation, it can foster good practices in the planning of the preprocessing phase. Moreover, when there are no *a priori* hypotheses on a particular preprocessing pipeline, the use of a multiverse analysis (Steegen et al., 2016), i.e., the use of multiple statistical and analytical approaches in the same study to reach consensus findings, would be a way to control the variability intrinsic to the preprocessing phase, yielding more solid results.

Data and code availability statement

BIDS-formatted data (Pernet et al., 2019) and code from this study are available on gin.g-node.org/CIMeC/TMS-EEG_brain_connectivity_BIDS.

Author contributions

Giacomo Bertazzoli: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization.

Romina Esposito: Methodology, Data Curation, Investigation, Writing – Review & Editing.

Tuomas P. Mutanen: Conceptualization, Methodology, Software, Validation, Writing - Review & Editing

Clarissa Ferrari: Methodology, Formal analysis, Writing - Review & Editing.

Risto J. Ilmoniemi: Conceptualization, Methodology, Writing - Review & Editing.

Carlo Miniussi: Conceptualization, Methodology, Resources, Writing - Review & Editing, Supervision

Marta Bortoletto: Conceptualization, Methodology, Validation, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision

Declaration of Competing Interest

The authors declare no competing financial interests.

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2021.118272.

References

Atluri, S., Frehlich, M., Mei, Y., Garcia Dominguez, L., Rogasch, N.C., Wong, W., Daskalakis, Z.J., Farzan, F., 2016. TMSEEG: a MATLAB-based graphical user interface for processing electrophysiological signals during transcranial magnetic stimulation. Front. Neural Circuits 10, 78. doi:10.3389/fncir.2016.00078.

Bagattini, C., Mutanen, T.P., Fracassi, C., Manenti, R., Cotelli, M., Ilmoniemi, R.J., Miniussi, C., Bortoletto, M., 2019. Predicting Alzheimer's disease severity by means of TMS–EEG coregistration. Neurobiol. Aging 80, 38–45. doi:10.1016/j.neurobiolaging.2019.04.008.

Beaulieu, L.D., Flamand, V.H., Massé-Alarie, H., Schneider, C., 2017. Reliability and minimal detectable change of transcranial magnetic stimulation outcomes in healthy adults: A systematic review. Brain Stimul. doi:10.1016/j.brs.2016.12.008.

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 experimental control. Brain Stimul. doi:10.1016/j.brs.2019.01.010.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300.

Biabani, M., Fornito, A., Mutanen, T.P., Morrow, J., Rogasch, N.C., 2019. Characterizing and minimizing the contribution of sensory inputs to TMS-evoked potentials. Brain Stimul 12, 1537–1552. doi:10.1016/j.brs.2019.07.009.

Bortoletto, M., Bonzano, L., Zazio, A., Ferrari, C., Pedull, L., Gasparotti, R., Miniussi, C., Bove, M., 2021. Asymmetric transcallosal conduction delay leads to finer bimanual coordination. 10.1016/j.brs.2021.02.002

Bortoletto, M., Bonzano, L., Zazio, A., Ferrari, C., Pedullà, L., Gasparotti, R., Miniussi, C., Bove, M., 2020. Asymmetric transcallosal conduction delay leads to finer bimanual coordination. bioRxiv doi:10.1101/2020.01.24.918102, 2020.01.24.918102.

Bortoletto, M., Veniero, D., Thut, G., Miniussi, C., 2015. The contribution of TMS-EEG coregistration in the exploration of the human cortical connectome. Neurosci. Biobehav. Rev. 49, 114–124. doi:10.1016/j.neubiorev.2014.12.014.

Botvinik-Nezer, R., Holzmeister, F., Camerer, C.F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J.A., Adcock, R.A., Avesani, P., Baczkowski, B.M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., Benoit, R.G., Berkers, R.M.W.J., Bhanji, J.P., Biswal, B.B., Bobadilla-Suarez, S., Bortolini, T., Bottenhorn, K.L., Bowring, A., Braem, S., Brooks, H.R., Brudner, E.G., Calderon, C.B., Camilleri, J.A., Castrellon, J.J., Cecchetti, L., Cieslik, E.C., Cole, Z.J., Collignon, O., Cox, R.W., Cunningham, W.A., Czoschke, S., Dadi, K., Davis, C.P., Luca, A.De, Delgado, M.R., Demetriou, L., Dennison, J.B., Di, X., Dickie, E.W., Dobryakova, E., Donnat, C.L., Dukart, J., Duncan, N.W., Durnez, J., Eed, A., Eickhoff, S.B., Erhart, A., Fontanesi, L., Fricke, G.M., Fu, S., Galván, A., Gau, R., Genon, S., Glatard, T., Glerean, E., Goeman, J.J., Golowin, S.A.E., González-García, C., Gorgolewski, K.J., Grady, C.L., Green, M.A., Guassi Moreira, J.F., Guest, O., Hakimi, S., Hamilton, J.P., Hancock, R., Handjaras, G., Harry, B.B., Hawco, C., Herholz, P., Herman, G., Heunis, S., Hoffstaedter, F., Hogeveen, J., Holmes, S., Hu, C.P., Huettel, S.A., Hughes, M.E., Iacovella, V., Iordan, A.D., Isager, P.M., Isik, A.I., Jahn, A., Johnson, M.R., Johnstone, T., Joseph, M.J.E., Juliano, A.C., Kable, J.W., Kassinopoulos, M., Koba, C., Kong, X.Z., Koscik, T.R., Kucukboyaci, N.E., Kuhl, B.A., Kupek, S., Laird, A.R., Lamm, C., Langner, R., Lauharatanahirun, N., Lee, H., Lee, S., Leemans, A., Leo, A., Lesage, E., Li, F., Li, M.Y.C., Lim, P.C., Lintz, E.N., Liphardt, S.W., Losecaat Vermeer, A.B., Love, B.C., Mack, M.L., Malpica, N., Marins, T., Maumet, C., McDonald, K., McGuire, J.T., Melero, H., Méndez Leal, A.S., Meyer, B., Meyer, K.N., Mihai, G., Mitsis, G.D., Moll, J., Nielson, D.M., Nilsonne, G., Notter, M.P., Olivetti, E., Onicas, A.I., Papale, P., Patil, K.R., Peelle, J.E., Pérez, A., Pischedda, D., Poline, J.B., Prystauka, Y., Ray, S., Reuter-Lorenz, P.A., Reynolds, R.C., Ricciardi, E., Rieck, J.R., Rodriguez-Thompson, A.M., Romyn, A., Salo, T., Samanez-Larkin, G.R., Sanz-Morales, E., Schlichting, M.L., Schultz, D.H., Shen, Q., Sheridan, M.A., Silvers, J.A., Skagerlund, K., Smith, A., Smith, D.V., Sokol-Hessner, P., Steinkamp, S.R., Tashjian, S.M., Thirion, B., Thorp, J.N., Tinghög, G., Tisdall, L., Tompson, S.H., Toro-Serey, C., Torre Tresols, J.J., Tozzi, L., Truong, V., Turella, L., van 't Veer, A.E., Verguts, T., Vettel, J.M., Vijayarajah, S., Vo, K., Wall, M.B., Weeda, W.D., Weis, S., White, D.J., Wisniewski, D., Xifra-Porxas, A., Yearling, E.A., Yoon, S., Yuan, R., Yuen, K.S.L., Zhang, L., Zhang, X., Zosky, J.E., Nichols, T.E., Poldrack, R.A., Schonberg, T., 2020. Variability in the analysis of a single neuroimaging dataset by many teams. Nature 582, 84–88. doi:10.1038/s41586-020-2314-9.

Casarotto, S., Lauro, L.J.R., Bellina, V., Casali, A.G., Rosanova, M., Pigorini, A., Defendi, S., Mariotti, M., Massimini, M., 2010. EEG responses to TMS are sensitive to changes in the perturbation parameters and repeatable over time. PLoS One 5. doi:10.1371/journal.pone.0010281.

Chung, S.W., Sullivan, C.M., Rogasch, N.C., Hoy, K.E., Bailey, N.W., Cash, R.F.H., Fitzgerald, P.B., 2019. The effects of individualised intermittent theta burst stimulation in the prefrontal cortex: a TMS-EEG study. Hum. Brain Mapp. 40, 608–627. doi:10.1002/hbm.24398.

Conde, V., Tomasevic, L., Akopian, I., Stanek, K., Saturnino, G.B., Thielscher, A., Bergmann, T.O., Siebner, H.R., 2019. The non-transcranial TMS-evoked potential is an inherent source of ambiguity in TMS-EEG studies. Neuroimage 185, 300–312. doi:10.1016/j.neuroimage.2018.10.052.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21. doi:10.1016/j.jneumeth.2003.10.009.

Delorme, A., Palmer, J., Onton, J., Oostenveld, R., Makeig, S., 2012. Independent EEG sources are dipolar. PLoS One 7. doi:10.1371/journal.pone.0030135.

Deng, Z.De, Lisanby, S.H., Peterchev, A.V., 2013. Electric field depth-focality tradeoff in transcranial magnetic stimulation: simulation comparison of 50 coil designs. Brain Stimul 6, 1–13. doi:10.1016/j.brs.2012.02.005.

Duecker, F., Sack, A.T., 2015. Rethinking the role of sham TMS. Front. Psychol. 6. doi:10.3389/fpsyg.2015.00210.

Gordon, P.C., Desideri, D., Belardinelli, P., Zrenner, C., Ziemann, U., 2018a. Comparison of cortical EEG responses to realistic sham versus real TMS of human motor cortex. Brain Stimul 11, 1322–1330. doi:10.1016/j.brs.2018.08.003.

Gordon, P.C., Zrenner, C., Desideri, D., Belardinelli, P., Zrenner, B., Brunoni, A.R., Ziemann, U., 2018b. Modulation of cortical responses by transcranial direct current stimulation of dorsolateral prefrontal cortex: a resting-state EEG and TMS-EEG study. Brain Stimul. 11, 1024–1032. doi:10.1016/j.brs.2018.06.004.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Netw. 13, 411–430. doi:10.1016/S0893-6080(00)00026-5.

Ilmoniemi, R.J., Kičić, D., 2010. Methodology for combined TMS and EEG. Brain Topogr. doi:10.1007/s10548-009-0123-4.

Kerwin, L.J., Keller, C.J., Wu, W., Narayan, M., Etkin, A., 2018. Test-retest reliability of transcranial magnetic stimulation EEG evoked potentials. Brain Stimul. 11, 536–544. doi:10.1016/j.brs.2017.12.010.

King, T.S., Chinchilli, V.M., Wang, K.L., Carrasco, J.L., 2007. A class of repeated measures concordance correlation coefficients. J. Biopharm. Stat. 17, 653–672. doi:10.1080/10543400701329455.

Komssi, S., Kähkönen, S., 2006. The novelty value of the combined use of electroencephalography and transcranial magnetic stimulation for neuroscience research. Brain Res. Rev. doi:10.1016/j.brainresrev.2006.01.008.

Lin, L.I.-K., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255. doi:10.2307/2532051.

Lindquist, M., 2020. Neuroimaging results altered by varying analysis pipelines. Nature doi:10.1038/d41586-020-01282-z.

Lioumis, P., Kičić, D., Savolainen, P., Mäkelä, J.P., Kähkönen, S., 2009. Reproducibility of TMS - Evoked EEG responses. Hum. Brain Mapp. 30, 1387–1396. doi:10.1002/hbm.20608.

Luck, S.J., Steven, J., 2014. An Introduction to the Event-Related Potential Technique. MIT Press.

Makeig, S., Jung, T.P., Bell, A.J., Ghahremani, D., Sejnowski, T.J., 1997. Blind separation of auditory event-related brain responses into independent components. Proc. Natl. Acad. Sci. U. S. A. 94, 10979–10984. doi:10.1073/pnas.94.20.10979.

Martínez-Cagigal, V., 2020. Multiple Testing Toolbox - File Exchange - MATLAB Central [WWW Document]. URL https://it.mathworks.com/matlabcentral/fileexchange/70604-multiple-testing-toolbox (accessed 12.7.20).

Mennemeier, M.S., Triggs, W.J., Chelette, K.C., Woods, A.J., Kimbrell, T.A., Dornhoffer, J.L., 2009. Sham transcranial magnetic stimulation using electrical stimulation of the scalp. Brain Stimul. 2, 168–173. doi:10.1016/j.brs.2009.02.002.

Metsomaa, J., Sarvas, J., Ilmoniemi, R.J., 2014. Multi-trial evoked EEG and independent component analysis. J. Neurosci. Methods 228, 15–26. doi:10.1016/j.jneumeth.2014.02.019.

Mokkink, L.B., Terwee, C.B., Patrick, D.L., Alonso, J., Stratford, P.W., Knol, D.L., Bouter, L.M., de Vet, H.C.W., 2010. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J. Clin. Epidemiol. 63, 737–745. doi:10.1016/j.jclinepi.2010.02.006.

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. doi:10.1016/j.neuroimage.2016.05.028.

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. doi:10.1016/j.neuroimage.2017.10.021.

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. doi:10.7554/eLife.51214.

Nikouline, V., Ruohonen, J., Ilmoniemi, R.J., 1999. The role of the coil click in TMS assessed with simultaneous EEG. Clin. Neurophysiol. 110, 1325–1328. doi:10.1016/S1388-2457(99)00070-X.

Nikulin, V.V., Kičić, D., Kähkönen, S., Ilmoniemi, R.J., 2003. Modulation of electroencephalographic responses to transcranial magnetic stimulation: evidence for changes in cortical excitability related to movement. Eur. J. Neurosci. 18, 1206–1212. doi:10.1046/j.1460-9568.2003.02858.x.

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. doi:10.1155/2011/156869.

Ozdemir, R.A., Tadayon, E., Boucher, P., Momi, D., Karakhanyan, K.A., Fox, M.D., Halko, M.A., Pascual-Leone, A., Shafi, M.M., Santarnecchi, E., 2020. Individualized perturbation of the human connectome reveals reproducible biomarkers of network dynamics relevant to cognition. Proc. Natl. Acad. Sci. U. S. A. 117, 8115–8125. doi:10.1073/pnas.1911240117.

Palmer, J.A., Makeig, S., Kreutz-Delgado, K., Rao, B.D., 2008. Newton method for the ICA mixture model. In: Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, pp. 1805–1808. doi:10.1109/ICASSP.2008.4517982.

Pernet, C.R., Appelhoff, S., Gorgolewski, K.J., Flandin, G., Phillips, C., Delorme, A., Oostenveld, R., 2019. EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. Sci. Data doi:10.1038/s41597-019-0104-8.

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, 5603–5612. doi:10.1523/JNEUROSCI.5089-13.2014.

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 doi:10.1016/j.neuroimage.2016.10.031.

Rogasch, N.C., Thomson, R.H., Farzan, F., Fitzgibbon, B.M., Bailey, N.W., Hernandez-Pavon, J.C., Daskalakis, Z.J., 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. doi:10.1016/j.neuroimage.2014.07.037.

Rogasch, N.C., Zipser, C., Darmani, G., Mutanen, T.P., Biabani, M., Zrenner, C., Desideri, D., Belardinelli, P., Müller-Dahlhaus, F., Ziemann, U., 2020. The effects of NMDA receptor blockade on TMS-evoked EEG potentials from prefrontal and parietal cortex. Sci. Rep. 10, 3168. doi:10.1038/s41598-020-59911-6.

Romero Lauro, L.J., Rosanova, M., Mattavelli, G., Convento, S., Pisoni, A., Opitz, A., Bolognini, N., Vallar, G., 2014. TDCS increases cortical excitability: direct evidence from TMS-EEG. Cortex 58, 99–111. doi:10.1016/j.cortex.2014.05.003.

Rosanova, M., Gosseries, O., Casarotto, S., Lanie Boly, M., Casali, A.G., Lie Bruno, M.-A., Mariotti, M., Boveroux, P., Tononi, G., Laureys, S., Massimini, M., 2012. Recovery of cortical effective connectivity and recovery of consciousness in vegetative patients. Brain 135, 1308–1320. doi:10.1093/brain/awr340.

Rossi, S., Hallett, M., Rossini, P.M., Pascual-Leone, A.Safety of TMS Consensus Group, T.S. of T.C., 2009. Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. Clin. Neurophysiol. 120, 2008–2039. doi:10.1016/j.clinph.2009.08.016.

Sammet, S., 2016. Magnetic resonance safety. Abdom. Radiol. 41, 444–451. doi:10.1007/s00261-016-0680-4.

Shrout, P.E., 1998. Measurement reliability and agreement in psychiatry. Stat. Methods Med. Res. 7, 301–317. doi:10.1177/096228029800700306.

Siebner, H.R., Conde, V., Tomasevic, L., Thielscher, A., Bergmann, T.O., 2019. Distilling the essence of TMS-evoked EEG potentials (TEPs): a call for securing mechanistic specificity and experimental rigor. Brain Stimul. doi:10.1016/j.brs.2019.03.076.

Steegen, S., Tuerlinckx, F., Gelman, A., Vanpaemel, W., 2016. Increasing transparency through a multiverse analysis. Perspect. Psychol. Sci. 11, 702–712. doi:10.1177/1745691616658637.

ter Braack, E.M., de Vos, C.C., van Putten, M.J.A.M., 2015. Masking the auditory evoked potential in TMS–EEG: a comparison of various methods. Brain Topogr. 28, 520–528. doi:10.1007/s10548-013-0312-z.

Tiitinen, H., Virtanen, J., Ilmoniemi, R.J., Kamppuri, J., Ollikainen, M., Ruohonen, J., Näätänen, R., 1999. Separation of contamination caused by coil clicks from responses elicited by transcranial magnetic stimulation. Clin. Neurophysiol. 110, 982–985. doi:10.1016/S1388-2457(99)00038-3.

Tremblay, S., Rogasch, N.C., 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. doi:10.1016/j.clinph.2019.01.001.

Veniero, D., Bortoletto, M., Miniussi, C., 2013. Cortical modulation of short-latency TMS-evoked potentials. Front. Hum. Neurosci. doi:10.3389/fnhum.2012.00352.

Veniero, D., Bortoletto, M., Miniussi, C., 2009. TMS-EEG co-registration: on TMS-induced artifact. Clin. Neurophysiol. 120, 1392–1399. doi:10.1016/j.clinph.2009.04.023.

Voineskos, D., Blumberger, D.M., Zomorrodi, R., Rogasch, N.C., Farzan, F., Foussias, G., Rajji, T.K., Daskalakis, Z.J., 2019. Altered transcranial magnetic stimulation–electroencephalographic markers of inhibition and excitation in the dorsolateral prefrontal cortex in major depressive disorder. Biol. Psychiatry 85, 477–486. doi:10.1016/j.biopsych.2018.09.032.

Wilcox, R.R., 2009. Comparing Pearson correlations: dealing with heteroscedasticity and nonnormality. Commun. Stat. - Simul. Comput. 38, 2220–2234. doi:10.1080/03610910903289151.

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, 1607–1625. doi:10.1002/hbm.23938.